

УНИВЕРЗИТЕТ У БЕОГРАДУ
ГРАЂЕВИНСКИ ФАКУЛТЕТ



ЈОВАН Д. КОВАЧЕВИЋ

**КАРТИРАЊЕ ШУМСКЕ ВЕГЕТАЦИЈЕ
НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ
ОСМАТРАЊА ЗЕМЉЕ КОРИШЋЕЊЕМ
ТЕХНИКА МАШИНСКОГ УЧЕЊА**

докторска дисертација

БЕОГРАД 2021.

**UNIVERSITY OF BELGRADE
FACULTY OF CIVIL ENGINEERING**



JOVAN D. KOVAČEVIĆ

**MAPPING FOREST VEGETATION
FROM SATELLITE EARTH
OBSERVATION DATA USING
MACHINE LEARNING TECHNIQUES**

doctoral dissertation

BELGRADE 2021

ПОДАЦИ О МЕНТОРУ И ЧЛАНОВИМА КОМИСИЈЕ

Ментор: В. проф. др Жељко Цвијетиновић, дипл. инж. геод.
Универзитет у Београду, Грађевински факултет

Комисија: В. проф. др Драган Михајловић, дипл. инж. геод.
Универзитет у Београду, Грађевински факултет

В. проф. др Љиљана Брајовић, дипл. инж. ел.
Универзитет у Београду, Грађевински факултет

Доц. др Милева Самарџић-Петровић, дипл. инж. геод.
Универзитет у Београду, Грађевински факултет

Проф. др Дмитар Лакушић
Универзитет у Београду, Биолошки факултет

В. проф. др Борис Радић, дипл. инж. пејз. арх.
Универзитет у Београду, Шумарски факултет

Датум одбране докторске дисертације: _____

*Посвећено мојој породици и
пријатељима.*

ЗАХВАЛНИЦА

Најпре желим да захвалим професорима др Жељку Цвијетиновићу и др Драгану Михајловићу. Поред помоћи и подршке приликом израде ове дисертације, њима дугујем и захвалност што увек радо деле своја знања и искуства са мном, како професионална, тако и животна. Они су ми помогли да препознам свој позив, али и да заволим посао „са друге стране катедре“.

Захвалност дугујем и свим члановима комисије на бројним корисним коментарима и сугестијама, који су учинили да ова дисертација буде квалитетнија. Истакао бих бројне конструктивне дискусије уз кафу са др Милевом Самарџић-Петровић. Посебно сам захвалан колеги и пријатељу Николи Станчићу на помоћи и бројним саветима, који је много времена одвојио за моје истраживање и текстове и увек био ту да саслуша моје дилеме и недоумице. Захваљујем и Момиру Митровићу, др Невени Кузмановић и фирми МапСофт, са којима сам, са задовољством, сарађивао у разним фазама овог истраживања.

Колегама са Катедре и Института за геодезију и геоинформатику захваљујем што су преузимали моје обавезе када је било неопходно и тиме олакшали писање ове дисертације. Посебно ми је значило уступање рачунара Лабораторије за развој геопросторних технологија отвореног кода приликом реализације захтевнијих експеримената овог истраживања.

Овим путем такође желим да изразим и поштовање тиму лекара и особља са Института за ортопедију Бањица који су омогућили да дисертацију дочекам и браним „на ногама“.

Захвалан сам и свим блиским пријатељима.

На крају, највећу захвалност упућујем својој породици: оцу Душану, мајци Добрили, сестрама Милици и Милани и вереници Данијели. Хвала вам за стрпљење, разумевање, љубав и подршку током свих ових година, не верујем да бих успео без вас!

Јован Ковачевић

САЖЕТАК

КАРТИРАЊЕ ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ КОРИШЋЕЊЕМ ТЕХНИКА МАШИНСКОГ УЧЕЊА

Потребе за квалитетним подацима квантитативних и квалитативних карактеристика шума се повећавају како је притисак на овај природни ресурс све већи. Подаци сателитског осматрања Земље су се показали као погодна алтернатива теренским методама за прикупљање информација о шумама. Основни циљеви истраживања у оквиру ове дисертације су анализа, развој и унапређење методологије за картирање типова шумске вегетације на основу података сателитског осматрања Земље. Предложена методологија се у основи ослања на издвајање тематских информација из сателитских снимака применом надзираних непараметарских метода класификације, посебно техника машинског учења. Истраживање обухвата утврђивање применљивости различитих техника машинског учења (стабла одлучивања, случајне шуме и градијентног појачавања) за картирање типова шумске вегетације, анализу различитих репрезентација података ради наглашавања просторних и временских аспеката у класификационом моделу, као и примену техника за креирање оптималног балансираног скупа података и избора атрибута од значаја за квалитетно картирање шумске вегетације. Развој и тестирање предложене методологије је извршено идентификацијом осам типова шумске вегетације на подручју Републике Србије. Коришћени су подаци сателитске мисије Sentinel-2 из 2019. године у комбинацији са додатним изворима података. Резултати спроведених експеримената се подударају са резултатима сличних истраживања заснованих на оптичким сателитским опажањима високе просторне резолуције. На основу спроведених експеримената је закључено да је предложена методологија применљива за картирање типова шумске вегетације у Србији. Иако је тестирање извршено на територији Србије, методологија има значајан потенцијал да се примени и на другим подручјима уз прилагођавање класификационе шеме и допуне података за тренирање модела. Такође, може се сматрати да је предложена методологија општег типа, тј. она може бити релативно једноставно прилагођена и другим класификационим проблемима. Додатна истраживања су предвиђена у будућности како би се испитале претходне претпоставке.

Кључне речи: класификација, машинско учење, шумска вегетација, картирање, даљинска детекција, подаци сателитског осматрања Земље, Србија

Научна област: Геодезија

Уже научне области: Фотограмetriја и даљинска детекција, Геоинформатика

ABSTRACT

MAPPING FOREST VEGETATION FROM SATELLITE EARTH OBSERVATION DATA USING MACHINE LEARNING TECHNIQUES

The necessity for quality data on quantitative and qualitative characteristics of forests is increasing as the pressure on this natural resource rises. Satellite Earth observation data has proven to be a suitable alternative to field methods when it comes to collecting information about forests. The main goals of the research within this dissertation are the analysis, development and improvement of the methodology for mapping forest vegetation types using satellite Earth observation data. The proposed methodology relies on the extraction of thematic information from satellite imagery using supervised non-parametric classification methods, especially machine learning techniques. The research includes determining the applicability of different machine learning techniques (decision tree, random forest and gradient boosting) for mapping forest vegetation types, analysis of different data representations for emphasizing spatial and temporal aspects in the classification model, and applying techniques to create optimal balanced datasets and to determine optimal features for quality mapping of forest vegetation. The development and testing of the proposed methodology were performed by identifying eight forest vegetation types in the Republic of Serbia. Sentinel-2 satellite imagery from the 2019 was used in combination with additional data sources. The results of the conducted experiments match the results of similar studies which were based on high-resolution optical satellite imagery. In accordance with the conducted experiments, it can be concluded that the proposed methodology is applicable for mapping forest vegetation types in Serbia. While the testing was performed on the territory of Serbia, the methodology has significant potential to be applied in other regions as well, with the adjustment of the classification scheme and the additional data for model training. Also, the proposed methodology can be considered to be of a generic type, since it can be relatively easily adapted to other classification problems. Additional research is planned in the future to test these assumptions.

Keywords: classification, machine learning, forest vegetation, mapping, remote sensing, satellite Earth observation, Serbia

Scientific field: Geodesy

Scientific subfields: Photogrammetry and Remote Sensing, Geoinformatics

САДРЖАЈ

ЗАХВАЛНИЦА	V
САЖЕТАК	VII
ABSTRACT	IX
САДРЖАЈ	XI
СПИСАК СЛИКА	XIV
СПИСАК ТАБЕЛА	XVI
1 УВОД	1
1.1 Поставка проблема	4
1.2 Циљеви	6
1.3 Приступ	6
1.4 Кратак преглед	8
2 ИЗДВАЈАЊЕ ИНФОРМАЦИЈА О ШУМАМА НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ	10
2.1 Сателити за осматрање Земље	10
2.1.1 Пасивни оптички сателитски системи	12
2.1.2 Активни радарски сателитски системи	14
2.2 Могућности сателитског осматрања Земље за прикупљање информација о шумама	16
2.2.1 Одређивање обухвата и типова шума	18
2.2.2 Издвајање биофизичких шумских карактеристика	21
2.2.2.1 Индекс површине листа	23
2.2.2.2 Надземна биомаса	24
2.2.2.3 Бруто и нето примарна производња	25
2.2.2.4 Карактеристике инвентуре шума	26

3	КЛАСИФИКАЦИЈА САТЕЛИТСКИХ СНИМАКА ТЕХНИКАМА МАШИНСКОГ УЧЕЊА..	29
3.1	Технике машинског учења.....	30
3.1.1	Стабло одлучивања	32
3.1.2	Случајна шума	36
3.1.3	Градијентно појачавање	37
3.2	Временска компонента у класификационом моделу	38
3.3	Просторна компонента у класификационом моделу	41
3.4	Креирање и избор репрезентативних атрибута	43
3.4.1	Филтер-методе за избор репрезентативних атрибута.....	44
3.4.1.1	Избор атрибута заснован на корелацији.....	44
3.4.2	Омотач-методе за избор репрезентативних атрибута	45
3.4.2.1	Избор репрезентативних атрибута применом генетских алгоритама	47
3.4.2.2	Секвенцијалне методе избора репрезентативних атрибута	50
3.5	Балансирање класа	50
3.6	Оцена квалитета класификације.....	51
3.6.1	Квантитативне мере перформанси класификационог модела.....	52
3.6.2	Крос-валидација са изостављањем комплетних локација	54
4	КАРТИРАЊЕ ТИПОВА ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ	56
4.1	Сажетак предложене методологије	56
4.2	Подручје истраживања.....	60
4.3	Подаци	62
4.3.1	Полигони типова шумске вегетације	62
4.3.2	Сателитска опажања мисије <i>Sentinel-2</i>	64
4.3.3	Допунски извори података.....	67

4.3.3.1	EU-DEM подаци висинске представе терена.....	68
4.3.3.2	Карта основног земљишног покривача Републике Србије ..	68
4.3.3.3	Коперникус паневропски слојеви високе резолуције о шумама	69
4.4	Припрема и иницијална анализа података	69
4.5	Софтверска подршка	75
5	РЕЗУЛТАТИ И ДИСКУСИЈА.....	79
5.1	Експериментално одређивање оптималних параметара методологије.....	79
5.1.1	Експеримент 1: Укључивање временске компоненте у класификациони модел	81
5.1.2	Експеримент 2: Избор репрезентативних атрибута.....	85
5.1.3	Експеримент 3: Утицај балансирања класа.....	95
5.1.4	Експеримент 4: Наглашавање просторних аспеката у класификационом моделу	97
5.2	Оптимални параметри методологије	99
5.3	Коначна карта шумске вегетације.....	104
5.4	Валидација коришћењем независног скупа података	109
6	ЗАКЉУЧАК	117
	ЛИТЕРАТУРА	124
	ПРИЛОЗИ	138

СПИСАК СЛИКА

Слика 1. Разврставање метода класификације на основу различитих критеријума	30
Слика 2. Општи изглед стабла одлучивања	33
Слика 3. Одабир подесног атрибута за нови чвор стабла одлучивања.....	34
Слика 4. Илустрација репрезентације података сателитског опажања са имплицитно и експлицитно дефинисаном временском компонентом	39
Слика 5. Илустрација Муровог суседства посматраног пиксела величине 3×3 и 5×5...	42
Слика 6. Омотач-методе за избор репрезентативних атрибута	46
Слика 7. Илустрација бинарно кодираних атрибута генетског алгоритма.....	47
Слика 8. Општи принцип функционисања генетског алгоритма.....	48
Слика 9. Илустрација процедуре петоструке крос-валидације.....	55
Слика 10. Графичка илустрација корака предложене методологије за картирање типова шумске вегетације	59
Слика 11. Локација подручја истраживања – Република Србија	60
Слика 12. Рељеф подручја истраживања	61
Слика 13. Просторни распоред креираних полигона типова шумске вегетације	64
Слика 14. Временске одреднице одабраних S2 гранула	66
Слика 15. Просторни распоред расположивих S2 опажања вегетације на подручју истраживања	67
Слика 16. Графици фенолошких спектралних образаца различитих типова шумске вегетације	73
Слика 17. Кутијаста дијаграм типова шумске вегетације и вредности висине терена...	75

Слика 18. Поређење укупне тачности различитих приступа укључивања временске компоненте	82
Слика 19. Поређење капа коефицијента слагања различитих приступа укључивања временске компоненте.....	82
Слика 20. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом стабла одлучивања.....	94
Слика 21. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом случајне шуме.....	94
Слика 22. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом градијентног појачавања.....	95
Слика 23. Допринос корака методологије постигнутим мерама перформанси оптималних класификационих модела са имплицитном временском репрезентацијом	101
Слика 24. Допринос корака методологије постигнутим мерама перформанси оптималних класификационих модела са експлицитном временском репрезентацијом	101
Слика 25. Коначна карте типова шумске вегетације применом модела са имплицитном временском компонентом.....	105
Слика 26. Коначна карте типова шумске вегетације применом модела са експлицитном временском компонентом.....	106
Слика 27. Референтни растер креиран за валидацију карата типова шумске вегетације	111
Слика 28. Просторни распоред тачно и погрешно класификованих пиксела карте са имплицитном временском дефиницијом у односу на референтни растер.....	113
Слика 29. Просторни распоред тачно и погрешно класификованих пиксела карте са експлицитном временском дефиницијом у односу на референтни растер.....	114

СПИСАК ТАБЕЛА

Табела 1. Приказ пасивних оптичких сателитских мисија/сензора са најважнијим карактеристикама (ПР – просторна резолуција, ВР – временска резолуција).....	13
Табела 2. Списак SAR система са најважнијим карактеристикама (ПР – просторна резолуција, ВР – временска резолуција).....	16
Табела 3. Општи изглед матрице конфузије	52
Табела 4. Дефиниција класа и њихова индивидуална заступљеност изражена кроз број креираних полигона и њихове укупне површине.....	63
Табела 5. Карактеристике мултиспектралног скенера <i>Sentinel-2</i> сателитске мисије...	65
Табела 6. Скуп карактеристика издвојених за сваки пиксел подручја истраживања ...	70
Табела 7. Дефиниција класа и број расположивих опажања.....	72
Табела 8. Најзначајније <i>R</i> библиотеке коришћене за потребе овог истраживања	76
Табела 9. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – временска компонента дефинисана као дан у години (<i>[day]</i>)	86
Табела 10. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – временска компонента дефинисана као дан и месец (<i>[day + month]</i>).....	87
Табела 11. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – имплицитно дефинисана временска компонента	88
Табела 12. Највише постигнуте мере перформанси и скупови атрибута који су коришћени за сваки приступ репрезентације временске компоненте	92
Табела 13. Комбинације параметара чијим коришћењем су постигнуте највише мере перформанси (скуп података је балансиран применом <i>SMOTE</i> алгоритма).....	96

Табела 14. Комбинација параметара којима су постигнуте највише мере перформанси уз наглашавање просторних компоненти у класификационом моделу.....	98
Табела 15. Комбинација параметара методологије којима су постигнуте највише мере перформанси	100
Табела 16. Вредности метрика оптималних класификационих модела	102
Табела 17. Површине и удео заступљености класа шумске вегетације у коначним картама	107
Табела 18. Информације о референтном скупу података.....	110
Табела 19. Матрица конфузије карте са имплицитном временском дефиницијом.....	112
Табела 20. Матрица конфузије карте са експлицитном временском дефиницијом ...	112

1

УВОД

Шуме представљају најдоминантнији тип копнених екосистема који покрива око 31 одсто укупне површине Земље (*Global Forest Resources Assessment 2020, 2020*). Ту се убрајају и природне и делимично природне шуме које представљају најсложеније и најразноврсније екосистеме на Земљи. Појам „шума“ се у овој докторској дисертацији користи у смислу *EUNIS* дефиниције (*Davies и Moss, 2002*), која је примењена и у националном Правилнику о стаништима (*Службени њласник РС, 35/2010-29, 2010*). Дакле, под шумом, односно шумским стаништем подразумева се „станиште са доминантном вегетацијом дрвећа са типично појединачним стаблима вишим од 5 m и покровношћу крошњи већом од 10%“.

Потребе за квалитетним подацима о квантитативним и квалитативним карактеристикама шума се повећавају како је притисак на овај природни ресурс све већи (*Панџић и Бороџа, 2015*). Инвентаризација шумских ресурса је државни интерес, због чега велики број земаља спроводи националне инвентуре шума (енгл. *National Forest Inventories*) (*Tomppo и др., 2010*). На овај начин се омогућава рационално коришћење свих шумских потенцијала, обезбеђују основе за привређивање у шумарству и пружа могућност планирања, контроле и управљања, како у шумарству, тако и у другим привредним гранама (*Банковић и др., 2009а; Панџић и Бороџа, 2015*).

Инвентаризација шума се претежно спроводи ради одрживог управљања и експлоатације шумских ресурса (*European Environmental Agency, 2007; Ferreira и др., 2020*). Заштита животне средине и биодиверзитета је свакако једна од подразумеваних

ставки одрживог управљања шумским ресурсима. Међутим, све је већи број кампања прикупљања карактеристика о шумама код којих је аспект заштите животне средине примаран и доминантан (Lucas и др., 2015; Richard и др., 2014; Zlinszky и др., 2015). Досадашња искуства у заштити биодиверзитета су показала да без адекватне заштите станишта нема адекватне заштите биодиверзитета на специјском и генетичком нивоу па, самим тим, нема ни адекватне заштите природе у целини. Сходно томе, ове истраживачке активности шумску вегетацију, пре свега, посматрају као основну структурну и функционалну компоненту шумских станишта. Овај приступ је формално промовисан у кључним међународним документима, као што су Конвенција о биодиверзитету (*The United Nations, Rio de Janeiro, 1992*), Бернска конвенција (*Council of Europe, Berne, 1979*) и Директива о стаништима (*Council of the European Union, 1992*). Додатно, шумска вегетација је један од виталних елемената структуре предела. Зато је утврђивање карактеристика њене „просторности“ (композиције и конфигурације), важна процедура у процесу карактеризације предела, као и имплементације концепта зелене инфраструктуре (Caudullo и др., 2013; García-Llamas и др., 2016; Proulx и Fahrig, 2010).

Подаци које се прикупљају приликом инвентура шума могу се груписати на: административно-географске податке, податке о станишту, податке о састојини и податке о стаблима (Панџић и Бороша, 2015). У зависности од усвојене методологије, различите државе приликом инвентаризације шума прикупљају различите податке о шумама или су исти подаци у различитим државама прикупљени коришћењем различитих метода и параметара (Gschwantner и др., 2016; Tomppo и др., 2010; Панџић и Бороша, 2015). Pause и др. (2016) су идентификовали скуп најчешће прикупљених вегетационих шумских карактеристика присутних у различитим европским инвентурама шума, заједно са методама за њихово прикупљање које су се показале као добра пракса.

Подаци о карактеристикама шума се традиционално прикупљају различитим теренским методама. Теренске методе се по правилу комбинују са неким видом стратификованог (кластеризованог) узорковања како би се на оптималан начин постигла покривеност на великим подручјима (Pause и др., 2016). Прикупљени теренски подаци служе како би се одредиле дрвне метрике попут површине под шумом и шумским земљиштем, запремине шумског фонда, али и метрике које се односе на

здравље и биодиверзитет шума (Zeng и др., 2015; Панџић и Бороџа, 2015). Велики број држава у оквиру инвентуре шума врши и процену биомасе, индекса површине листа (енгл. *Leaf Area Index – LAI*) и залиха угљеника присутних у живим шумским стаблима (Tomppo и др., 2010, 2008; Wulder, 1998).

Поред инвентура шумских ресурса због којих се примарно прикупљају, подаци о карактеристикама шума имају и бројне друге примене. Из перспективе даљинске детекције, овако креиране базе података омогућавају калибрацију и валидацију модела за добијање различитих сателитских продуката (Hansen и др., 2017; Kirchhoefer и др., 2017; Novo-Fernández и др., 2019). Велики број активних сателитских сензора намењених за осматрање Земље пружа широк спектар могућности када је реч о прикупљању информација о карактеристикама шума. Комбиновање теренских и сателитских извора података доводи и до додатних информација које могу бити посебно погодне у случајевима када појединачни извор није довољно тачан (Tomppo и др., 2008; Zeng и др., 2015). Поједине државе, попут Финске и Шведске, већ врше инвентуру шума засновану на више извора података, где се стратификовани теренски узорци комбинују са сателитским подацима и другим расположивим изворима (Tomppo и др., 2008).

Boyd и Danson (2005) препознају три нивоа карактеристика шума које се могу одредити на основу података сателитског осматрања Земље. На првом нивоу, подаци сателитског осматрања Земље се могу искористити за одређивање обухвата (граница) шума. Следећи ниво јесте одређивање информација о типовима шума до нивоа доминантних врста дрвећа. Последњи, трећи ниво, подразумева издвајање биофизичких и биохемијских шумских карактеристика (Boyd и Danson, 2005). Поред тога, треба узети у обзир и читав низ додатних информација које се могу одредити о шумама, попут идентификације врста и/или крошњи дрвећа, детекције промена и издвајања ивичних шумских зона (Lutz и др., 2008). Овако издвојене информације имају бројне примене, као што су одређивање количина и продуктивности шумског фонда, праћење здравља и промена шумског покривача, одређивање оштећења шума услед пожара или других непогода и штеточина, праћење „ивичног ефекта“ шуме као индикатора и генератора биодиверзитета и стабилности предела (Ahamed и др., 2011; Bayr и др., 2016; Galiatsatos и др., 2020; Li и др., 2018; Matlack, 1994; Pause и др., 2016; Xiao и др., 2019).

Технике класификације садржаја сателитских снимака су доминантне када је у питању одређивање обухвата и типова шума. О томе сведочи велики број успешних примена,

које се ослањају на широк дијапазон сателитских сензора различитих карактеристика (Cheng и Wang, 2019; Martone и др., 2018; Sheeren и др., 2016). Треба разликовати надзиране, ненадзиране и хибридне методе класификације. Надзиране методе подразумевају априори знање о појавама које се картирају, а које се најчешће обезбеђује кроз скуп тзв. „тренинг“ података из којих алгоритам учи о циљној појави. Ненадзиране методе врше груписање (кластерисање) појава према одређеним мерама сличности без икаквог претходног знања о самим појавама. Све методе класификације се могу разврстати и у параметарске и непараметарске методе, у зависности од тога да ли уводе одређене статистичке претпоставке. Параметарске методе се ослањају на теорију статистичког одлучивања, којом се преко скупа параметара претпоставља одређена статистичка расподела у подацима. Непараметарске методе, с друге стране, одлучују на основу правилности уочених у подацима без статистичких претпоставки и ограничења. Методама класификације које користе технике машинског учења, и то посебно непараметарским методама, се поклања посебна пажња у литератури о даљинској детекцији (Ferreira и др., 2020; Grabska и др., 2020; Marrs и Ni-Meister, 2019). Главне предности ове групе метода класификације су способност да процесирају изузетно велике количине података из различитих извора, могућност укључивања нумеричких и категоричких атрибута у класификациони модел, као и робусност када је у подацима присутан шум (Sun и Scanlon, 2019).

Ово истраживање је примарно фокусирано на решавање проблема картирања типова шумске вегетације на основу података сателитског осматрања Земље. У фокусу су надзиране непараметарске методе класификације које спадају у групу техника машинског учења. Дисертација даје систематичан преглед и могућности употребе сателитских опажања Земље ради издвајања других информација о шумама. Анализом релевантне литературе обезбеђено је сагледавање актуелних трендова истраживања, што омогућава добро разумевање ширег контекста и места посматраног проблема у односу на сродне области истраживања.

1.1 Поставка проблема

Директива о стаништима је кључни законски инструмент која државе Европске уније (ЕУ) обавезује да на својој територији идентификују и одреде простор за кључна станишта која ће бити укључена у међународну *Natura 2000* мрежу (Council of the

European Union, 1992). Подручја укључена у еколошке мреже треба да обезбеде услове за опстанак на хиљаде ретких и угрожених биљних и животињских врста које зависе од ових станишта. Исти законски инструменти обавезују да се врши праћење стања и да се у регуларним временским интервалима објављују извештаји о статусу очувања идентификованих заштићених подручја и њихових станишта. Република Србија у оквиру приступних преговора за чланство у Европској унији има обавезу да изврши идентификацију и успостављање еколошки значајних подручја која ће приступањем ЕУ постати део мреже *Natura 2000*. Најважнији национални правни документи који се баве овом темом су Уредба о еколошкој мрежи (*Службени њласник РС, 102/2010-10, 2010*) и Правилник о стаништима Србије (*Службени њласник РС, 35/2010-29, 2010*).

У последњих неколико година је спроведен већи број теренских кампања прикупљања података о дистрибуцији типова станишта на територији Републике Србије. Распожива база података представља веома важан извор података. Комбиновање података из ове базе података са подацима сателитског осматрања Земље нуди интересантне могућности. Идејно, на овај начин је могуће извршити квалитетну идентификацију потенцијално значајних еколошких подручја на целој територији Србије. Овај задатак је практично немогуће спровести само уз теренске кампање. С тим у вези, ово истраживање је фокусирано на развој методологије за издвајање и категоризацију типова шумске вегетације на територији Републике Србије. Шумска вегетација је одабрана као циљна због великих подручја које она покрива на територији Републике Србије, али и зато што тренутно скоро трећину типова станишта *Natura 2000* мреже чине управо шумска станишта. Идентификација типова шумске вегетације зато мора бити усклађена са Правилником о стаништима Србије (*Службени њласник РС, 35/2010-29, 2010*), што омогућава директну конверзију номенклатуре из правилника у типологију која се користи за потребе успостављања мреже *Natura 2000*.

Ипак, картирање типова шумске вегетације на основу података сателитског осматрања Земље и даље представља изазов. Шумска подручја могу бити веома хетерогена, што се манифестује кроз веома комплексне и преклапајуће спектралне одзиве, посебно изражене у тропским подручјима (*Grabska и др., 2019*). Разноликост спектралних карактеристика је последица структуре и карактеристика присутне вегетације, али и разлика у условима раста, густини и старости вегетације, све у комбинацији са топографским ефектима. Како се подручје истраживања повећава, ови ефекти су

присутнији и израженији, резултирају израженим нивоима хетерогености предела, што додатно отежава разликовање типова шума (Lutz и др., 2008; Wu и др., 2000). Последично, универзално решење за картирање типова шумске вегетације на глобалном нивоу и за било који ниво детаљности и даље не постоји. Уместо тога, изналажење оптималног решења подразумева дефинисање локално прилагођене методологије, којом се обухватају све специфичности и карактеристике подручја истраживања.

1.2 Циљеви

Основни циљеви истраживања у оквиру ове дисертације су анализа, развој и унапређење методологије за картирање типова шумске вегетације на основу података сателитског осматрања Земље. Наведени приступ се базира на техникама машинског учења, уз респектовање просторних и временских аспеката.

Из наведеног циља произилазе следећи парцијални циљеви истраживања:

- 1) анализа јавно доступних података сателитских мисија који су употребљиви за картирање шумске вегетације;
- 2) идентификација карактеристика шума које је могуће одредити и пратити на основу података сателитског осматрања Земље;
- 3) анализа постојећих метода за картирање шумске вегетације на основу података сателитског осматрања Земље;
- 4) идентификација и анализа проблема који се јављају приликом картирања шумске вегетације на основу података сателитског осматрања Земље;
- 5) успостављање методологије за картирање шумске вегетације на основу података сателитског осматрања Земље коришћењем техника машинског учења;
- 6) израда карте шумске вегетације за шумска подручја Републике Србије применом предложене методологије.

1.3 Приступ

Методолошки, приступ за картирање шумске вегетације се базира на техникама класификације садржаја сателитских снимака. Као примарни извор су коришћени подаци сателитског осматрања Земље, а у фокусу су непараметарске надзиране методе класификације које припадају техникама машинског учења. Конкретно, у оквиру

дисертације је развијена методологија картирања шума на основу података *Sentinel* сателитских мисија. Коришћене су традиционалне технике машинског учења које су се показале као најефикаснији класификатори у досадашњим проблемима класификације, као и скупови више модела машинског учења (енгл. *ensemble learning*). Поред оригиналних података *Sentinel* сателитских мисија, укључивање и креирање и додатних извора података као атрибута у класификационом моделу, попут података о топографији терена, се показало као неопходно и веома важно. Истовремено, ради смањења димензионалности података, из комплетног скупа креираних атрибута је потребно спровести и селекцију најинформативнијих атрибута са аспекта картирања типова шумске вегетације.

Велики број истраживања је показао да коришћење временских серија сателитских података повећава могућност раздвајања различитих типова вегетације у поређењу с коришћењем само једне епохе сателитских опажања. Укључивањем временских серија сателитских снимака користе се разлике у фенологији вегетације, што има позитиван ефекат на могућност међусобног разликовања вегетације (*Cheng и Wang, 2019; Grabska и др., 2020; Sheeren и др., 2016*). Класификација коришћењем техника машинског учења у општем случају не укључује временске аспекте у класификационом моделу, а најчешће решење подразумева креирање таквог модела где се спектрални канали сваке серије сателитских снимака представљају као независни атрибути. На овај начин се број атрибута значајно повећава како расте број расположивих временских серија, што доводи до проблема са меморијским и процесорским захтевима. Додатно, овакав приступ подразумева да не сме бити недостајућих података у коришћеним/одабраним временским серијама. Када се ради о оптичким сателитским снимцима код којих се може јавити проблем недостајућих података због заклоњености облацима, овај проблем постаје доминантан, нарочито када се врши картирање на већем подручју. Како се у овом истраживању као извор података подразумевају и оптички подаци *Sentinel-2* сателитске мисије, предложена методологија мора да узме у обзир и временске аспекте. Међутим, истовремено се мора предвидети и решење за проблем недостајућих података услед заклоњености облацима.

Поред временских аспеката, неопходно је разматрати и нагласити и просторне аспекте у процедури класификације. *Aonpong и др. (2016)* су постигли повећање квалитета класификације земљишног покривача укључивањем суседних пиксела у модел, а

Samardžić-Petrović и др. (2016) су показали да информација о најчешћој класи у суседству посматраног пиксела представља један од најинформативнијих атрибута у случају одређивања краткорочних промена начина коришћења земљишта. У складу са наведеним примерима истраживања, очекивано је да укључивање суседства пиксела у класификациону процедуру може помоћи приликом картирања шумске вегетације.

Практични рад и тестирање методологије је извршено на подручју Републике Србије, а циљни типови шумске вегетације су усклађени са националним Правилником о стаништима Србије (*Службени гласник РС, 35/2010-29, 2010*). То практично значи да је на почетку неопходно дефинисање оптималне класификационе шеме, тј. издвајање и груписање оних типова шумске вегетације чије је картирање могуће извршити на основу података сателитског осматрања Земље. Избор шумских класа мора бити рађен итеративно до коначног усвајања једне или више класификационих шема, а за сваку издвојену класу су неопходни репрезентативни тренинг подаци.

1.4 Кратак преглед

Резултати истраживања су у овој дисертацији изложени кроз шест поглавља. Поглавља су конципирана тако да се надовезују једно на друго, пратећи логичан след излагања материје и најважнијих аспеката спроведеног истраживања.

Поглавље 1 садржи кратак преглед садржаја и циљева дисертације. Ово поглавље на почетку упознаје читаоца са мотивацијом, идентификацијом и поставком проблема истраживања. У наставку су изложени обухват и главни циљеви истраживачког рада, као и примењени приступ. Поглавље се завршава кратким прегледом дисертације.

Поглавље 2 је посвећено могућностима издвајања информација о шумама на основу података сателитског осматрања Земље. То је урађено кроз детаљан преглед литературе у вези са претходним истраживањима у овој и сродним областима истраживања. У првом делу поглавља су представљене најважније карактеристике сателита за осматрање Земље, приказане су две главне групе сензора и наведени и описани њихови најзначајнији представници. Други део поглавља је фокусиран на представљање могућности, достигнућа и актуелних трендова употребе података сателитског осматрања Земље за одређивање обухвата и типова шума, као и за одређивање биофизичких карактеристика шума.

Теоретска позадина и најважније карактеристике класификације садржаја сателитских снимака техникама машинског учења су изложени у оквиру **поглавља 3**. Описана је процедура класификације података сателитског опажања Земље коришћењем техника машинског учења, уз посебан нагласак на респектовању временске и просторне компоненте у класификационом моделу. Важни обрађени аспекти класификационе процедуре јесу и креирање и избор скупа репрезентативних атрибута, као и решавање проблема небалансираности (неравномерне заступљености) класа у подацима. Поглавље се завршава описом коришћене процедуре за оцену квалитета извршене класификације као кључног апарата за одређивање оптималних параметара предложене методологије.

Поглавље 4 описује структуру и најважније кораке предложене методологије за картирање типова шумске вегетације на основу података сателитског осматрања Земље коришћењем техника машинског учења. На почетку је описано подручје истраживања над којим су вршена експериментална истраживања. У наставку су наведени коришћени скупови података, а затим је описана њихова припрема и иницијална анализа. Последњи сегмент овог поглавља описује софтверску подршку, односно најзначајније библиотеке и програме на које се ослања предложена методологија.

Највећи део **поглавља 5** чини опис процедуре, резултата и дискусије експерименталног одређивања оптималних параметара предложене методологије. Ово је реализовано кроз четири експеримента, којима су покривени и анализирани главни аспекти предложене методологије. У засебном делу су утврђени и коментарисани оптимални параметри предложене методологије. Они су затим коришћени за креирање коначних карата шумске вегетације на комплетној територији Републике Србије, што је последњи сегмент овог поглавља.

Последње **поглавље 6** приказује сажетак најважнијих закључака ове дисертације, уз препоруке за будућа истраживања.

2

ИЗДВАЈАЊЕ ИНФОРМАЦИЈА О ШУМАМА НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ

Друго поглавље дисертације обухвата литературни преглед могућности и актуелних трендова и праваца истраживања за издвајање информација о шумама на основу података сателитског осматрања Земље. Представљени су и анализирани сателитски системи за осматрање Земље, где су посебно разматрани пасивни оптички сателитски системи и активни радарски сателитских системи. Најважније карактеристике сваке групе сензора и њихови најзначајнији представници су наведени, при чему су разматране и активне и историјске сателитске мисије. Остатак поглавља се односи на могућности употребе података сателитског осматрања Земље за одређивање обухвата и типова шума, као и за одређивање биофизичких карактеристика шума. Извршена је синтеза достигнућа и актуелних трендова у овим областима истраживања, уз велики број референци на истраживања новијег датума која користе отворене и јавно доступне сателитске податке.

2.1 Сателити за осматрање Земље

Епоха сателитског осматрања Земље започиње 4. октобра 1957. год. када је Совјетски Савез лансирао први вештачки сателит Спутњик-1. Од тада се број сателита у орбити експоненцијално увећава, тако да данас постоји више од 2000 активних сателита који врше осматрање Земље (*Ferreira* и др., 2020). Различити материјали другачије

рефлектују и апсорбују зрачење у различитим деловима електромагнетног (ЕМ) спектра, што представља основу за њихово међусобно разликовање (нпр. одређивање типа земљишног покривача) (*Artiola* и др., 2004; *Ma* и др., 2017). Сателити за осматрање Земље су опремљени различитим сензорима који се разврставају на активне и пасивне. Активни сензори емитују зрачење ка Земљиној површи и затим региструју повратно зрачење које се рефлектовало или расејало. Супротно њима, пасивни сензори не емитују сопствено зрачење, већ региструју зрачење које Земљина површина емитује као сопствену енергију или коју рефлектује од Сунца (*Andries* и др., 2019).

Податке прикупљене сателитским сензорима карактеришу четири различите резолуције: просторна, спектрална, радиометријска и временска. Просторна резолуција дефинише величину површи терена (величину странице квадратног елемента површи терена) коју у природи покрива сваки пиксел сателитског снимка и практично дефинише минималне објекте које ће бити могуће разлучивати на снимцима. Спектрална резолуција се односи могућност разликовања појединих опсега таласних дужина у електромагнетном спектру у којима сателитски сензор прикупља зрачење, док радиометријска резолуција представља детаљност (финоћу) у нивоима зрачења које сензор може да разликује. Последња резолуција је временска и односи се на период до поновног прелета, тј. фреквенцију којом сателитски сензор поново осмотри исто подручје Земље. Уколико се као критеријум разврставања сателитских сензора посматра спектрална резолуција, треба разликовати: оптичке/инфрацрвене системе, термалне инфрацрвене системе и радарске системе (*Kuenzer* и др., 2014). Сателитски сензори се категоризују и на основу просторне резолуције. Тако се сензори могу разврстати на оне грубе просторне резолуције ($> 1 \text{ km}$), средње просторне резолуције (100 m – 1 km), високе просторне резолуције (5 m – 100 m) и веома високе просторне резолуције ($< 5 \text{ m}$) (*Liang* и *Wang*, 2020).

У наставку ће укратко бити приказане две главне групе сензора и наведени и описани њихови најзначајнији представници. Као допунски извор се препоручује веома опширна и детаљна анализа сателитских сензора коју су урадили *Kuenzer* и др. (2014). Посебна предност истраживања ове групе аутора је што је представљена применљивост сателитских сензора за праћење стања биодиверзитета, где је велика пажња поклоњена управо шумским стаништима и вегетацији.

2.1.1 Пасивни оптички сателитски системи

Пасивни оптички сателитски системи региструју зрачење у оквиру видљивог (400 nm – 700 nm), блиског инфрацрвеног (700 nm – 1000 nm), краткоталасног инфрацрвеног (1000 nm – 2500 nm), па све до дуготаласног инфрацрвеног (8000 nm – 14000 nm) дела ЕМ спектра. Сензори неких сателитских мисија, поред оптичког/инфрацрвеног дела ЕМ спектра, региструју зрачење и у термалним инфрацрвеним (енгл. *Thermal InfraRed - TIR*) каналима. *TIR* спектрални опсег подразумева подручје таласне дужине између 3000 nm и 14000 nm, али сензори због атмосферске продорности зрачење најчешће региструју у опсезима 3000 nm – 5000 nm и 8000 nm – 14000 nm (*Jensen, 2006*). Број спектралних опсега (канала) у којима се региструје зрачење разликује се од сензора до сензора. Мултиспектрални скенери (МСС) се конструишу са неколико (најчешће несуседних) спектралних канала, док хиперспектрални скенери региструју зрачење у оквиру одабраног дела ЕМ спектра издељеног на стотине уских спектралних канала (*Kivinen и др., 2020*). Сателитски сензори који поседују *TIR* канале омогућавају одређивање термалне емисивности (тј. температуре зрачења) Земљине површине, а поред емисивности се често креирају и додатни продукти, као што су: температура површине земљишта (енгл. *Land Surface Temperature - LST*), термална инерција, термалне аномалије, жаришта итд. (*Liang и Wang, 2020; Jensen, 2006*).

Карактеристике пасивних оптичких система драстично варирају по питању просторне, спектралне и временске резолуције. Због неопходних технолошких компромиса, просторна резолуција пасивних оптичких система диктира ширину подручја обухвата сензора и на тај начин утиче и на временску резолуцију (*Fagan и DeFries, 2009*). Неке мисије се пројектују са посебним путањама (геостационарни сателити) које обезбеђују кратак период обрта, али само над одређеним локалним/регионалним подручјима (*Liang и Wang, 2020*). Често решење за претходни проблем јесте усмеравање обухвата снимања сензора на специфичне делове Земље или формирање констелације већег броја сателита која се састоји од једног пара сателита, више десетина, па чак и стотина сателита (*Liang и Wang, 2020*). Додатно, због опсега коришћених таласних дужина и пасивног приступа, ови системи немају продорност кроз облаке или дим, погођени су измаглицом и не могу бити коришћени ноћу. Ове особености посебно треба имати у виду уколико се планира прикупљање информација у подручјима веома изражене и честе облачности (нпр. у околини екватора).

У наредној табели (**Табела 1.**) приказане су најважније пасивне оптичке сателитске мисије, односно сензори. Као главни критеријум избора је разматрана тренутна и будућа доступност података, али и значај и присутност мисија/сензора у пракси.

Табела 1. Приказ пасивних оптичких сателитских мисија/сензора са најважнијим карактеристикама (ПР – просторна резолуција, ВР – временска резолуција)

Сателитска мисија/сензор	Број канала	ПР [m]	ВР [дана]	Обухват [km]	Отворен приступ	Период доступности
AVHRR *	4-6	1100	0.5	2399	Да	1979-
VIIRS *	22	375-750	1	3060	Да	2011-
MODIS *	36	250-1000	0.25-0.50	2330	Да	2000-
Sentinel-3 *	30	300-1000	1-2	1440	Да	2016-
FIREBIRD *	5	42.4-356	3-4	178-211	Не	2012-
Landsat *	4-11	15-100	16	185	Да	1972-
ASTER *	14	15-90	16	60	Да	2000-
Sentinel-2	13	10-60	5	290	Да	2015-
SPOT	4-5	2.5-20	26	120	Не	1986-
RapidEye	5	6.5	1-5.5	77	Не	2008-2020
ZiYuan III	4	2.1-5.8	5	50	Не	2012-
PlanetScope	4	3.7	1	24	Не	2016-
IKONOS	5	1-4	1.5-3	11.3	Не	1999-2015
QuickBird	5	0.61-2.24	2.7	16.5	Не	2001-
Worldview	4-17	0.31-2.40	1-4	17.6	Не	2007-
SkySat	5	0.9-2.0	0.15	8	Не	2013-
Pleiades	5	0.5-2	4-5	20	Не	2011-
GeoEye	5	0.46-1.84	2.1-8.3	15.2	Не	2008-

* поседују термалне канале

Ипак, без обзира на наведена ограничења, пасивни оптички системи омогућавају осматрање Земљине површине од глобалног до локалног нивоа (*Kivinen* и др., 2020), и то са вишедеценијским временским обухватом. Радиометар AVHRR (скраћено од *Advanced Very-High-Resolution Radiometer*) опремљен МСС-ом обезбеђује глобално дневно (а за велики део Земље и два пута дневно) осматрање земљине површине од 1979. године (*Hastings* и *Emery*, 1992). Међутим, његова примена је ограничена због грубе просторне резолуције од 1 km. Радиометри MODIS (скраћено од *Moderate*

Resolution Imaging Spectroradiometer) на *Terra* и *Aqua* сателитима обезбеђују податке осматрања Земље и до 4 пута дневно (временска резолуција *Terra* и *Aqua* сателита је 1-2 дана). Мисија је активна од 2000. године и то са просторном резолуцијом од 250 m до 1 km у зависности од спектралног канала (*Justice* и др., 1998). Једино флота *Landsat* сателита обезбеђује податке високе просторне резолуције (у просеку 30 m, па све до 10 m за панхроматске канале) за период од преко 30 година (*Straub* и др., 2019). *Landsat-7*, *Landsat-8* и *Landsat-9* су тренутно активни сателити ове мисије чији су снимци доступни. Временска резолуција *Landsat* сателита је 16 дана. *Sentinel-2* сателитска мисија је део Коперникус програма који је развила Европска свемирска агенција (енгл. *European Space Agency - ESA*). Констелацију од 2015. године чине два идентична сателита опремљена МСС-ом високе просторне резолуције који прикупљају податке у 13 спектралних канала са комбинованом временском резолуцијом од 5 дана (*Drusch* и др., 2012). Као што је раније већ напоменуто, немају све мисије и *TIR* канале. Уколико постоје, њихов број и спектрални опсег варира, где нпр. *Landsat-8* и *Landsat-9* имају по два, а *ASTER* чак пет термалних канала. Међутим, уколико је неопходно обезбедити вишедеценијски обухват података, једино су *AVHRR* и *Landsat* мисије које обезбеђују термалне податке дуже од три деценије у континуитету.

Посебну групу пасивних оптичких система представљају мисије са веома високом просторном резолуцијом, односно чији је пиксел мањи од 1 m. Подаци оваквих система су за сада једино доступни у комерцијалним аранжманима. Најважнији представници ове групе су *QuickBird*, *Worldview*, *RapidEye*, *ZiYuan III* и *GaoFen* сателитске мисије. Међутим, поред бројних погодности које овако висока просторна резолуција омогућава, треба имати на уму да ови сензори често немају инфрацрвене, а још ређе термалне канале, који су посебно значајни код сателитског опажања вегетације (*Kuenzer* и др., 2014).

2.1.2 Активни радарски сателитски системи

Активни радарски сателитски системи, најчешће радари са синтетичким отвором (енгл. *Synthetic Aperture Radar - SAR*), емитују зрачење ка Земљиној површини у микроталасном делу ЕМ спектра и затим мере повратно зрачење (енгл. *backscatter*). У питању су системи који могу да прикупљају податке без обзира на временске услове или доба дана (*Liang* и *Wang*, 2020). У употреби је већи број микроталасних опсега ЕМ

спектра, при чему централне таласне дужине ових опсега варирају од сензора до сензора. Генерално, SAR системи се конструишу тако да емитују и региструју ЕМ зрачење у следећим опсезима: X-каналу (~3 cm), C-каналу (~6 cm), L-каналу (~25 cm) или P-каналу (~70 cm) (Jensen, 2006; Liang и Wang, 2020). Разлике у таласним дужина практично дефинишу који ће објекти на површи терена доминантно утицати на одбијање емитованог зрачења. Тако се нпр. краће таласне дужине више одбијају од мањих објеката попут грања и крупнијег лишћа дрвећа, док се веће таласне дужине више одбијају од великих објеката попут стабала (Boyd и Danson, 2005).

SAR системи емитују поларизовано и кохерентно ЕМ зрачење и затим региструју амплитуду и фазу повратног зрачења. Највећи број радарских система снима са бока (бочни SAR системи), па се због косе геометрије и топографије јављају проблеми издужења/скраћења и сенке. Регистрована фаза и амплитуда зависе од растојања објекта од сензора, а амплитуда додатно и од карактеристика таласа (поларизације, таласне дужине, упадног угла), али и карактеристика површи терена (храпавости, влажности, топографије, вегетације, вештачких објеката итд.) (Jensen, 2006; Liang и Wang, 2020). Напреднији SAR системи могу да емитују и региструју ЕМ зрачење у више поларизација (хоризонталној и вертикалној), што омогућава квалитетније сагледавање храпавости и текстуре површине. Због кохерентног зрачења, јавља се и ефекат зрнастог шума (енгл. *speckle noise*), који захтева посебно просторно или временско филтрирање (енгл. *speckle filter*). Просторно филтрирање подразумева неки вид осредњавања суседних пиксела, што практично значи да долази до деградације просторне резолуције (Ananto и др., 2019; Yuan и др., 2018).

SAR системи су, као и оптички системи, присутни у оквиру сателитских мисија већ неколико деценија, са различитим просторним, спектралним и временским резолуцијама. У наредној табели (**Табела 2.**) приказани су најзначајнији SAR системи. У табели су углавном наведене активне сателитске мисије. Ипак, поред њих су наведене и неке мисије које нису више активне, али које су се показале као значајне или су усаглашене са неком од тренутно активних.

Табела 2. Списак SAR система са најважнијим карактеристикама (ПР – просторна резолуција, ВР – временска резолуција)

Сателитска мисија/сензор	Канал	Поларизација	ПР [m]	ВР [дана]	Обухват [km]	Отворен приступ	Период доступности	
ERS-1/2	C	VV	20	35	10-100	Да	1991-2011	
JERS	L	HH	20	44	175	Да	1992-1998	
RADARSAT	C	HH/HV/VV **	1-100	5-10	18 - 500	Не	1995-	
TerraSAR-X и TanDEM-X	X	HH/HV/VV *	1-20	11	5-100	Не	2007-	
COSMO-SkyMed	X	HH/HV/VV *	1-100	0.5	10-200	Не	2007-	
SAOCOM	L	HH/HV/VV **	7-100	10-100	30 - 350	Не	2015-	
Sentinel-1	C	HH/HV/VV *	5-100	12	80-400	Да	2015-	
ALOS-2	L	HH/HV/VV **	3-100	14	25-350	Не	2015-	
ICEYE	X	VV	1-3	18-22	5-30	Не	2018-	
PAZ	X	HH/HV/VV **	0.25-40	11	4-200	Не	2018-	
Biomass	P	тек треба да буде објављено						2022-

* прикупљају зрачење у једној или двама поларизацијама

** прикупљају зрачење у свим или било којој поларизацији (полариметрични)

2.2 Могућности сателитског осматрања Земље за прикупљање информација о шумама

Традиционалне теренске методе обезбеђују веома квалитетне информације о шумама, али често по цену великог утршка времена и много ручног рада. Коришћење података сателитског осматрања Земље може да надомести ове недостатке смањењем потребне количине времена и рада, уз очување тачности и богатства информација (Grabska и др., 2019; Madonsela и др., 2018; Sheeren и др., 2016). Сателитска опажања Земље често могу пружити и информације које ће индиректно описивати шумске карактеристике или пружити допунске информације за њихово успешније разумевање. Комбиновањем спектралних канала могуће је креирање различитих вегетационих индекса, који су се показали као добар индиректан начин праћења земљине површине (Хуе и Су, 2017; Huete, 2012; Tucker, 1979). Сателитски подаци практично немају алтернативу када је реч о глобалној покривености, нудећи податке и од неколико деценија уназад и најчешће са великим распоном просторне, спектралне и временске резолуције. Цена сателитског

осматрања је веома мала у односу на друге методе прикупљања података, уколико се цена изражава према јединици површине (*Kuenzer* и др., 2014; *Mathieu* и *Aubrecht*, 2018; *Straub* и др., 2019). Додатно, на овај начин је омогућено комплетно и у великој мери хомогено сагледавање подручја од интереса, јер сателитски сензори врше систематско узорковање података на растојању дефинисаном просторном резолуцијом пиксела. Супротно томе, теренске методе се из практичних и економских разлога ослањају на систематско узорковање са знатно већим растојањем између узорака, при чему се и процедуре узорковања разликују од државе до државе (*Lechner* и др., 2020; *Pause* и др., 2016).

Већ је раније наведено да су сателитске мисије и њихови сензори конструисани са одређеним ограничењима и уз компромисе по питању резолуција. Не постоји идеалан свемогућ сензор за све примене, већ избор одговарајућег сателитског сензора пре свега мора бити прилагођен циљу и потребама за који ће се користити (*Kuenzer* и др., 2014; *Lechner* и др., 2020). Овакав приступ свакако важи и када је реч о шумској вегетацији. Величина подручја које се истражује често је главни фактор који утиче на избор сателитске мисије и сензора. Поред тога, веома су важни и практични аспекти као што су расположиви буџет, могућности обраде и дефинисани рокови (*Boyd* и *Danson*, 2005; *Iverson* и др., 1989). Последично, подаци грубе просторне резолуције су често добар избор када се ради о великим подручјима истраживања. Финија просторна резолуција доноси богатство детаљности, али подразумева и веће трошкове набавке, веће меморијске и процесорске захтеве и најчешће нижу временску резолуцију (осим у случају констелације већег броја сателита) (*Liang* и *Wang*, 2020; *Zhang* и др., 2020). Све се то може показати као пресудни фактор, при чему је за пасивне оптичке системе облачност додатни проблем, који је нарочито изражен у тропским подручјима. У том случају, избор се сужава на радарске системе или пасивне оптичке системе грубе просторне резолуције. Насупрот коришћењу једног сензора, синергијски ефекат комбинације више различитих сензора је више пута доказиван (*Boyd* и *Danson*, 2005; *Franklin* и *Blodgett*, 1993; *Zhu* и др., 2018). Посебно је атрактивно комбиновање активних радарских SAR података са пасивним оптичким опажањима, где се поред спектралне синергије решава и проблем заклоњености услед облачности (*Gargiulo* и др., 2020; *Joshi* и др., 2016; *Palubinskas* и др., 2011).

Различити истраживачи на другачије начине категоришу примене података сателитског осматрања Земље за одређивање информација о шумама. У оквиру овог истраживања усвојена је подела према *Boyd и Danson (2005)*. У складу са таквом поделом, у наредним поглављима су приказане могућности коришћења сателитског осматрања Земље за одређивање обухвата и типова шума (**поглавље 2.2.1**), као и за одређивање биофизичких карактеристика шума (**поглавље 2.2.2**).

2.2.1 Одређивање обухвата и типова шума

Одређивање обухвата шума има за циљ картирање подручја Земљине површине која покривају шуме. Проблем може бити посматран и тако да се врши диференцирање шумског од нешумског подручја (*Ganz и др., 2020; Martone и др., 2018; Pekkarinen и др., 2009; Waser и др., 2015*). Обухват шума такође може бити одређен и индиректно, где је примарни циљ картирање типова шума или типова вегетације, а до конкретне границе обухвата шума се долази агрегацијом издвојених шумских класа (*Tsendbazar и др., 2020*). Информације о обухвату и типовима шума су се показале као важне ради прецизније оцене продуктивности шуме и код одређивања запремине дрвне масе и базалне површине (*Lutz и др., 2008; Xiao и др., 2019*). Одређивање обухвата шума или типова шума се често врши и као део одређивања карте земљишног покривача или карте коришћења земљишта, где су шуме или више типова шума само неке од циљних тематских класа (*Nguyen Trong и др., 2020*).

Типови шума могу бити различито дефинисани, од грубих класа као што су листопадне, зимзелене и мешовите шуме, до знатно детаљнијих класификационих шема изграђених до нивоа доминантних врста дрвећа (*Boyd и Danson, 2005*). Који типови шума ће бити издвојени пре свега зависи од циљне примене, али и од комплексности подручја истраживања и расположивих података. Није неубичајено ни усвајање класификационе шеме са класама различитог нивоа детаљности, где су заступљене и грубе и детаљније шумске класе. Ипак, детаљно разликовање типова шума и даље представља комплексан проблем чије решавање методама даљинске детекције заснованим на сателитском осматрању Земље има ограничени домет. Разлог треба тражити пре свега у великој хетерогености шумских подручја која се манифестује кроз веома комплексне и преклапајуће спектралне карактеристике, посебно изражене у тропским подручјима (*Grabska и др., 2019*). На разноликост спектралних

карактеристика, поред структуре и карактеристика различитих типова вегетације, утичу и разлике у условима раста, густина и старост вегетације. У планинским подручјима додатно су изражени и топографски ефекти и утицај висинске промене. Како се подручје истраживања повећава, ови ефекти су присутнији и отежавају разликовање типова шума (*Lutz* и др., 2008).

Ране студије одређивања обухвата и типова шума су се ослањале на визуелну интерпретацију садржаја сателитских снимака (*Asrat* и др., 2018; *Lechner* и др., 2020; *Schepaschenko* и др., 2019). Интерпретацију су по правилу вршили експерти, користећи знање о подручју истраживања у комбинацији са спектралним и просторним обрасцима уочљивим на снимцима. Повећањем просторне резолуције снимака, на располагању су и информације о текстури ради квалитетнијег разликовања шумског од нешумског подручја и/или различитих типова шума (*Schepaschenko* и др., 2019). Предност овог приступа је што оператер лако превазилази проблеме различитог угла осматрања сателита, временских услова и евентуалних шума на снимцима. Међутим, овај приступ је истовремено и веома захтеван, подразумева велику количину мануелног рада и подложен је субјективности. Због дуготрајног периода израде, посебно када се ради о већим подручјима, неретко се дешава да овако креиране карте већ у току израде застаревају, па је њихова употребљивост за поједине примене ограничена. Зато се визуелна интерпретацију данас преваходно користи у сврху прелиминарне интерпретације садржаја сателитских снимака и у комбинацији са другим методама (*Boyd* и *Danson*, 2005; *Schepaschenko* и др., 2019).

Технике класификације сателитских снимака су доминантне када је у питању одређивање обухвата и типова шума. То потврђује велики број успешних примена, које користе широк дијапазон сателитских сензора различитих карактеристика (*Cheng* и *Wang*, 2019; *Martone* и др., 2018; *Sheeren* и др., 2016). Просторна резолуција сензора и усвојена класификациона шема су уско повезани, па се тако подаци грубљих просторних резолуција уобичајено користе за одређивање обухвата шума и/или типова шума на грубљим нивоима (*Clerici* и др., 2012; *Nelson* и *Horning*, 1993; *Xin Fang Yui* и др., 2004). Подаци грубље просторне резолуције су природно погоднији уколико се одређивање обухвата и типова шума спроводи на већим подручјима (националном, регионалном, континенталном или глобалном нивоу) (*Lechner* и др., 2020). У таквим случајевима је посебно изражен проблем помешаних пиксела, који умногоме може

утицати на погрешну (умањену) процену површина (Boyd и Danson, 2005). Потенцијална решења у таквим случајевима подразумевају увођење корекције одређених површина на основу процене обухвата спроведене на локалним подручјима користећи податке у вишој просторној резолуцији (Tomppo и др., 2008). Други приступ подразумева увођење расплинуте логике (енгл. *fuzzy logic*), где се као резултат одређује проценат заступљености шума или типова шума за сваки пиксел. На оба начина се могу обезбедити квалитетније и тачније оцене површина под шумама, али ће тачност одређених граница обухвата и типова шума и даље бити ограничена просторном резолуцијом сателитских података.

Повећањем просторне резолуције омогућава се финије диференцирање типова шума. Издвајање типова шума на нивоу врста дрвећа представља виши ниво разликовања. Међутим, ово издвајање је за сада и даље ограниченог домета због веома велике спектралне блискости појединих врста. Због тога се оваква истраживања уобичајено фокусирају на локална подручја и то са мањим бројем циљних врста дрвећа (Ferreira и др., 2019; Immitzer и др., 2012; Kim и др., 2011). Ипак, поред просторне резолуције, пресудни фактор за разликовање типова шума је спектрална резолуција сензора, тј. број расположивих спектралних канала у којима се прикупљају подаци. У случају мултиспектралног опажања ту се пре свега мисли на канале изван видљивог дела ЕМ спектра, као што су блиски инфрацрвени (енгл. *Near-InfraRed* – *NIR*) и краткоталасни инфрацрвени канали (енгл. *Short-Wavelength InfraRed* – *SWIR*) у којима спектрални одзиви вегетације зависе од њене структуре и количине воде (Huete, 2004). Хиперспектрални сензори генерално омогућавају успешније разликовање типова шума у односу на мултиспектралне сензоре (Boyd и Danson, 2005; Lutz и др., 2008). Међутим, за сада је доступан тек незнатан број оваквих сателитских сензора са грубљом просторном и временском резолуцијом и често само у комерцијалним аранжманима. Коришћење радарских опажања у комбинацији са изведеним радарским карактеристикама се такође показало као успешан приступ за разликовање различитих типова шума (Bae и др., 2019; Mazza и др., 2019; Pulella и др., 2020; Rignot и др., 1994). Додатно, коришћење радарских опажања у различитим поларизацијама се показало као корисније у односу на само једну поларизацију (Boyd и Danson, 2005; Ruiz-Ramos и др., 2020).

Укључивање помоћних извора података у процедуру класификације је такође постала уобичајена пракса. Укључивање информације о текстури (посебно код сателитских података више просторне резолуције) значајно доприноси поузданијем разликовању типова шума (*Ferreira* и др., 2019). Подаци висинске представе терена и изведени производи из њих су се исто тако показали као значајни (*Borlaf-Mena* и др., 2020; *Liu* и др., 2018). Последњих година је све израженији тренд коришћења више извора података приликом одређивања обухвата и/или типова шума. Овакав приступ се назива фузијом података. Под тиме се подразумева, како комбинација опажања сензора из исте групе, тако и упаривање радарских и оптичких опажања (*Liu* и др., 2018). Истраживања показују и да коришћење серије сателитских снимака распоређених у различитим периодима године погодује издвајању типова шума. На овај начин се постиже обухватање фенолошких вегетационих циклуса, што наглашава различитост спектралних карактеристика појединих шумских класа (*Lutz* и др., 2008; *Wulder* и др., 2020). Ипак, обухватање специфичних фенолошких догађаја (нпр. почетак сезоне) може бити отежано услед заклоњености облацима. Сем тога, ово прате и већи меморијски, процесорски и новчани захтеви (ако снимци нису бесплатни), уз обавезно коришћење снимака вишег нивоа обраде - са спроведеним геометријским, радиометријским и атмосферском корекцијама (*Kovačević* и др., 2020).

2.2.2 Издајање биофизичких шумских карактеристика

Традиционалне теренске методе одређивања биофизичких и биохемијских карактеристика шума су веома захтевне по питању времена и људских ресурса, а прикупљање података за велика подручја и у већем броју епоха је практично неизводљиво (*Boyd* и *Danson*, 2005; *Lechner* и др., 2020; *Lutz* и др., 2008). Ипак, ради се о веома важним карактеристикама шума које варирају просторно и временски и које су од великог значаја за бројне примене. Селективна сеча и проређивање шума која је нарочито изражена у Африци утиче на структуру шуме, коју је могуће идентификовати и квантификовати помоћу биофизичких параметара (*Gourlet-Fleury* и др., 2013; *Hirschmugl* и др., 2020; *Laporte* и *Lin*, 2003). Индекс површине листа заједно са биохемијском структуром листа карактерише функције шуме као што су интеракције са зрачењем, процес кружења материје и продуктивност шуме (*Zheng* и *Moskal*, 2009). Покровност шумских крошњи дефинише микроклиматске услове који владају у ивичним шумским подручјима (*Pinto* и др., 2010), односно утиче на ширину ових

подручја и може послужити за праћење стања биодиверзитета и стабилности предела (Li и др., 2018; Matlack, 1994). Карактеристике шуме као што су старост, висина стабала, базална површина стабла и количина дрвне масе могу бити изведене из биофизичких параметара, а представљају информације које су неопходне и присутне у националним инвентурама шумских ресурса (Boyd и Danson, 2005; Tomppo и др., 2010, 2008). Због свега тога је јасно зашто је значајан напор усмерен ка развоју метода даљинске детекције којима би се омогућило ефикасно и прецизно одређивање управо ових карактеристика на великим подручјима које покривају шуме.

Генерално, постоје два основна приступа одређивању биофизичких и биохемијских карактеристика шума на основу података сателитског осматрања Земље. Физичко моделирање подразумева математичко симулирање процеса интеракције зрачења са вегетацијом, користећи знања о основним факторима који утичу на везу између регистрованог зрачења сателитских сензора и биофизичких и биохемијских својстава вегетације. Ипак, и даље је јако тешко квалитетно моделирати факторе који утичу на спектралне карактеристике вегетације, попут хетерогености вегетационог покривача и спољашњих услова (атмосферско расејање и апсорпција) (Boyd и Danson, 2005; Zheng и Moskal, 2009). Компликације прави и велика временска и просторна варијација ових фактора, што у великој мери ограничава успешност физичког моделирања (Liang и Wang, 2020). Због тога је веома популаран и често операционо применљивији приступ заснован на емпиријском моделирању. У том случају се биофизички и биохемијски параметри шума доводе у везу са сателитским опажањима Земље (и изведеним карактеристикама попут спектралних индекса). Ово се најчешће изводи креирањем регресионих модела на основу кореспондентних теренских и сателитских података (Boyd и Danson, 2005). Обезбеђивање квалитетних теренских података је императив за примену поменутог приступа, што често може бити отежавајући фактор. Додатно, емпиријски модели подразумевају коришћење радиометријски и атмосферски ректификованих сателитских података, како би се омогућила њихова примена и на другим просторним и временским резолуцијама. Радиометријска и атмосферска корекција ипак често не обезбеђује довољну робусност модела због локалних специфичности и разноликости вегетације, па је у том случају неопходно обезбеђивање и додатних локалних теренских података, у циљу постизања врхунских резултата (Boyd и Danson, 2005; Lutz и др., 2008; Zheng и Moskal, 2009).

2.2.2.1 Индекс површине листа

Индекс површине листа (енгл. *Leaf Area Index - LAI*) је бездимензиона величина која представља однос површине листа према јединици површине терена. Као таква, може бити доведена у везу са процесима кружења материје у биљкама као што су фотосинтеза (*Duchemin* и др., 2006), испаравање и транспирација (*Chen* и др., 2005; *Cleugh* и др., 2007), упијање падавина (*Anderson* и др., 2015; *Li* и др., 2017; *Tesemma* и др., 2015) и циклус угљеника (*Soegaard*, 1999). *LAI* је такође један од кључних улазних параметара за одређивање надземне биомасе шумских подручја (*Boyd* и *Danson*, 2005; *Lutz* и др., 2008), али се показао и као важан индикатор стреса шуме (*Anderson* и др., 2015) и метода за разумевање динамике промена продуктивности и климатских утицаја на шумске екосистеме (*Zheng* и др., 2019).

Уочена је значајна позитивна корелација између *LAI*-ја и следећих индекса: индекса једноставног односа (енгл. *Simple Ratio Index - SR*), вегетационог индекса нормализоване разлике (енгл. *Normalized Difference Vegetation Index - NDVI*), унапређеног вегетационог индекса (енгл. *Enhanced Vegetation Index - EVI*) и индекса умањеног једноставног односа (енгл. *Reduced Simple Ratio - RSR*) (*Boyd* и *Danson*, 2005; *Lutz* и др., 2008; *Zheng* и *Moskal*, 2009). Вредности *LAI*-ја су такође довођене у везу са расположивим спектралним каналима и бројним другим вегетационим индексима коришћењем различитих линеарних и нелинеарних модела. Ипак, истраживања показују да „јачина“ ових веза зависи од композиције биљних врста и карактеристика подручја истраживања, као и да постоји велики утицај рефлексије зрачења од земљишта и атмосферских услова (*Zheng* и *Moskal*, 2009). Ово се обично манифестује кроз засићење вредности *LAI*-ја, где се односи мењају нелинеарно, посебно код већих вредности *LAI*-ја. Додатни проблем коришћења вегетационих индекса као посредника је што они зависе од сателитског сензора, првенствено од његових спектралних карактеристика и угла осматрања, што захтева калибрацију модела независно за различите сензоре (*Boyd* и *Danson*, 2005).

Хиперспектрални сензори генерално омогућавају тачније моделирање *LAI*-ја (*Boyd* и *Danson*, 2005; *Lutz* и др., 2008; *Zheng* и *Moskal*, 2009). Због великог броја уских канала, ови сензори омогућавају одређивање првих или других извода спектралне рефлектансе, који су мање осетљиви на утицаје густине вегетације и омогућавају прецизнију оцену *LAI*-ја (*Liu* и др., 2016; *Xiao-chen* и др., 2008). Хиперспектрални

подаци се могу искористити и за одређивање локације тзв. „црвене ивице“ (енгл. *red edge*), места нагле промене вредности рефлектансе у црвеном делу ЕМ спектра која се јавља око таласне дужине 720 nm. Истраживања су показала да се локација „црвене ивице“ може успешно довести у везу са вредношћу LAI-ја и да је ова повезаност снажнија него када је реч о мултиспектралним каналима или на основу њих изведених спектралних индекса (*Darvishzadeh* и др., 2009; *Xie* и др., 2018). Истовремено, постоји велико интересовање за примене радарских опажања (и то најпре SAR опажања) ради оцене LAI-ја. Истраживања су доминантно усмерена ка интеграцији оптичких и радарских опажања (*Beerli* и др., 2020; *Pipia* и др., 2019), али има и примера коришћења само радарских опажања (*Bériaux* и др., 2015; *Manninen* и др., 2003; *Stankevich* и др., 2017).

Прегледни рад који су урадили *Zheng* и *Moskal* (2009) представља сјајан извор информација које се односе на одређивање LAI-ја. У овом раду је обухваћен детаљан и свеобухватан приказ теорије, метода и сензора за одређивање LAI-ја, са посебним нагласком на могућности и проблеме који се односе на коришћење сателитских сензора (*Zheng* и *Moskal*, 2009).

2.2.2.2 Надземна биомаса

Надземна биомаса (енгл. *aboveground biomass*) подразумева количину суве материје која је присутна код живе вегетације изнад површи терена. Информација о количини надземне биомасе је пресудна за управљање шумским ресурсима, за разумевање и управљање глобалним циклусима угљеника, али и као улазни параметар за одређивање бруто примарне производње шуме, праћење промена шумског покривача, алокације горива и хранљивих састојака итд. (*Lutz* и др., 2008; *Xiao* и др., 2019).

Велики број истраживања се фокусирао на одређивање надземне биомасе применом пасивних оптичких сателитских опажања, тако што се количина надземне биомасе емпиријским моделима доводи у везу са одређеним спектралним каналима сензора или изведеним вегетационим индексима (*Askar* и др., 2018; *Koju* и др., 2017). Истраживања су показала да је за постизање веће тачности оцењених количина надземне биомасе неопходно укључивати и информације о текстури и допунске информације о шумском покривачу, попут индекса површине листа, структуре крошње и процента сенке, старости, висине и типова шумске вегетације (*Lutz* и др., 2008; *Xiao*

и др., 2019). Пасивни оптички сензори су, уопштено говорећи, оптерећени проблемом засићења, што проузрокује потцењене количине у случају велике надземне биомасе, односно прецењене количине у подручјима мале надземне биомасе (Xiao и др., 2019). Алтернативни приступ подразумева коришћење SAR опажања различитих спектралних карактеристика и поларизације (Berninger и др., 2018; Huang и др., 2018). Истраживања су показала скоро линеарну везу регистрованог повратног зрачења и надземне биомасе, али и даље уз ефекат засићења при већим количинама биомасе (Boyd и Danson, 2005; Lutz и др., 2008). Ефекат засићења зависи од фреквенције ЕМ зрачења и износи око 200 t/ha за P-канал, око 100 t/ha за L-канал, док је зрачење у C-каналу много мање употребљиво за оцену надземне биомасе (Xiao и др., 2019). Додатно, временски услови у комбинацији са структуром и композицијом шуме могу утицати на квалитет оцене биомасе. Количина биомасе се директно може одредити на основу густине стабала, запремине стабла (висине дрвећа и пречника стабла у висини груди) и композиције дрвних врста (Xiao и др., 2019). На основу тога, предложен је и велики број сателитских алометријских модела за оцену надземне биомасе (Aabeyir и др., 2020; Daba и Soromessa, 2019; Miah и др., 2020).

2.2.2.3 Бруто и нето примарна производња

Бруто примарна производња (енгл. *Gross Primary Production – GPP*) представља укупну количину угљеника који биљка угради у себе током године, без узимања у обзир количине утрошене на респирацију. Нето примарна производња (енгл. *Net Primary Production – NPP*) је мера брзине којом биљке одвајају угљеник и складиште енергију, али након што су и количине утрошене на респирацију узете у обзир. Познавање *GPP* и *NPP* указује на разлике у продуктивности екосистема и омогућава увид како различита подручја на Земљи апсорбују угљеник у различитим животним фазама (Yu и др., 2009; Zhang и др., 2016). Овиме се омогућава праћење испуњености стандарда дефинисаних Кјото протоколом и другим протоколима за контролу емисије угљеника (Lutz и др., 2008; Xiao и др., 2019). *GPP* и *NPP* су биофизичке карактеристике које се тешко одређују и за одређивање њихових прецизних вредности на већим подручјима неопходни су посебни торњеви за мерење флукса (Gao и др., 2013; Lutz и др., 2008). Због тога уклапање података сателитског опажања Земље са овим релативно ретким теренским подацима омогућава веома ефикасно и прецизно обезбеђивање бруто и нето примарне производње за већа подручја.

Једноставнији приступ за одређивање *GPP* и *NPP* подразумева калибрацију емпиријских модела заснованих на вегетационим индексима са теренски одређеним *GPP* и *NPP* вредностима. Истраживања су показала да је на овај начин могуће одредити *GPP* са релативно високом тачношћу (*Saleska* и др., 2016; *Xiao* и *Moody*, 2005). С друге стране, ови модели ипак имају потешкоће да моделују *GPP* у случају метеоролошки базираних несезонских екосистема (попут тропских шума) или екосистема код којих не постоји метеоролошка и фенолошка синхронизација (попут медитеранских екосистема) (*Xiao* и др., 2019). Други приступ одређивања *GPP* и *NPP* се ослања на моделирање ефикасности коришћења светлости (енгл. *Light Use Efficiency* – *LUE*). Код таквих приступа, подаци сателитског опажања Земље имају важну улогу у обезбеђивању информација о типовима вегетације, статусу раста вегетације и условима животне средине (*Xiao* и др., 2019). Ипак, смањивање несигурности оцене *GPP* и *NPP* коришћењем *LUE* модела и даље представља изазов, јер су ови модели осетљиви на временске услове, влажност земљишта, као и на ефекат засићења присутног код густе вегетације.

2.2.2.4 Карактеристике инвентуре шума

Иако се приликом инвентуре шума традиционално користе теренске методе прикупљања, за прикупљање неких карактеристика шума истовремено постоје и веће или мање могућности одређивања на основу података сателитског осматрања Земље. Потреба за проналажењем алтернативе теренским методама је очигледна, због чега је велики број истраживања усмерен ка унапређивању постојећих и развоју нових метода за прикупљање (одређивање) широког распона различитих параметара шума. У том смислу, у оквиру ове дисертације неће бити обрађене могућности сателитског осматрања Земље за одређивање свих карактеристика шума присутних у инвентурама шума, него ће само укратко бити покривене оне карактеристике које су најприсутније и које имају посебан значај.

Запремина стабла је најчешће коришћена информација која се одређује приликом инвентуре шумских ресурса и она представља запремину дрвне масе свих живих стабала на посматраном подручју (*Lutz* и др., 2008; *Tomppo* и др., 2008). За сада су донети одређивања ове карактеристике на основу сателитских опажања Земље крајње ограничени. Постоје појединачни покушаји издвајања на основу пасивних оптичких

снимака (*Gu* и др., 2006; *Mohammadi* и др., 2010; *Tanaka* и др., 2015) и радарских опажања (*Alappat* и др., 2011; *Chen* и др., 2020; *Vastaranta* и др., 2014), при чему је несигурност остварених резултата и даље превелика да би тако добијени подаци могли да се користе за управљање шумама.

Висина крошњи је важан индикатор бројних структурних карактеристика шума и може бити доведена у везу са другим шумским метрикама попут базалне површине, укупне биомасе (*Fassnacht* и др., 2021) и карактеристика биодиверзитета (*Bergen* и др., 2009; *Ghosh* и др., 2020). Оштећења шуме услед најезде инсеката или шумских пожара такође се могу пратити на основу висине крошњи (*Ranson* и др., 2003). SAR сензори су доминантни када је реч о одређивању висине крошње. Висина крошњи може бити одређена коришћењем принципа интерферометрије и кохеренције радарског зрачења, која је директно повезана са 3Д структуром вегетације (*Lutz* и др., 2008). За постизање високе тачности, радарска опажања је неопходно ректификовати за утицаје рељефа и временских услова као што су ветар, падавине и температура. Са друге стране, подаци пасивних оптичких сензора се не могу користити самостално, већ једино у комбинацији са SAR опажањима.

Одређивање старости шуме на основу сателитских опажања је такође ограниченог домета. Постоје засићења спектралних одзива код шума старости 30-40 година, што практично значи да није могуће прецизно одредити старост старијих шума (*Jensen* и др., 1999; *Lutz* и др., 2008). Истовремено, проблем прави и то што дрвеће на неком подручју због различитих микроуслова раста може бити различито без обзира на исту старосну доб. Проблем компликују и секундарне шумске врсте које у великој мери могу утицати на спектралне одзиве. Због тога, старост појединих типова шума је могуће веома квалитетно одредити, док постоје типови шума чију старост практично није могуће одредити на основу сателитског осматрања Земље (*Lutz* и др., 2008; *Zhang* и др., 2014). Ипак, услед велике важности одређивања старости шума за квалитетно управљање шумским ресурсима, за разумевање циклуса угљеника и за праћење напредовања шума након оштећења од инсеката, пожара или сече у шумама, и даље су значајни напори усмерени ка решавању овог проблема. Већина таквих приступа се фокусира на индиректно одређивање старости шуме, најчешће коришћењем вегетационих или индекса заснованих на текстури (*Lutz* и др., 2008).

Издвајање појединачних крошњи стабала је такође атрактивна тема истраживања која се истовремено може посматрати и као највиши и најдетаљнији ниво разликовања типова шума. Ове информације имају велике употребне вредности за креирање дигиталног катастра зелених површина или за планирање зелене инфраструктуре на локалној размери урбаних предела (*Matasci* и др., 2018; *Shackleton* и др., 2015; *Vasiljević* и др., 2016). За такве потребе се подразумева коришћење снимака веома високе просторне резолуције. Поред приступа базираног на визуелној интерпретацији садржаја која је оптерећена већ раније наведеним проблемима, веома су атрактивне и полуаутоматска и аутоматска детекција крошњи стабала. Ови приступи се ослањају на коришћење посебних алгоритама базираних на сегментацији сателитских снимака и најчешће су прилагођени за појединачне врсте дрвећа (*Hirschmugl* и др., 2007; *Tianyang* и др., 2018).

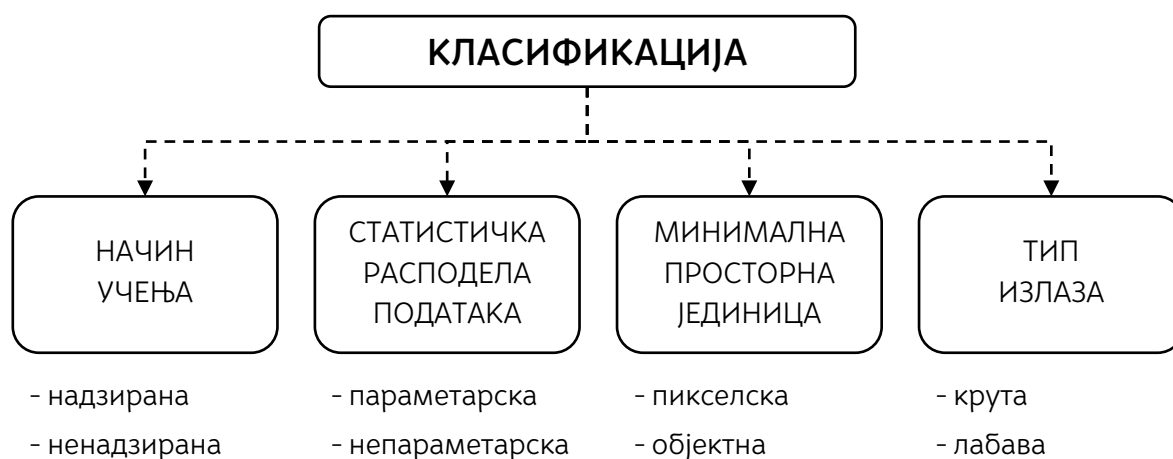
3

КЛАСИФИКАЦИЈА САТЕЛИТСКИХ СНИМАКА ТЕХНИКАМА МАШИНСКОГ УЧЕЊА

У оквиру трећег поглавља извршен је приказ теоријских основа процедуре класификације садржаја сателитских снимака коришћењем техника машинског учења. Нагласак је на надзираним непараметарским класификационим техникама машинског учења, а детаљније су приказане теоријске основе алгоритама стабла одлучивања (енгл. *decision tree*), случајне шуме (енгл. *random forest*) и градијентног појачавања (енгл. *gradient boosting*). Разматрана је процедура просторно-временске класификације, где је посебна пажња поклоњена могућностима и приступима за укључивање временске компоненте, као и поступцима за наглашавање просторних аспеката у класификационим моделима. Креирање и избор репрезентативних атрибута је такође један од анализираних аспеката класификационе процедуре. Такође је разматран проблем небалансираности (неравномерне заступљености) класа у подацима и његово превазилажење креирањем додатних примерака (инстанци) мањински заступљених класа. Поглавље се завршава описом процедуре за оцену квалитета извршене класификације. Приказане су усвојене квантитативне мере перформанси класификационог модела и описана је коришћена процедура крос-валидације са изостављањем комплетних локација.

3.1 Технике машинског учења

Класификација слике подразумева груписање пиксела који имају сличне обрасце дигиталних вредности у коначан број класа. Примењена на податке сателитског опажања Земље, класификација покушава да уочи везу између спектралних образаца сателитских опажања и присутних типова земљишног покривача. У случају картирања шумске вегетације, крајњи циљ класификације јесте израда карте са онолико типова (класа) шумске вегетације колико је то предвиђено усвојеном класификационом шемом. Методе класификације сателитских снимака се могу разврстати на више начина, и то према: 1) начину учења, 2) претпоставкама о дистрибуцији података, 3) минималној просторној јединици која се класификује и 4) типу излаза за сваку просторну јединицу (Слика 1.).



Слика 1. Разврставање метода класификације на основу различитих критеријума

Према начину учења треба разликовати надзиране и ненадзиране методе класификације. Надзиране методе класификације захтевају скуп примера за сваку циљну класу („податке за учење“ или „тренинг податке“), на основу којих алгоритам класификације формира скуп правила за разликовање класа. Обезбеђивање свеобухватног репрезентативног скупа примера за учење је предуслов за постизање врхунских резултата, што често може бити веома захтеван посао. Самим тим, перформансе надзираних метода могу да варирају од врло ниских до врхунских, према томе колико је скуп података за учење обухватнији и репрезентативнији. Насупрот њима, ненадзиране методе класификације не захтевају податке за учење, већ врше природно груписање пиксела/објеката на основу сличности њихових дигиталних вредности (спектралних карактеристика). Због тога, ненадзирана класификација

подразумева додатни корак у коме се врши интерпретација идентификованих природних класа и њихово мапирање у стварне класе.

Претпоставке о статистичкој расподели података категоришу методе класификације на параметарске и непараметарске методе. Параметарске методе се ослањају на теорију статистичког одлучивања која претпоставља одређену статистичку расподелу података. Ову расподелу података је могуће описати унапред дефинисаним коначним скупом параметара. Комплексност модела је тада унапред фиксирана, што овакве моделе чини мање флексибилним. Непараметарске методе су, с друге стране, ослобођене претпоставки о статистичкој расподели података. У њиховом случају се расподела података не може описати коначним скупом параметара, већ се најчешће скуп параметара посматра као функција која се мења према расположивом скупу податка. Перформансе параметарских метода класификације у великој мери зависе од тога колико подаци заиста подлежу усвојеним статистичким моделима, као и од квалитета оцене параметара модела. Поред тога, параметарске методе су посебно осетљиве на Хјузов феномен (енгл. *Hughes phenomenon*) или „проклетство димензионалности“, где долази до драматичног пада перформанси класификатора када број димензија (атрибута, спектралних канала) достигне одређену вредност (Ма и др., 2013).

Минимална просторна јединица која се обрађује класификационим алгоритмом може бити пиксел или група пиксела који чине неки објекат уочљив на слици. Објектно-оријентисана класификација је релативно новији приступ у односу на пикселски базиран. Овај приступ покушава да имитира људску визуелну интерпретацију. Потреба за увођењем објектно-оријентисане парадигме се јавља када појединачни пиксел више не носи довољно информација за квалитетно разврставање пиксела у неку од класа и где примена пикселски базираних алгоритама уобичајено резултује појавом „со и бибер шума“ (енгл. *salt-and-pepper noise*). Ово је карактеристичан проблем у раду са сателитским опажањима високе просторне резолуције. Објектно-оријентисана класификација уводи додатан корак, где се пре класификације врши сегментација слике, односно груписање пиксела у објекте према њиховој сличности. На овај начин се омогућава класификационом алгоритму да, поред дигиталних вредности (спектралних карактеристика) пиксела, искористи и додатне карактеристике које потичу од креираних објеката (облик, величина, текстура итд.).

Класификација може бити крута (енгл. *hard*) или лабава (енгл. *soft*), у зависности од тога како се изражава припадност класи сваке минималне просторне јединице слике. Крута класификација подразумева да ће свака просторна јединица припадати једној и само једној класи, док лабава класификација омогућава припадност више од једној класи и/или делимичну припадност класама. Лабава класификација је предложена у литератури као алтернатива крутој класификацији због њене способности да моделира помешане пикселе (енгл. *mixed pixels*). Проблем помешаних пиксела се јавља као последица теселације простора и просторне резолуције сензора, где се у резултујућим пикселима не јавља само једна циљна класа. Примена лабавог класификационог приступа омогућава изражавање степена припадности пиксела свакој класи, што може осликавати релативну пропорцију присутности различитих класа на подручју које тај пиксел покрива.

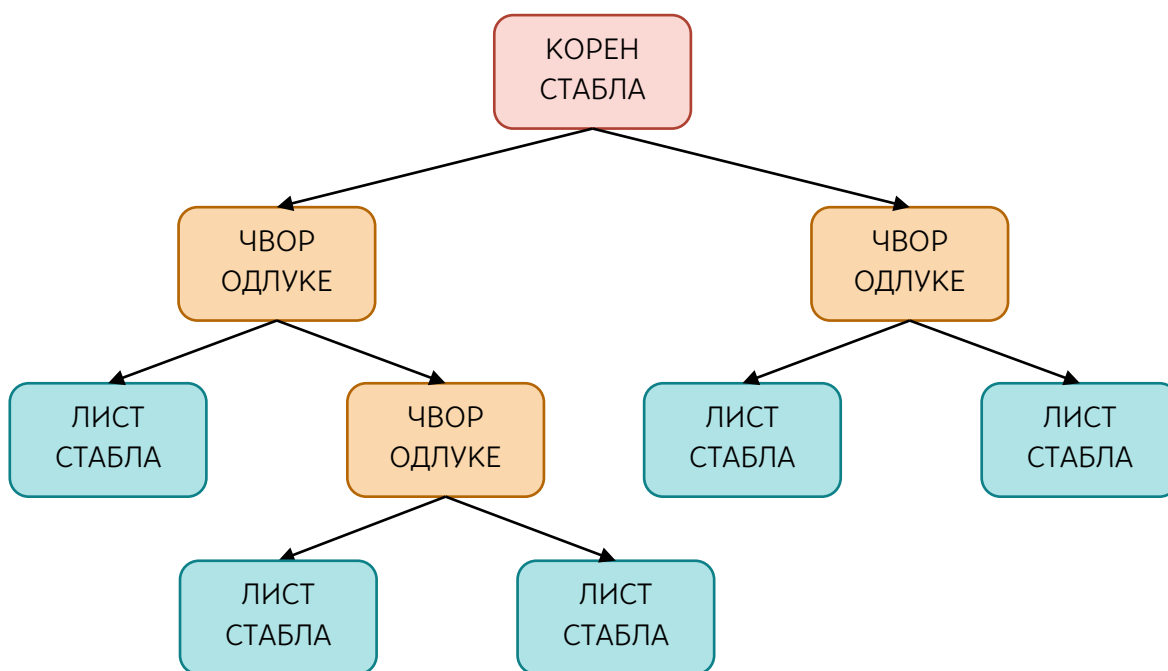
Машинско учење представља део области истраживања који се бави употребом рачунара ради симулирања процеса људског учења. То подразумева истраживање самоунапређујућих метода помоћу којих рачунари могу да идентификују постојећа и стичу нова знања и способности и константно повећавају своје перформансе и могућности (Wang и др., 2009). Машинско учење се може посматрати и као учење на основу примера, јер је реч о аутоматском приступу креирања емпиријских модела само на основу података. Том приликом није неопходно увођење било каквих претпоставки о подацима, њиховој расподели вероватноћа или функционалној форми, што су карактеристике непараметарског приступа. Технике машинског учења се могу применити и на решавање проблема регресије и класификације. За сваки од ових проблема на располагању је читав низ различитих техника, а неки од најчешће коришћених обухватају стабло одлучивања (енгл. *decision tree*), случајне шуме (енгл. *random forest*), неуронске мреже (енгл. *neural network*) и метод потпорних вектора (енгл. *support-vector machine*). У оквиру ове дисертације, нагласак је на надзираним непараметарским класификационим техникама машинског учења, где су у наредним поглављима детаљније приказане теоријске основе примењених алгоритама.

3.1.1 Стабло одлучивања

Стабло одлучивања (енгл. *decision tree*) представља веома популарну технику машинског учења намењену за решавање проблема класификације. Могућност

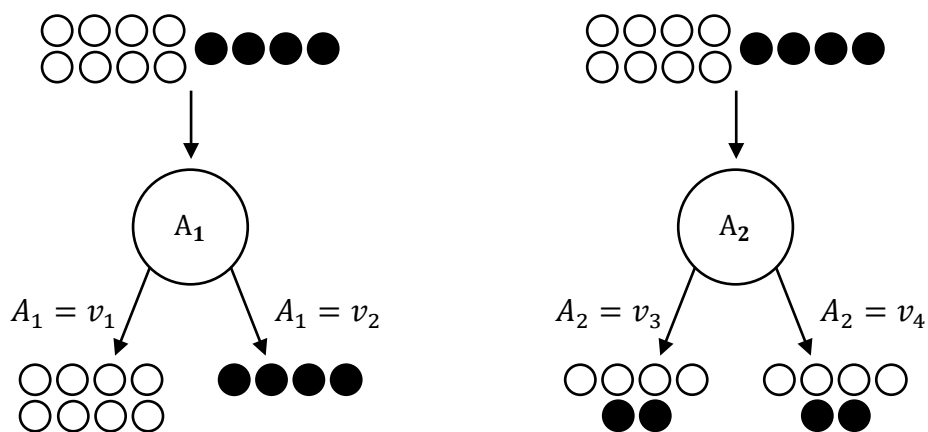
примене ове технике постоји и код проблема регресије. Реч је о једноставној техници која се ефикасно имплементира на основу скупа података (надзирана метода) и која је лака за разумевање и интерпретацију чак и корисницима који нису експерти. Постоји већи број алгоритама стабла одлучивања, од којих су најпознатији: *CART* (Breiman и др., 1984), *ID3* (Quinlan, 1986) и *C4.5* (Quinlan, 1993).

Стабло одлучивања се састоји од чворова одлуке и завршних чворова (лишће стабла), а почетни чвор одлуке се назива корен стабла (Слика 2.). Класификација стаблом одлучивања подразумева испитивање вредности атрибута посматраног примерка (инстанце) у чворовима одлуке, почевши од корена стабла. У сваком чвору одлуке се анализира вредност неког конкретног атрибута, а резултат анализе усмерава даљу путању кретања. На овај начин се пролази кроз стабло одлучивања док се не стигне до било код завршног чвора (листа) стабла у којем се налази циљна класа, тј. резултат класификације посматраног примерка. Свака путања од корена до листа стабла се може интерпретирати као конјункција (логичко „и“) тестова над атрибутима присутним на тој путањи. Пошто се иста циљна класа може наћи у више листова стабла, свака циљна класа се може интерпретирати као дисјункција (логичко „или“) свих путања које се завршавају том циљном класом, односно дисјункција над конјункцијама свих ограничења вредности атрибута примерака из расположивог скупа података (Samardžić-Petrović, 2014).



Слика 2. Општи изглед стабла одлучивања

Стабло одлучивања се креира користећи алгоритам учења од врха ка доле (енгл. *top-down induction of decision trees*), а процедура је позната и као „рекурзивно дељење“ (енгл. *recursive partitioning*) или „подели и освоји“ (енгл. *divide-and-conquer*). Алгоритам одабира најподеснији атрибут који ће представљати корен стабла и на основу њега дефинише критеријум поделе скупа података на непреклапајуће подскупове. У случају корена стабла, погодан атрибут је онај атрибут који ће разделити комплетан скуп података на што је могуће хомогеније подскупове (према циљној класи). Када се подела на основу одабраног атрибута изврши, процедура се рекурзивно понавља над резултујућим подскуповима података, односно комплетан скуп података се сукцесивно дели на све мање и мање непреклапајуће подскупове. Увек се тежи проналажењу најподеснијег атрибута, како би нови подскупови били што хомогенији (**Слика 3.**). Свака подела резултује новим чворовима одлуке и процедура се наставља све док се не направе подскупови у којима се налазе примерци само једне класе. Тада се завршава раст стабла, а резултујући чвор се проглашава листом стабла и означава одговарајућом циљном класом.



Слика 3. Одабир подесног атрибута за нови чвор стабла одлучивања

а) Идеалан избор атрибута A_1 , б) Непогодан избор атрибута A_2 (Samardžić-Petrović, 2014)

Критичан корак приликом креирања стабла одлучивања је одабир најподеснијег атрибута и критеријума на основу којих се врши подела на непреклапајуће подскупове. У случају дискретних атрибута, критеријум поделе је најчешће облика $A = v_i$, где v_i представља расположиве вредности атрибута. Код нумеричких атрибута се уобичајено користи бинарни критеријум поделе облика $A < v_t$, којим се постиже подела на подскупове са вредностима атрибута изнад и испод дефинисаног прага v_t . Ипак, треба имати на уму да неке варијанте стабла одлучивања врше тестирање и над више атрибута истовремено. Дobar одабир атрибута и критеријума поделе обезбеђује

поделу на непреклапајуће подскупове који су што је могуће хомогенији по питању циљне класе, тј. који су у идеалном случају потпуно хомогени. Одабир најподеснијег атрибута се најчешће утврђује користећи неку од функција којом се изражава нехомогеност или нечистоћа резултујућих подскупова (енгл. *impurity*). Најчешће мере нехомогености резултујућих подскупова које су у употреби су ентропија (Quinlan, 1986) и Гинијев индекс (енгл. *Gini index*) (Breiman и др., 1984). Обе функције су засноване на уделима примерака различитих класа у укупном скупу и дефинисане су изразима:

$$Entropy(S) = E(S) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad 3.1$$

$$Gini(S) = G(S) = 1 - \sum_{i=1}^c \left(\frac{|S_i|}{|S|} \right)^2 \quad 3.2$$

где је S посматрани почетни скуп података, а S_i су резултујући подскупови са примерцима циљне класе c_i . Обе функције имају максималне вредности када су резултујући подскупови са једнако дистрибуираним класама (када су сви S_i исте величине), а минималну вредност када неки S_i садржи све примерке класе ($S_i = S$). Дакле, потребно је одабрати онај атрибут који обезбеђује највеће умањење средње нечистоће чвора, тзв. информациони добитак (енгл. *Information Gain* – *IG*):

$$IG(S, A) = Impurity(S) - \sum_t \frac{|S_t|}{|S|} \cdot Impurity(S_t) \quad 3.3$$

где S_t представљају резултујуће непреклапајуће подскупове података S који настају поделом по атрибуту A , а $Impurity(S)$ може бити било која мера нечистоће чвора. Уобичајени проблем коришћења информативности атрибута је што атрибути који имају већи број различитих вредности лакше креирају хомогене подскупове, због чега се они преферирају у односу на атрибуте са мањим бројем вредности. Ради решавања овог проблема, често се врши нормализација информативности атрибута ентропијом $E(S)$, што се назива степеном информативности атрибута (енгл. *Information Gain Ratio* – *IGR*):

$$IGR(S, A) = \frac{IG(S, A)}{E(S)} = \frac{Impurity(S) - \sum_t \frac{|S_t|}{|S|} \cdot Impurity(S_t)}{\sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}} \quad 3.4$$

Применом описане процедуре гарантује се креирање стабла одлучивања за било који неконтрадикторан скуп података, што може довести до превелике прилагођености модела подацима (енгл. *overfitting*) и резултирати веома комплексним стаблима. Због тога се код већине алгоритама након креирања стабла врши орезивање (енгл. *pruning*) којим се уклањају поједини чворови и листови стабла у циљу ограничавања комплексности. Процедура неке од унутрашњих чворова одлуке практично проглашава листовима, а све наредне сегменте стабла одбацује. У том случају, новонастали листови стабла више нису потпуно хомогени (садрже примерке различитих класа), што се најчешће решава проглашавањем најчешће класе за излазну класу листа стабла. Истовремено је могуће одредити и грубу оцену вероватноће одређивања коначне класе, изражену као проценат примерака излазне класе у посматраном листу стабла.

3.1.2 Случајна шума

Случајна шума (енгл. *random forest*) је једна од најпопуларнијих техника машинског учења. Припада групи метода које користе скуп модела машинског учења, на основу чијих појединачних излаза се доносе заједничке одлуке (тзв. ансамбл техника машинског учења). Базирају на приступу да ансамбл више „слабих модела“ може бити комбинован тако да чини нови „моћни модел“. Код случајних шума ансамбл чини већи број стабала одлучивања (*Breiman, 2001*), где се свако стабло одлучивања креира над различитим подскуповима података. Важна карактеристика је да се не врши „орезивање“ креираних стабала. Када је шума стабала одлучивања креирана, излазна вредност случајне шуме се одређује агрегацијом излаза свих појединачних стабала. Агрегација подразумева средњу вредност резултата у случају регресије или најчешћу класу у случају класификације, где се најчешћа класа одређује већинским гласањем појединачних стабала. Подскупови података за креирање стабала одлучивања се уобичајено креирају користећи униформно узорковање са враћањем, познато и као „бутстрап узорковање“ (енгл. *bootstrapping*). На овај начин се постиже да су грешке појединачних стабала међусобно слабије корелисане, што резултује робуснијим решењем са мањом вероватноћом претераног прилагођавања модела подацима.

3.1.3 Градијентно појачавање

Градијентно појачавање (енгл. *gradient boosting*) такође припада групи ансамбл техника машинског учења, са стаблом одлучивања као најчешћом градивном јединицом ансамбла. Међутим, за разлику од случајних шума где се сваки модел креира независно, основна идеја појачавања (енгл. *boosting*) је да се модели у ансамблу креирају узимајући у обзир понашање других модела. На овај начин се, модел по модел $h_m(x)$, врши итеративно побољшавање (појачавање) текућег ансамбла модела F_m :

$$F_{m+1} = F_m(x) + h_m(x) \quad 3.5$$

где x представља скуп примерака (инстанци) над којима се креира модел.

Побољшавање се одређује преко разлике оцењених $F(x)$ и познатих вредности y , изражене преко неке функције губитка $L(y, F(x))$. Нови модел се креира и додаје у ансамбл тако да његови резидуали представљају негативни градијент функције губитка:

$$h_m(x) = y - F_m(x) = -\frac{\partial L(y, F(x))}{\partial F(x)} \quad 3.6$$

Већина алгоритама појачавања почиње од модела $F_0(x)$ и затим врши итеративно проширивање постојећег ансамбла F_{m-1} додавањем новог модела $h_m(x)$, тако да ансамбл F_m настаје на следећи начин:

$$F_m(x) = F_{m-1}(x) + \arg \min_{h_m \in \mathcal{H}} \left[\sum_{i=1}^N L(y_i, F_{m-1}(x_i) + h_m(x_i)) \right] \quad 3.7$$

где је $h_m \in \mathcal{H}$ базна функција учења (енгл. *base learner function*).

Пошто је проналажење најбоље функције h за произвољну функцију губитка L у сваком кораку прорачунски практично неизводљиво, врши се поједностављење претходног проблема применом опадајућег градијента (енгл. *gradient descent*):

$$F_m(x) = F_{m-1}(x) + \gamma_m \sum_{i=1}^N \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)), \quad 3.8$$
$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)))$$

где је γ_m корак опадајућег градијента.

У највећем броју случајева се користи стабло одлучивања као градивна јединица ансамбла над којим се примењује градијентно појачавање. У сваком кораку m , над резидуалима се креира стабло одлучивања $h_m(x)$ са J_m листова. Креираним стаблом одлучивања се врши сегментација атрибутског простора на J_m -непреклапајућих региона $\{R_{jm}\}_{j=1}^{J_m}$, где су излази модела у тим регионима константни. Излаз стабла одлучивања се може изразити као:

$$h_m(x) = \sum_{j=1}^{J_m} b_{jm} I(x \in R_{jm}) \quad 3.9$$

где је b_{jm} излаз стабла одлучивања у региону R_{jm} . У том случају, појачавање градијента подразумева да се коефицијенти b_{jm} множе неком вредношћу γ_m којом се умањује функција губитка, тако да резултујући модел гласи:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \quad 3.10$$

Фридман је предложио модификацију тако да се уместо јединственог γ_m за цело стабло, одређују засебна γ_m за сваки регион одлучивања (Hastie и др., 2009). Због тога коефицијенти b_{jm} могу бити занемарени и такав модел градијентног појачавања гласи:

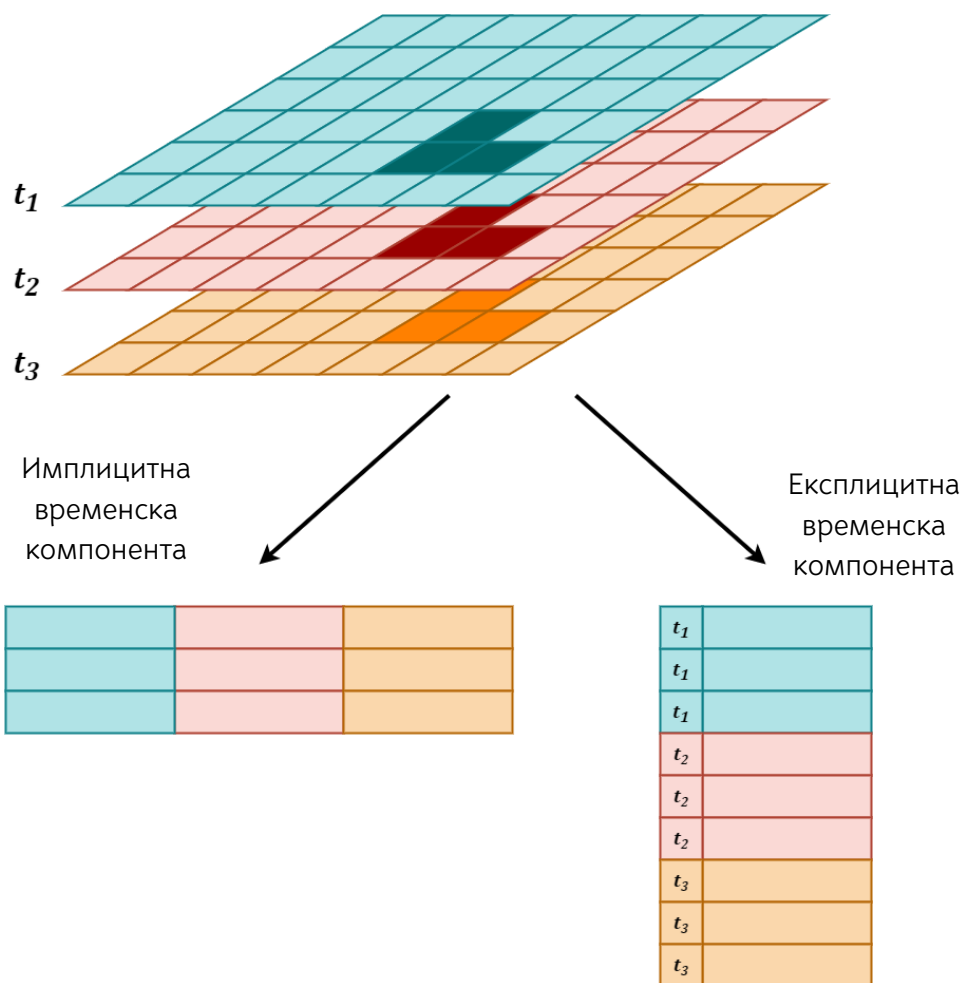
$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$$

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad 3.11$$

3.2 Временска компонента у класификационом моделу

Велики број истраживања је показао да коришћење временских серија сателитских података обезбеђује робуснија решења и повећава тачност и квалитет класификационог модела. Међутим, највећи број техника машинског учења нису предодређене за укључивање временске компоненте у модел класификације, већ је ове карактеристике неопходно укључити на неки други начин. Ако се изузму методе које су већ дизајниране са уграђеним временским аспектима, већина техника машинског

учења може да буде надограђена тако да респектује временске компоненте у класификационом моделу. То се може решити имплицитно, тако што се сваки временски тренутак/интервал за који постоје подаци репрезентује скупом независних атрибута у моделу класификације (Grabska и др., 2019; Sheeren и др., 2016). Алтернативно, временски аспекти могу бити и експлицитно дефинисани, тако што се за свако опажање води један или више временских атрибута (нпр. дан у години, кумулативни дан у години, датум итд.) (He и др., 2016) (Слика 4.).



Слика 4. Илустрација репрезентације података сателитског опажања са имплицитно и експлицитно дефинисаном временском компонентом

Када је реч о класификацији садржаја серије сателитских снимака, уобичајен је приступ имплицитног дефинисања временских компоненти, где се за сваку просторну локацију (пиксел) воде одвојени скупови атрибута свих спектралних канала за сваки временски тренутак/интервал (Grabska и др., 2019; Sheeren и др., 2016). Ово подразумева да се број атрибута у моделу повећава како је број временских одредница већи. Када је реч о дугачким временским серијама података, ово повећање атрибута може

проузроковати проблеме са меморијским захтевима и перформансама рачунара. Истовремено, треба имати на уму да се различите технике машинског учења различито носе са великим број атрибута, што ограничава применљивост појединих техника. Ипак, много већи проблем оваквог приступа је што он захтева да су за сваки део подручја истраживања на располагању опажања за сваку временску одредницу која је репрезентована у класификационом моделу. Овај захтев је посебно проблематичан када је реч о оптичким сателитским опажањима, где је неопходно избегавати временске тренутке са облацима. Како се подручје истраживања повећава, постаје све теже обезбедити податке без облака са макар сличним временским одредницама. За нека изузетно велика подручја са специфичним климатским условима је практично немогуће обезбедити опажања која нису захваћена облацима. Постоји више начина за превазилажење овог проблема, сваки са одређеним предностима и манама. Један од приступа подразумева увођење грубље временске резолуције у модел и коришћење месечних, сезонских или годишњих композита сателитских опажања. Оваква генерализација подразумева специфичне радиометријске корекције и филтере приликом израде композита, што уводи додатне несигурности у модел. Грубља временска резолуција је посебно ограничавајући фактор када је реч о сателитском праћењу вегетације, јер може довести до губитка нијанси у фенолошким обрасцима, што на крају може утицати на грешке у класификацији. Други приступи су усмерени ка креирању посебних процедура и метода за попуњавање недостајућих података услед облачности коришћењем различитих техника (*Meraner* и др., 2020; *Pipia* и др., 2019; *Yan* и *Roy*, 2018). Ови приступи значајно усложњавају процедуру класификације, смањују просторну резолуцију или нису довољно ефикасни када су рупе у подацима превише велике.

У оквиру овог истраживања, разматран је алтернативни приступ и предности и могућности експлицитног коришћења временских компоненти као атрибута у класификационом моделу. Главна последица оваквог приступа је што се на овај начин производи скуп излаза класификације за сваки пиксел подручја од интереса, при чему број излаза зависи од броја расположивих сателитских опажања. То практично значи да је у процедуру класификације неопходно увести додатан корак којим ће бити извршена агрегација вишеструких излаза класификације, како би се одредила коначна класа за сваки пиксел подручја истраживања. Тестирано је више функција за агрегацију, где је коначна класа одређена коришћењем једног од следећих правила:

- 1) најчешће присутна класа (енгл. *Most Common* – *MC*);
- 2) класа са највећом вероватноћом – проста аритметичка средина (енгл. *Simple Mean* – *SM*);
- 3) класа са највећом вероватноћом – геометријска средина (енгл. *Geometric Mean* – *GM*).

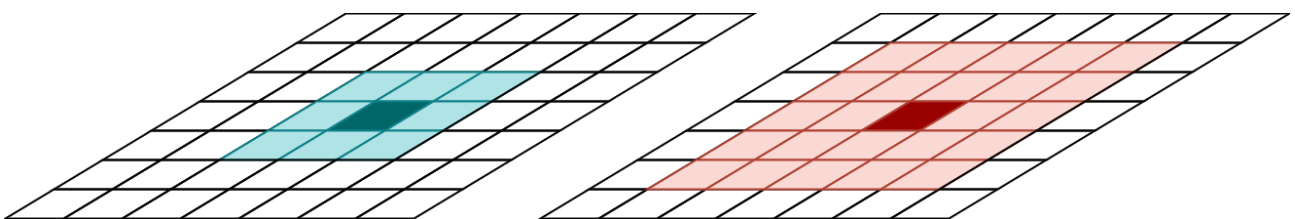
Правило најчешће присутне класе за посматрани пиксел гласи да класификациони модел креира излаз за свако опажање и даје по један глас оној класи за коју има највећу оцењену вредност вероватноће. Затим се коначна класа пиксела одређује као она класа која има највећи број гласова. Преостала два приступа су базирана на осредњавању експертских предвиђања (енгл. *averaging expert predictions*) (Kivinen и Warmuth, 1999). То подразумева да се излаз класификационог модела за свако опажање посматра као једно експертско предвиђање. Оцењене вредности вероватноће се затим осредње применом просте аритметичке или геометријске средине. Не врши се трансформација вредности оцењених вероватноћа нити се уводе тежине приликом рачунања средњих вредности. Коначна класа пиксела се на крају одређује као она класа која има највећу средњу вредност оцењене вероватноће.

3.3 Просторна компонента у класификационом моделу

Класификација садржаја сателитских снимака је у великој мери базирана на првом Тоблеровом закону географије који гласи: „Све је повезано са свиме, али су блиске појаве међусобно више повезане него оне удаљене“ (Tobler, 1970). Наравно, у случају сателитских опажања се уместо просторне блискости нагласак ставља на спектралну блискост. Свакако, и просторне карактеристике података морају бити узете у обзир у класификацији, макар имплицитно. Просторна компонента сваког пиксела (локација и просторна резолуција) је неизоставни део комбиновања скупа растерских података, а усаглашавање просторне резолуције и просторно поравнавање растера је саставни део већине класификационих процедура. Ипак, уколико појава која је предмет класификације показује одређене просторне правилности, постоји оправдана бојазан да класификациони модел неће моћи квалитетно да измоделира такве промене када је просторна компонента имплицитно дефинисана.

Као што је то случај и са временском компонентом, технике машинског учења нису предодређене ни за просторне аспекте података. Због тога је потребно применити неки

од алтернативних приступа како би се просторна компонента нагласила у класификационом моделу. Најједноставнији приступ подразумева експлицитно вођење просторне компоненте кроз скуп координата пиксела као додатних атрибута у моделу (He и др., 2016; Li и др., 2011). На располагању су и сложенији приступи, где се као додатни атрибути модела воде просторно најближа опажања (Sekulić и др., 2020). Проблем са претходним приступима је што су претежно конципирани за примену над векторским структурама података (нпр. тачкаста мерења) и што не узимају у обзир карактеристике растерске структуре података. Укључивање суседства посматраног пиксела у процедуру класификације је једноставна имплементација Тоблеровог закона географије, а која се ослања на растерску структуру података. Aonpong и др. (2016) су повећали квалитет класификације земљишног покривача укључивањем суседства пиксела у процедуру гласања циљне класе код случајне шуме. Samardžić-Petrović и др. (2016) су користили најчешћу класу у Муровом суседству (енгл. *Moore's neighbourhood*) за одређивање краткорочних промена начина коришћења земљишта, што се показало као један од најинформативнијих атрибута у креираном моделу. У оквиру ове дисертације, идеја о наглашавању просторне компоненте је инспирисана комбинацијом претходно наведених приступа (Aonpong и др., 2016; Samardžić-Petrović и др., 2016, 2017). Примењен је приступ агрегације оцењених вредности вероватноћа у Муровом суседству посматраног пиксела (Слика 5.), при чему су примењена иста правила агрегације описана код респектовања временске компоненте (најчешћа класа, класа са највећом вероватноћом срачуната простом и геометријском средином).



Слика 5. Илустрација Муровог суседства посматраног пиксела величине 3x3 и 5x5

Основна идеја приступа заснованог на суседству је решавање ситуација у којима околни пиксели показују велику вероватноћу припадности одређеној класи, која се разликује од класе централног пиксела. У овим случајевима постоји оправдана сумња да и централни пиксел такође треба придружити класи околних пиксела, јер је вероватније да је извршена погрешна класификација као последица шума присутног у спектралним информацијама, него да је таква ситуација заиста и на терену. Идеја предложене

методологије је да ће управо применом просторно-временске агрегације овакви случајеви бити успешно препознати и решени.

3.4 Креирање и избор репрезентативних атрибута

Већ је раније наведено да картирање типова шумске вегетације на основу сателитских опажања Земље и даље представља изазов, пре свега због великог броја разнородних утицаја који мењају спектралне карактеристике шумске вегетације (**поглавље 2.2.1**). Због тога се, поред спектралних канала сателитске мисије, у процедуру класификације често укључују и бројни додатни подаци. Подаци о висинској представи терена, климатски подаци, педолошки и геолошки подаци, постојеће карте земљишног покривача и бројни други извори података су се показали као корисни када је реч о квалитетнијем издвајању типова шума. Поред тога, често се креира и скуп изведених карактеристика применом одговарајућих функција над постојећим подацима. Тако се комбинацијом одговарајућих спектралних канала може креирати већи број различитих спектралних индекса, док се на основу висинске представе терена може израчунати скуп топографских карактеристика попут нагиба, аспекта, храпавости итд.

Укључивањем додатних извора података и креирањем скупа изведених карактеристика као атрибута класификационог модела увећава се димензионалност података. Истовремено, неки од тих атрибута су неинформативни или редундантни. Уколико се на почетку располаже са скупом од n атрибута, то значи да је на располагању 2^n потенцијалних скупова атрибута. Испитивање свих комбинација гарантује избор најбољег скупа атрибута, али је јасно да је ово неизводљиво када год је број расположивих атрибута иоле већи. Основна идеја избора репрезентативних атрибута (енгл. *feature selection*) јесте „паметно“ претраживање потенцијалних комбинација атрибута, тако да се креира скуп атрибута од користи за класификациони модел, а да се притом искључе они атрибути који мало или нимало не доприносе квалитету класификације (Liu, 2010). На овај начин се повећава тачност модела, смањује се комплексност модела уз краће време тренирања (учења) и избегава се превелика прилагођеност модела подацима (Samardžić-Petrović, 2014). Генерално се могу разликовати филтер-методе (енгл. *filter methods*) и омотач-методе (енгл. *wrapper methods*) за избор репрезентативних атрибута. Не постоји универзално решење за избор репрезентативних атрибута зато што свака од поменутих група метода

располаже великим бројем метода, свака са својим предностима и манама. У оквиру овог истраживања је испитано више метода за избор атрибута из обе групе, а у наставку ће бити укратко описана свака од њих.

3.4.1 Филтер-методе за избор репрезентативних атрибута

Филтер-методе врше рангирање важности атрибута на основу статистичких метрика и затим издвајају подскуп оних најважнијих. То се уобичајено изводи у два корака. Прво се врши рангирање атрибута на основу одређеног критеријума, а затим се у скуп репрезентативних атрибута бирају високо ранжирани атрибути. Предложен је велики број критеријума за рангирање атрибута који се заснивају на:

- 1) могућности разликовања класа на основу варијансе између класа и унутар-класне варијансе;
- 2) зависности између атрибута и циљне класе;
- 3) корелацији између атрибута и циљне класе и између атрибута међусобно;
- 4) способности да се очува структура многострукости (енгл. *manifold structure*);
- 5) заједничким информацијама између атрибута итд.

Предност филтер-метода је што се ради о веома ефикасном и робусном приступу у смислу управљања превелике прилагођености модела подацима. Мана је што се не узима у обзир пристрасност и хеуристика технике машинског учења, што значи да коначни скуп одабраних атрибута не мора бити оптималан за конкретни алгоритам машинског учења. У оквиру ове дисертације, пажња је усмерена на филтер-методу за избор атрибута базирану на корелацији (енгл. *Correlation-based Feature Selection - CFS*).

3.4.1.1 Избор атрибута заснован на корелацији

Избор атрибута заснован на корелацији врши рангирање атрибута користећи хеуристичку функцију базирану на корелацији (Hall, 2000). CFS метода покушава да креира скуп атрибута где сваки одабрани атрибут има јаку способност класификације, али тако да истовремено у скупу нема сувишних атрибута. Метода претпоставља да су мање информативни атрибути слабо корелисани са циљном променљивом и да као такви могу бити занемарени. Истовремено, атрибути који су јако корелисани са једним или више других атрибута такође могу бити занемарени, јер је реч о редундантним

атрибутима. Критеријумска функција која формализује претходно описану хеуристику гласи:

$$Merit_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \quad 3.12$$

где је $Merit_S$ хеуристичка мера скупа атрибута S који садржи k атрибута, \bar{r}_{cf} је средња вредност корелације између атрибута и циљне класе и \bar{r}_{ff} је средња вредност крос-корелације између атрибута. Како би се обезбедила основа за одређивање корелација у претходној једначини (израз 3.12), сви континуални атрибути се претходно преводе у дискретне атрибуте. Ако су X и Y независне дискретне променљиве, као мера асоцијације између ових двеју променљивих може се искористити информациони добитак (израз 3.3). Међутим, проблем представља што је информациони добитак по природи пристрастан према категоричким атрибутима који имају већи број вредности и што корелације изражене на овај начин нису нормализоване, чиме је онемогућено њихово поређење. Због тога се уводи мера симетричне неодређености (енгл. *Symmetrical Uncertainty – SU*) (Hall и Smith, 1999):

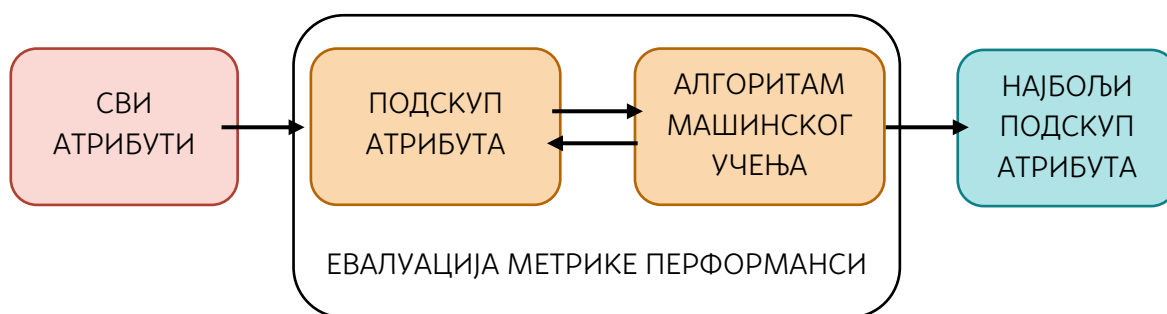
$$SU(Y, X) = 2 \times \left[\frac{IG(Y, X)}{E(Y) + E(X)} \right] \quad 3.13$$

где $IG(Y, X)$ представља прираст информација изражен преко ентропије, где су $E(Y)$ и $E(X)$ вредности ентропије атрибута X и Y (израз 3.1). На овај начин, *CFS* метода користи меру која компензује недостатке информационог добитка и обезбеђује непристрасну меру корелације која је додатно и нормализована у интервалу $[0, 1]$. *CFS* метода практично тестира информативност већег броја скупова атрибута и покушава да одреди скуп који је најпогоднији. Процедура почиње од празног скупа атрибута, а затим се атрибути додају применом принципа претраге првог најбољег унапред (енгл. *forward best first search*), где се за критеријум за престанак узима пет узастопних проширења скупа атрибута без побољшања (Hall и Smith, 1999).

3.4.2 Омотач-методе за избор репрезентативних атрибута

Омотач-методе за избор репрезентативних атрибута одређују скуп погодних атрибута на основу усвојене метрике и применом конкретног алгоритма машинског учења (Слика 6.). Идеја је да екстремној вредности усвојене метрике одговара

репрезентативни скуп атрибута који треба одредити одређеним механизмом претраге. Метрике за оцену перформанси скупа атрибута се најчешће одређују крос-валидацијом, а на располагању је већи број различитих метрика у зависности од области примене и проблема који се решава. Циљ је да се усвоји онај скуп атрибута за који класификација резултира екстремним вредностима усвојене метрике. Код класификационих проблема као метрика се може користити укупна тачност (енгл. *overall accuracy*), капа коефицијент слагања (енгл. *kappa coefficient of agreement*), површина испод *ROC* криве (енгл. *area under the ROC curve*) итд. Методе из ове групе испитују велики број подскупова атрибута итеративним укључивањем и искључивањем атрибута из скупа док се не задовољи услов завршетка претраге. Идеално, претрага ће бити прекинута када је пронађена глобална екстремна вредност метрике, којој одговара и најрепрезентативнији скуп атрибута. Мана је што се претрага може завршити локалним екстремом, што може негативно утицати на перформансе модела. Зато примена ових приступа прорачунски може бити веома скупа. Ипак, на овај начин се узима у обзир пристрасност и хеуристика конкретне методе класификације и превазилазе главни недостаци филтер-метода.



Слика 6. Омотач-методе за избор репрезентативних атрибута

Кључна карактеристика омотач-метода обухвата дефинисану стратегију претраге репрезентативних атрибута, тј. механизам укључивања/искључивања атрибута из скупа репрезентативних атрибута. Према *El Aboudi и Benhlina (2016)* омотач-методе за избор репрезентативних атрибута се према стратегији претраге разврставају на:

- 1) експоненцијално захтевне методе;
- 2) еволуционе методе;
- 3) секвенцијалне методе.

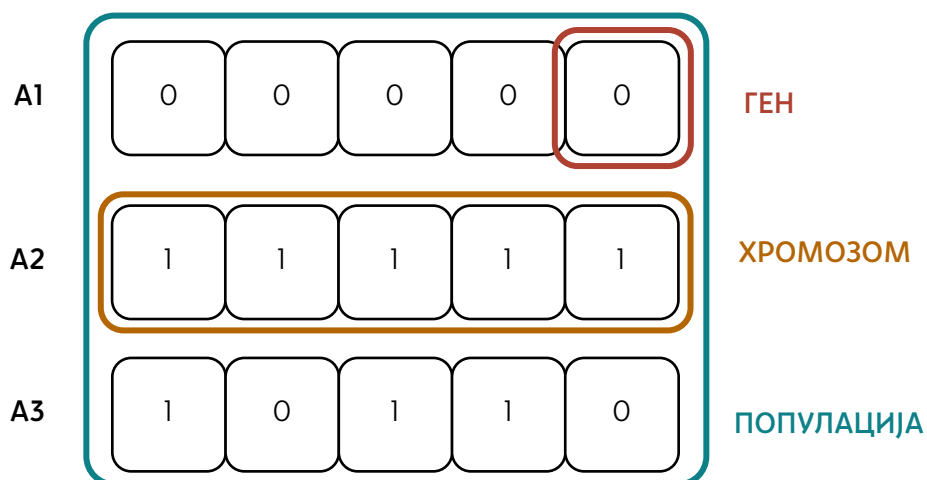
Експоненцијално захтевне методе разматрају све могуће комбинације атрибута, због чега гарантују проналажење оптималног решења. Ипак, као што је раније већ

напоменуто, ове методе су непрактичне за веће скупове атрибута због огромних захтева у погледу прорачуна. Еволуционе методе, познате и као методе базиране на популацији, користе одређену оптимизациону метахеуристику која покушава да опонаша принцип еволуције ка бољим решењима који постоји у природи. Неки од најчешће коришћених еволуционих метода за избор репрезентативних атрибута су генетски алгоритми (енгл. *genetic algorithms*), алгоритам колоније мрава (енгл. *ant colony algorithm*), оптимизација ројем честица (енгл. *particle swarm optimization*) итд. Секвенцијалне методе су базиране на приступу „похлепне“ претраге (енгл. *greedy search approach*). Методе ове групе на неки начин иницијализују почетни скуп атрибута, а затим се атрибути итеративно убацују или избацују из скупа атрибута.

3.4.2.1 Избор репрезентативних атрибута применом генетских алгоритама

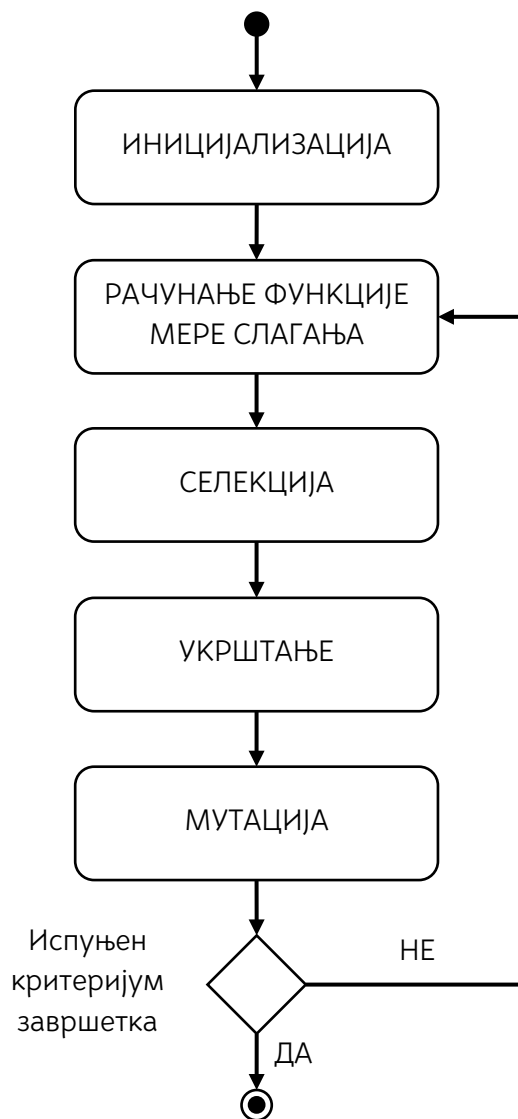
Генетски алгоритми су алгоритми засновани на популацији који се воде идејом о биолошкој еволуцији и природној селекцији, где најподобнији појединци преживљавају. Основне принципе алгоритма је поставио *Holland* седамдесетих година 20. века, а касније су *De Jong* и *Goldberg* показали примену генетских алгоритама за решавање комплексних проблема (*De Jong, 1975; Goldberg, 1989; Holland, 1975*).

Решења се у генетским алгоритмима представљају хромозомима (**Слика 7.**). Решења се састоје од гена чија је улога да представе различите особине које решење има. У случају избора репрезентативних атрибута, гени у хромозому представљају укључивање и искључивање атрибута из репрезентативног скупа атрибута.



Слика 7. Илустрација бинарно кодираних атрибута генетског алгоритама
A1 – празан скуп атрибута, A2 – сви атрибути, A3 – први, трећи и четврти атрибут

Генетски алгоритам на почетку креира насумичан скуп решења из популације, што представља прву генерацију. У свакој генерацији се дешава репродуктивни процес, који креира наредну генерацију решења. Резултат су потомци решења који задржавају неке од карактеристика сваког родитеља. Квалитет сваког решења се мери вредношћу усвојене метрике, која се често назива функција мере слагања (енгл. *fitness function*). На основу те функције се проверава подобност појединих решења, као и изгледи да то решење опстане кроз генерације. Аналогно биолошким еволутивним процесима, већа је вероватноћа да ће потомци с релативно добром метриком преживети и даље се репродуковати. Основни принцип је да су добри изгледи да ће се рекомбинацијом добрих гена добити још боље решење. Идејно, кроз много генерација ће се добре карактеристике проширити кроз читаву популацију (Beasley и др., 1993). Графичка шема функционисања генетског алгоритма приказана је на наредној слици (Слика 8.).



Слика 8. Општи принцип функционисања генетског алгоритма

Основни оператори генетских алгоритама који дефинишу креирање наредних генерација решења су:

- 1) селекција (енгл. *selection*);
- 2) укрштање (енгл. *crossover*);
- 3) мутација (енгл. *mutation*).

Улога оператора селекције јесте да се одреде најпогоднији хромозоми (тј. решења) над чијим ће се генима у наредној генерацији алгорита применити други оператори. Најбољи хромозоми се одређују коришћењем неке критеријумске функције. Резултат може бити било која комбинација n гена (где је n предефинисана дужина хромозома), али је већа вероватноћа да ће бити кодирано оно решење које има бољу вредност критеријумске функције. Избор се најчешће врши методом точка среће (пропорционалне селекције) или методом турнира.

Укрштање је најзначајнији оператор генетског алгоритама и подразумева да се од два постојећа селекцијом одабрана хромозома из претходне генерације (родитељи) укрштањем креирају два нова хромозома (потомци). Јединке се замењују потомцима, док се остале преузимају из претходне генерације. Размена делова гена се обавља на случајно одабраном месту у хромозому. Очекује се да ће се овом разменом претрага оптимума усмерити ка делу простора за претраживање с погоднијим решењима.

Оператор мутације потенцира малу генетичку разноликост међу решењима. Мутирање (најчешће једног гена) подразумева доделу различите случајне вредност из расположивог домена. Предност мутације је да она, за разлику од укрштања, може да уведе нови квалитет у решење. Међутим, треба бити свестан да постоји ризик да се добро решење поквари или, као и у реалности, појаве дегенеративне јединке.

Успешност генетског алгоритама највише зависи од одговора на питања: како се формира иницијална популација, шта се користи као критеријумска функција, на који начин се врши селекција, на који начин се врши укрштање, на који начин се врши мутација и који је услов за прекид алгоритама. Избор параметара којима се описују ове карактеристике захтева посебну пажњу, јер од њих у великој мери могу зависити и постигнути резултати. Највећа мана генетских алгоритама је што захтевају велики број прорачуна, с обзиром на то да се креира и испитује модел за свако решење. Поред тога,

због своје стохастичке природе, некада је потребан велики број генерација како би решење конвергирало.

3.4.2.2 Секвенцијалне методе избора репрезентативних атрибута

Секвенцијалне методе за избор репрезентативних атрибута се међусобно разликују по начину одабира почетног скупа атрибута, као и по смеру претраге. Међу популарнијим приступима из ове групе су: приступ секвенцијалне претраге унапред (енгл. *Sequential Forward Search – SFS*), секвенцијалне претраге уназад (енгл. *Sequential Backward Search – SBS*) и претрага планинарењем (енгл. *Hill Climbing Search – HCS*). Секвенцијална претрага унапред почиње од празног скупа атрибута, затим итеративно процењује сваки расположиви атрибут и на крају најбољи одабира у скуп репрезентативних атрибута. Процедура се понавља тако да се у свакој итерацији додаје по један атрибут, докле год се мере перформанси модела побољшавају. Секвенцијална претрага уназад функционише по сличном принципу, с тим што почиње од комплетног скупа атрибута и врши избацивање једног по једног, докле год долази до побољшања мера перформанси. У оба случаја претрага тече у само једном правцу, тј. атрибути се или додају или избацују из коначног скупа репрезентативних атрибута. Претрага планинарењем представља другачији приступ, који започиње случајним скупом атрибута, а затим процењује све своје суседе и бира најбољег. Недостатак свих ових метода је што врше доста ограничену претрагу комбинација атрибута, што не гарантује проналажење оптималног решења нити брзо конвергирање решења (El Aboudi и Benhlina, 2016).

3.5 Балансирање класа

Скуп података је оптерећен проблемом небалансираних класа (енгл. *class imbalance problem*) који се јавља када је дистрибуција класа изразито различита, односно када постоји једна или више класа које доминирају по питању броја примерака у односу на друге класе. Овај проблем је често последица небалансираности класа у стварном свету, али може бити и последица (не)доступности података или начина узорковања. Без обзира на узрок, коришћење небалансираног скупа података за креирање класификационог модела може значајно утицати на његове перформансе. По правилу се то манифестује тако што модел мањинске класе разврстава као већинске класе.

У случају када узорковање додатних примерака мање заступљених класа није могуће, што је често и случај, најједноставнији приступ је подузорковање (енгл. *undersampling*) већинских класа или дуплирање примерака мањинских класа. Прикладнија алтернатива је да се примерци мањинских класа вештачки креирају применом одговарајућих процедура над већ расположивим скупом података. Техника *SMOTE* (енгл. *Synthetic Minority Oversampling TEchnique*) представља једну од популарнијих метода за решавање проблема небалансираних класа (*Chawla* и др., 2002; *Torgo*, 2010). Ова метода функционише тако што се примерци мањинских класа вештачки креирају на основу најсличнијих расположивих примерака те класе. Конкретно, за насумично одабран примерак мањинске класе, *SMOTE* прво проналази његових k најближих суседа у атрибутском простору. Насумичним одабиром једног од k суседа, формира се линијски сегмент у атрибутском простору између одабраног примерка и суседа, а затим се вештачки примерци креирају као комбинација претходно одабраног примерка и суседа. Показано је да комбинација *SMOTE*-а и подузорковања већинских класа резултује вишим перформансама него применом само подузорковања (*Chawla* и др., 2002). Предност овог приступа што је могуће креирати онолико примерака мањинских класа колико је потребно, при чему се може очекивати да су вештачки примерци веома слични стварним због блискости у атрибутском простору. Генерални недостатак поменуте методе је што се приликом креирања нових примерака мањинске класе уопште не узимају у обзир подаци других класа, што може увести двосмислене примерке у случају када постоји значајно преклапање између класа.

3.6 Оцена квалитета класификације

Непристрасна и квалитетна оцена квалитета извршене класификације је веома важна компонента сваке класификационе процедуре. У оквиру овог истраживања, оцена квалитета класификације је пресудна ради компаративне анализе различитих аспеката предложене методологије картирања шумске вегетације. На овај начин је омогућено одређивање оптималних параметара методологије, али и извођење оцене квалитета коначне излазне карте шумске вегетације израђене на подручју истраживања. Због тога је у наредним поглављима описана примењена процедура крос-валидације и усвојених квантитативних мера квалитета класификације.

3.6.1 Квантитативне мере перформанси класификационог модела

Уобичајена процедура оцене тачности класификације подразумева идентификовање два скупа података, оба са познатим излазним класама. На првом скупу података се врши креирање класификационог модела (тренинг скуп), који се затим примењује на другом скупу (валидациони скуп). Мере перформанси се одређују поређењем класификованих и референтних (познатих) излазних класа на валидационом скупу података, што се може приказати матрицом конфузије (Табела 3.). Вредности у матрици конфузије представљају проценат примерака валидационог скупа који су класификовани као класа i и имају референтну класу j . Дакле, дуж дијагонале матрице конфузије се налазе добро класификована опажања (у овом случају пиксели), док вандијагоналне вредности одговарају грешкама изостављања и уврштавања (енгл. *omission and commission errors*).

Табела 3. Општи изглед матрице конфузије

		Референтне вредности					
		Класа	1	2	3	...	N
Класификација	1	p_{11}	p_{12}	p_{13}	...	p_{1N}	p_{1*}
	2	p_{21}	p_{22}	p_{23}	...	p_{2N}	p_{2*}
	3	p_{31}	p_{32}	p_{33}	...	p_{3N}	p_{3*}

	N	p_{N1}	p_{N2}	p_{N3}	...	p_{NN}	p_{N*}
	Σ	p_{*1}	p_{*2}	p_{*3}	...	p_{*N}	1

На основу матрице конфузије се могу израчунати различите квантитативне мере перформанси класификационог модела, од којих су у оквиру ове дисертације усвојене: укупна тачност (енгл. *Overall Accuracy – OA*), произвођачева тачност (енгл. *Producer’s Accuracy – PA*), корисникова тачност (енгл. *User’s Accuracy – UA*) и капа коефицијент слагања (енгл. *kappa coefficient of agreement*) (Olofsson и др., 2014).

Укупна тачност представља проценат добро класификованих опажања у односу на укупан број опажања, који се у случају са N класа рачуна:

$$OA = \sum_{j=1}^N p_{jj} \quad 3.14$$

Произвођачева тачност за класу j представља проценат референтних примерака класе j које су и класификоване као класа j :

$$PA = \frac{p_{jj}}{p_{*j}} \quad 3.15$$

док корисникова тачност за класу i представља проценат свих примерака класификованих као класа i , а који и имају референтну класу i :

$$UA = \frac{p_{ii}}{p_{i*}} \quad 3.16$$

Капа коефицијент слагања представља квантитативну меру перформанси базирану на укупној тачности кориговану за случајне утицаје, која се рачуна следећим изразом (Cohen, 1960):

$$Kappa = \frac{OA - SU}{1 - SU} \quad 3.17$$

где SU представља удео слагања који је последица случајности и који се рачуна:

$$SU = \sum_{j=1}^N p_{j*} \cdot p_{*j} \quad 3.18$$

Капа коефицијент слагања може имати вредности у опсегу од -1 до 1 , где негативне вредности показују на одсуство слагања, док се слагање сматра бољим што су вредности овог коефицијента ближе 1 .

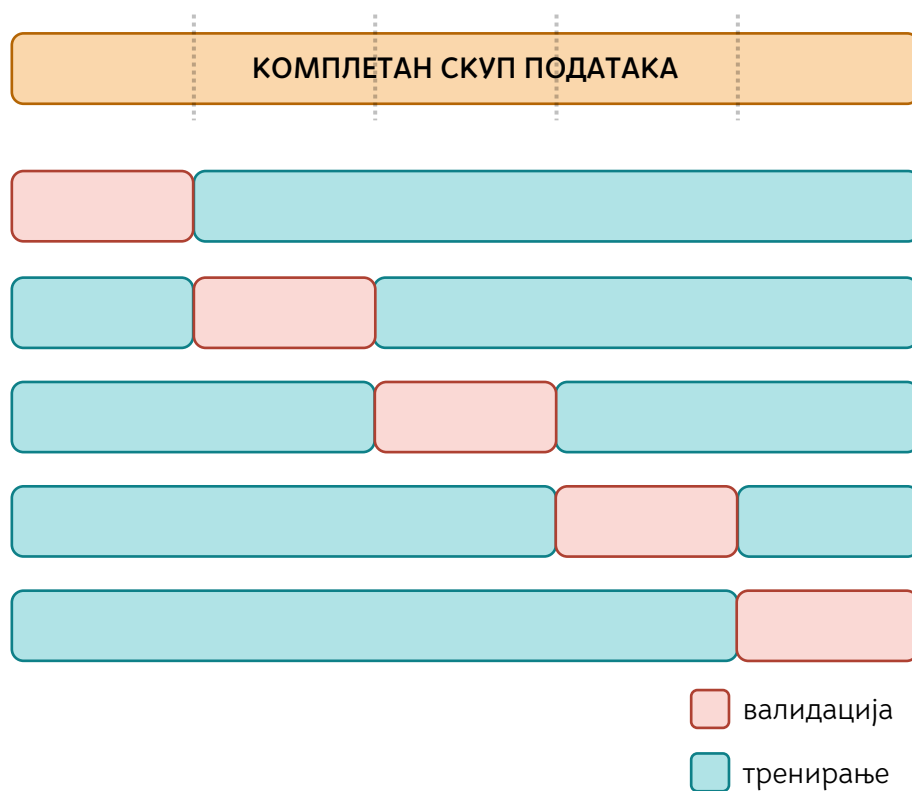
3.6.2 Крос-валидација са изостављањем комплетних локација

Крос-валидација (енгл. *cross-validation*) представља веома популаран приступ оцене квалитета извршене класификације у случајевима када за те намене није на располагању други, независан репрезентативан скуп података. Уместо тога, процедура крос-валидације врши поделу комплетног постојећег скупа у K непреклапајућих подскупова (енгл. *K-folds*) сличне величине. У сваком кораку се један подскуп посматра као скуп за оцену квалитета (валидациони скуп), а примерци из преосталих $K - 1$ подскупова се обједињују и користе за креирање класификационог модела. Креирани класификациони модел се затим примењује на валидационом скупу. На овај начин се креира K класификационих модела, који се независно примењују на валидационим скуповима. Излази класификације модела се упоређују са референтним вредностима класа валидационих скупова, што се затим користи за оцену квалитета извршене класификације.

Недостатак крос-валидације је што квалитет оцене тачности зависи од тога како је извршена подела комплетног скупа података на непреклапајуће подскупове. Пре свега је важно очувати, колико је то могуће, однос дистрибуције класа у сваком резултујућем подскупу који је присутан и у оригиналном скупу података. Када се ради о анализама у којима постоји просторни аспект, пожељно је и тај аспект обухватити приликом креирања подскупова. За те потребе су предложене посебне варијанте крос-валидација. Крос-валидација са изостављањем комплетних локација (енгл. *Leave-Location Out Cross-Validation - LLOCV*) представља једну такву надградњу стандардне крос-валидације. У тој варијанти се разврставање примерака у подскупове спроводи на такав начин да се сви примерци који припадају некој просторној локацији/региону (нпр. полигон са свим припадајућим опажањима) увек комплетно размешта у само један подскуп.

У оквиру ове дисертације је примењена петострука крос-валидација са изостављањем комплетних локација (**Слика 9.**). Приликом креирања подскупова вођено је рачуна да се сачува оригинална дистрибуција класа комплетног скупа података. Додатно је уведен још један ниво случајности, где је цела процедура крос-валидације понављана пет пута. То практично значи да је свако опажање 20 пута коришћено за учење класификационог модела, а пет пута за валидацију. Коначна оцена тачности је затим

изведена над обједињеним валидационим скупом података, рачунајући претходно наведене квантитативне мере за оцену перформанси класификационог модела.



Слика 9. Илустрација процедуре петоструке крос-валидације

4

КАРТИРАЊЕ ТИПОВА ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ

Најважније карактеристике предложене методологије за картирање типова шумске вегетације на основу података сателитског осматрања Земље су изложене у оквиру четвртог поглавља дисертације. На почетку је сажето представљена методологија, приказом и образлагањем најважнијих идентификованих корака. У наставку је описано подручје истраживања које је коришћено за тестирање и примену предложене методологије. Представљени су коришћени скупови података, заједно са описом и резултатима спроведене припреме и иницијалне анализе података. Поглавље се завршава навођењем софтверске подршке на коју се предложена методологија ослања.

4.1 Сажетак предложене методологије

Главни циљ ове дисертације је развој методологије за картирање шумске вегетације на основу података сателитског осматрања Земље. У фокусу предложене методологије су надзиране непараметарске методе класификације које спадају у групу техника машинског учења. Методологија се у основи ослања на стандардну процедуру издвајања тематских информација на основу података сателитских опажања, при чему су кораци стандардне процедуре анализирани и прилагођени специфичностима картирања шумске вегетације.

Развој методологије обухвата:

- тестирање применљивости различитих техника машинског учења за креирање класификационог модела за картирање шумске вегетације;
- анализу различитих репрезентација података ради наглашавања просторних и временских компоненти у класификационом моделу и
- примену различитих техника за креирање оптималног скупа података и атрибута од значаја за квалитетно картирање шумске вегетације.

Предложена методологија се састоји од следећих корака (**Слика 10.**):

- 1) дефиниција проблема/циљева;
- 2) прикупљање и припрема података;
- 3) одређивање оптималних параметара методологије;
 - репрезентација података (временска компонента и просторни аспекти);
 - креирање и избор оптималног скупа атрибута;
 - балансирање података;
- 4) креирање модела – примена техника машинског учења;
- 5) валидација креираних модела;
- 6) изналажење и примена оптималног модела.

Проблем картирања типова шумске вегетације може бити формулисан као класификациони проблем, где типови шумске вегетације представљају излазе креираног класификационог модела. У том смислу, дефиниција проблема и циљева подразумева дефинисање оптималне класификационе шеме, тј. идентификацију оних типова шумске вегетације које је потребно/могуће картирати на подручју од интереса. Оптимална класификациона шема мора респектовати доминантне класе (типове шумске вегетације) на подручју истраживања, али и разматрати реалне могућности њихове идентификације. Ова два захтева су често опречна и зависе од расположивих података. Због тога је оптимална класификациона шема и резултат компромиса.

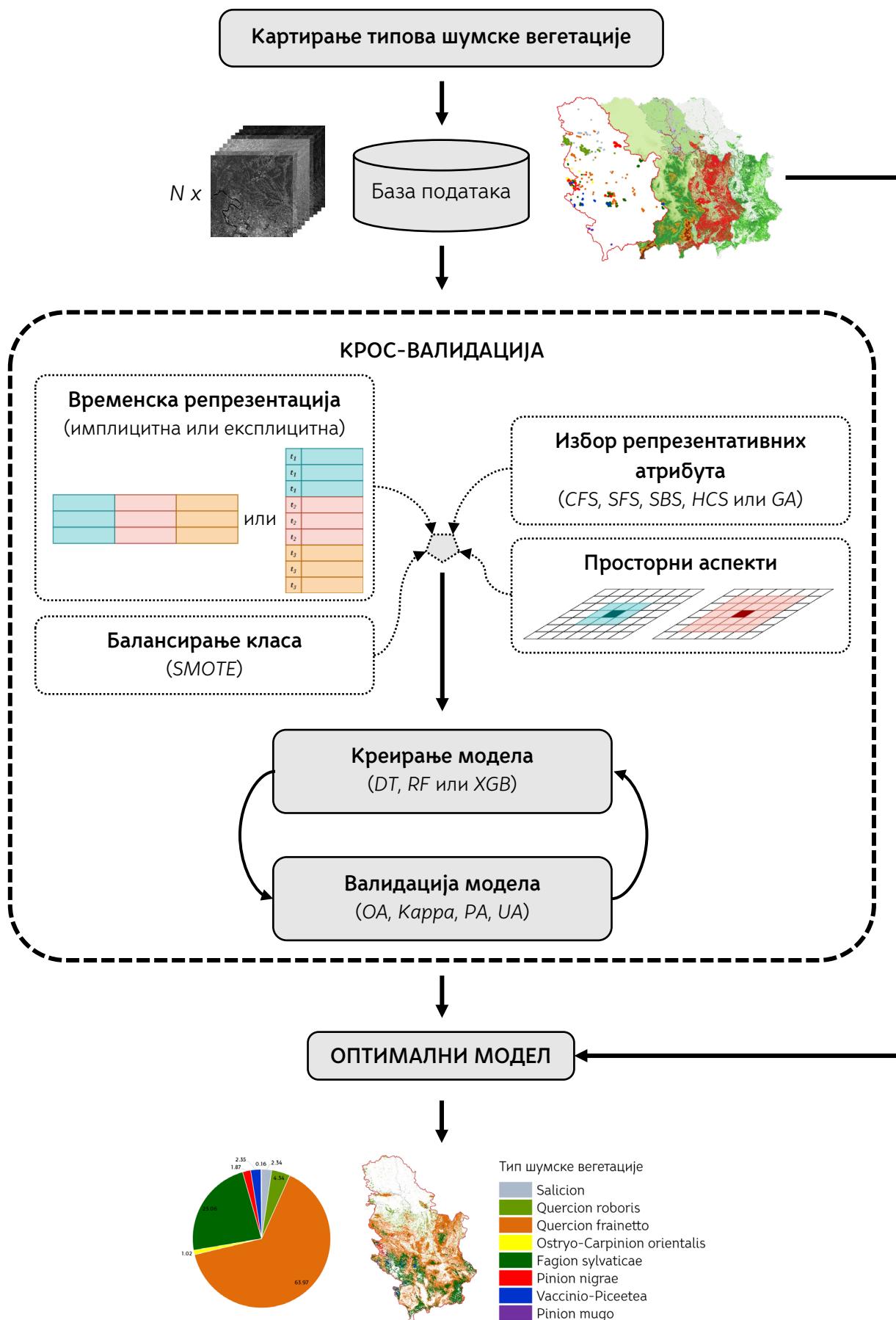
Распрострањеност шумске вегетације на неком подручју је последица великог броја различитих утицаја. Због тога је потребно обезбедити скуп података који ће на што свеобухватнији начин описати ове утицаје и нагласити специфичности различитих типова шумске вегетације. Окосницу таквог скупа података чини серија снимака сателитског опажања Земље која може бити допуњена и бројним додатним просторним подацима (нпр. подаци висинске представе терена, карта земљишног покривача итд.).

Како ови подаци често имају различите карактеристике, долазе из различитих извора, налазе се у различитим форматима и пројекцијама, на почетку је потребно извршити њихово усаглашавање и смештање у одговарајућу просторну базу података. Ово је предвиђено у другом кораку методологије.

Након просторног преклапања тренинг полигона (полигона са познатим класама шумске вегетације) и обезбеђеног скупа података (сателитска опажања и додатни просторни подаци), трећи корак подразумева одређивање скупа оптималних параметара методологије. Фокус је на више аспеката: репрезентацију података, избор оптималног скупа атрибута и балансирање скупа података. Врши се разматрање различитих могућности и начина репрезентације временске компоненте, као и наглашавање просторних аспеката у подацима. Истовремено се врши и одабир скупа најинформативнијих атрибута. За те потребе, на располагању је већи број различитих техника, а у оквиру овог истраживања је примењено и анализирано пет различитих приступа за избор репрезентативних атрибута (избор атрибута заснован на корелацији *[CFS]*, секвенцијална претрага унапред *[SFS]*, секвенцијална претрага уназад *[SBS]*, претрага планинарењем *[HCS]*, избор атрибута применом генетских алгоритама *[GA]*). Ако се ради о скупу података где нису приближно подједнако заступљене све класе, у овој фази методологије се врши и балансирање скупа података применом *SMOTE* (енгл. *Synthetic Minority Oversampling TEchnique*) алгорита.

Наредне фазе укључују креирање и валидацију већег броја класификационих модела, са циљем проналажења оптималног. Ово подразумева примену конкретне технике машинског учења, где су у оквиру овог истраживања тестиране три технике: стабло одлучивања *[DT]*, случајна шума *[RF]* и градијентно појачавање *[XGB]*. Након тога следи валидација модела која је у оквиру овог истраживања вршена применом петоструке крос-валидације са изостављањем комплетних опажања и рачунањем квантитативних мера за оцену квалитета класификације (укупна тачност *[OA]*, капа коефицијент слагања *[Kappa]*, произвођачева тачност *[PA]* и корисникова тачност *[UA]*). Као резултат ове фазе је одређен оптималан скуп параметара методологије.

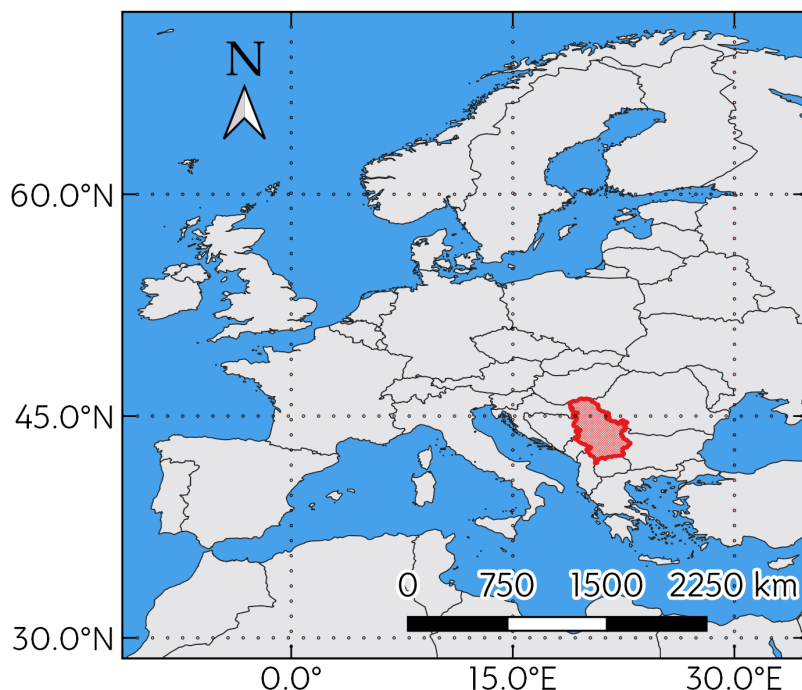
Последња фаза подразумева идентификацију и практичну употребу оптималног модела. На овај начин ће се извршити одређивање карте типова шумске вегетације на комплетном подручју истраживања, чиме се испуњава на почетку дефинисан циљ предложене методологије.



Слика 10. Графичка илустрација корака предложене методологије за картирање типова шумске вегетације

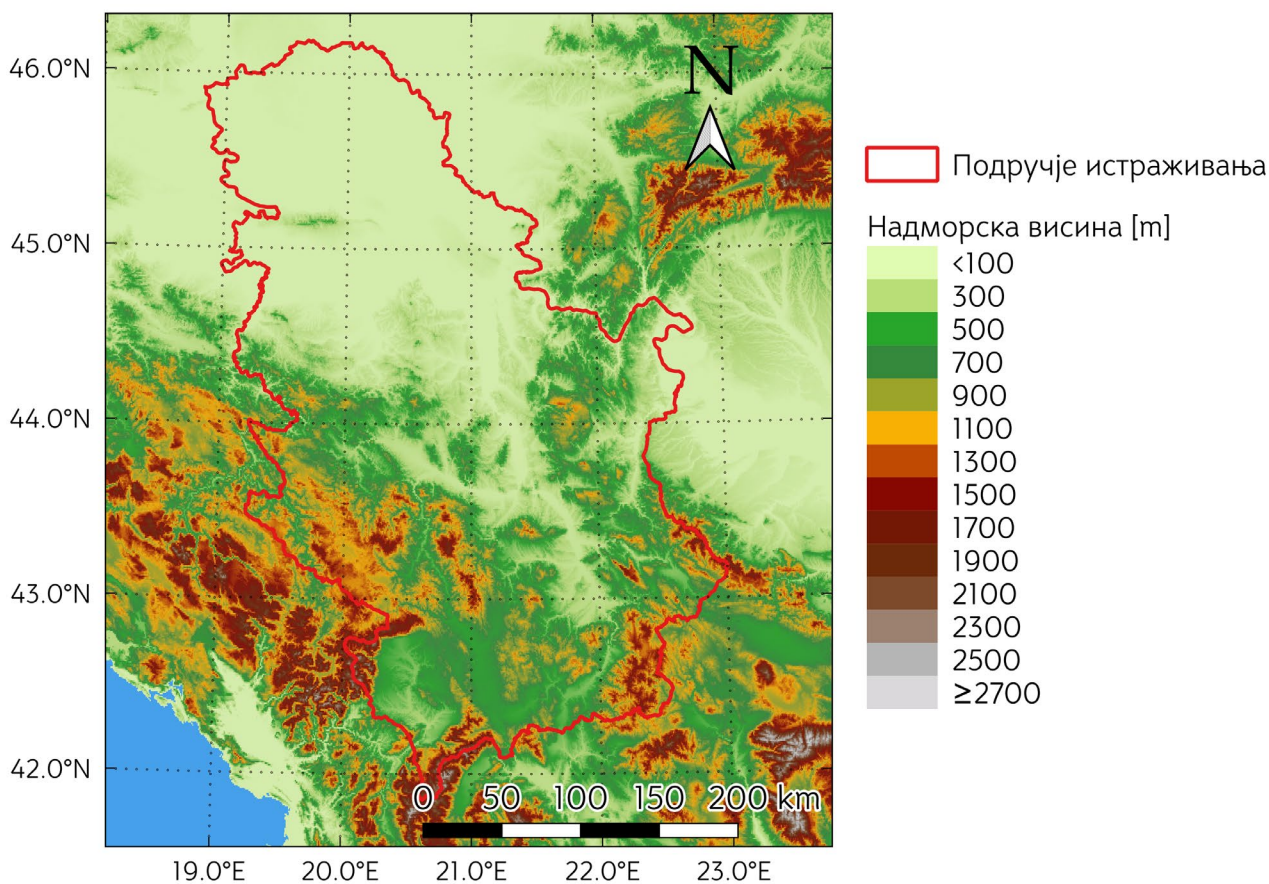
4.2 Подручје истраживања

Подручје истраживања обухвата целокупну територију Републике Србије. У питању је континентална земља југоисточне Европе која се налази у центру Балканског полуострва (Слика 11.). Укупна површина подручја истраживања износи 88 361 km².



Слика 11. Локација подручја истраживања – Република Србија

Србија има разнолик рељеф (Слика 12.). Северни део одликују низијски предели Панонске низије, где надморска висина углавном не прелази 200 m. Изузетак су две ниске планине, Фрушка гора (539 m) и Вршачки брег (641 m), као једина планинска подручја на северу. Надморска висина расте како се иде јужније. У централном делу доминирају брдовита подручја, која на југу земље прерастају у планине са више од 30 врхова изнад 2000 m надморске висине. Влажна континентална клима преовладава највећим делом земље, док је на вишим планинама присутна алпска клима. Просечна годишња количина падавина износи 896 mm.



Слика 12. Рељеф подручја истраживања

Земљишни покривач Србије је у великој мери предодређен рељефом, где обрадиво земљиште покрива највећи део Панонске низије, као и низијска подручја у околини већих река. Шумски и травнати земљишни покривач је најчешће присутан у осталим деловима земље, а на појединим подручјима постоје и мочварна и гола земљишта. Шуме покривају више од трећине Србије, при чему су најчешће присутне врсте храста, букве, бора и јеле (Банковић и др., 2009b). Према Bohn и др. (2000), у Србији су најчешће присутне термофилне мешовите листопадне широколисне шуме (шуме у којима доминирају *Quercus cerris*, *Q. frainetto*, *Q. petraea*, *Q. pubescens*, *Carpinus orientalis*, *Ostrya carpinifolia*) и мезофитне широколисне листопадне шуме (F.3, F5 – шуме у којима доминирају *Fagus sylvatica*, *F. moesiaca*, *Carpinus betulus*). Поред њих, значајна подручја покривају мезофилне четинарске шуме (D.4, D.5, D.6 – шуме у којима доминирају *Abies alba*, *Picea abies*, *P. omorika*, *Pinus sylvestris*, *Pinus peuce*), ксерофилне четинарске шуме и шикаре (K.1 – шуме у којима доминирају *P. nigra agg*, *P. heldreichii*), шумске степе (L.1, L.2 – шуме у којима доминирају *Quercus pubescens*, *Q. robur*, *Q. pedunculiflora*, *Acer tataricum*) и алувијалне шуме меког дрвета (U.5 – шуме у којима доминирају *Populus nigra*, *P. alba*, *Salix alba*).

4.3 Подаци

За дефинисано подручју истраживања је формирана база података која што боље описује релевантне типове шумске вегетације. То је подразумевало креирање репрезентативних подручја (полигона) типова шумске вегетације и обезбеђивање скупова података који ће послужити као атрибути у класификационом моделу. Главницу скупа података атрибута чини серија снимака сателитског опажања Земље, а којој су придодати и допунски скупови података. Овако обезбеђен скуп података је послужио за тестирање предложене методологије, а затим и израду карте типова шумске вегетације на целокупном подручју истраживања. У наредним поглављима је сваки од ових скупова података детаљније описан.

4.3.1 Полигони типова шумске вегетације

Као основни извор за одређивање репрезентативних подручја типова шумске вегетације коришћена је База података о распрострањености потенцијално угрожених врста и станишта Србије (Лакушић, 2019) . У питању је база података која је настала као резултат опсежних теренских кампања током више протеклих сезона. Просторне локације типова шумске вегетације су издвојене из базе података, а затим је у њиховом окружењу од око 500 m извршена мануелна векторизација полигона шумске вегетације. Векторизација је рађена помоћу *Google Earth* софтвера и сателитских снимака високе и веома високе резолуције. Полигони су креирани коришћењем довољно великог увећања да би ти полигони били што је могуће хомогенији и да би се умањиле грешке векторизације. У случајевима где није била могућа једнозначна идентификација, коришћена је серија историјских сателитских снимака из различитих периода и годишњих доба. За ово истраживање идентификовано је укупно осам типова шумске вегетације, који одговарају главним класама шумске вегетације присутним на подручју истраживања. Резултујући скуп података се састоји од 743 полигона који представљају седам шумских класа и једну класу жбунасте вегетације. Ради лакше интерпретације резултата и дискусије, свим класама су додељена и асоцијативна неформална фитосоциолошка имена, без синтаксономског значења. Истовремено су за сваку идентификовану класу одређени и одговарајући кодови дефинисани националним правилником о стаништима Србије (Службени *Гласник РС*, 35/2010-29, 2010) (Табела 4.).

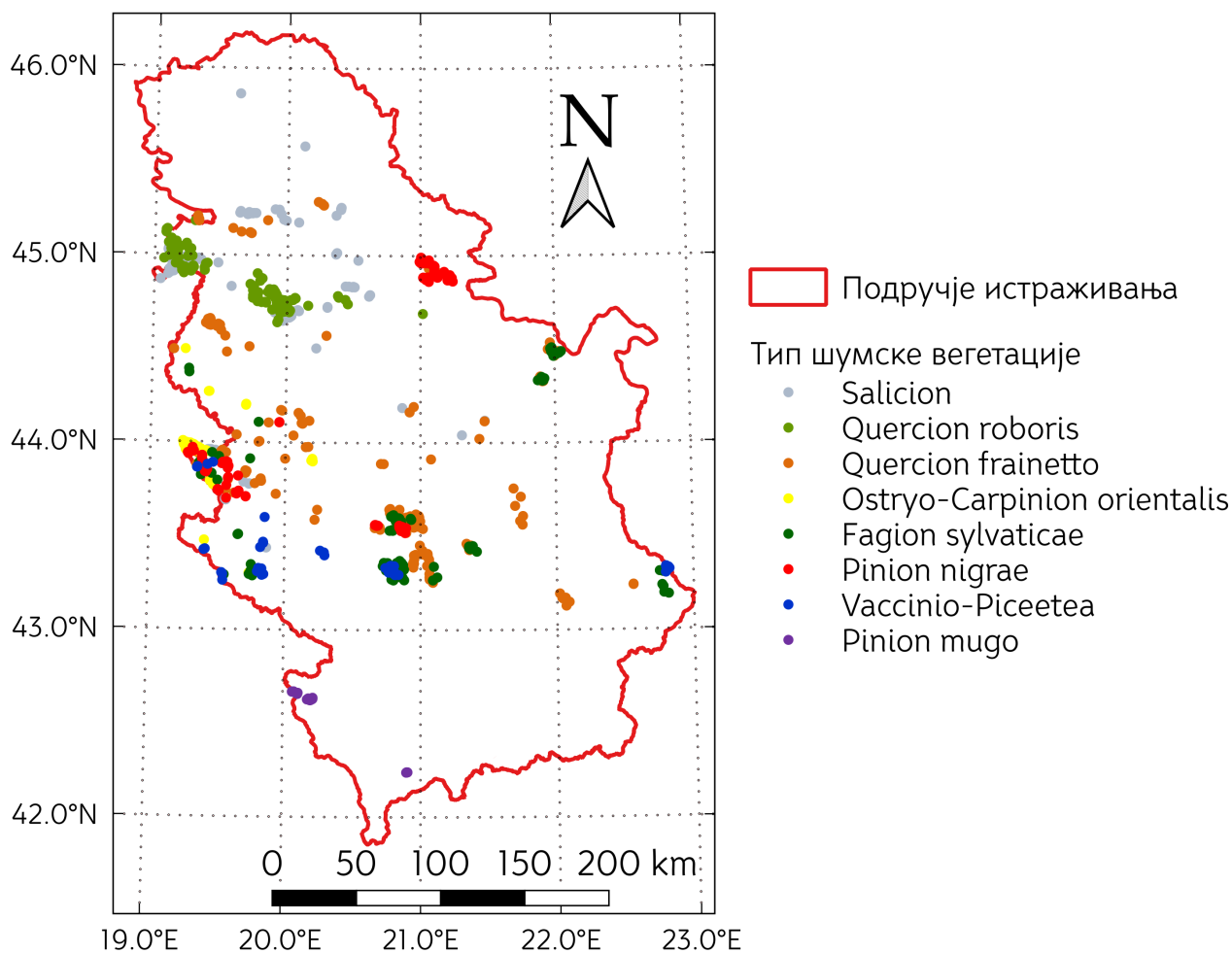
Табела 4. Дефиниција класа и њихова индивидуална заступљеност изражена кроз број креираних полигона и њихове укупне површине

<i>ID</i> класе	Тип шумске вегетације	Асоцијативно фитосоциолошко име	<i>SrbHab</i> ознака	Број полигона	Површина [km ²]
1	Хигрофилне шуме врба и топола	<i>Salicion</i>	A1.1; A1.2	123	9.82
2	Хигрофилне шуме лужњака и пољског јасена	<i>Quercion roboris</i>	A1.3; A1.4	120	36.99
3	Ксерофилне шуме храстова	<i>Quercion frainetto</i>	A2.1; A2.2; A2.3; A2.4; A2.5; A2.6; A2.7; A2.8	196	14.35
4	Ксерофилне шуме белограбића и црног граба	<i>Ostryo-Carpinion orientalis</i>	A2.9; A2.A	40	1.73
5	Мезофилне шуме букве	<i>Fagion sylvaticae</i>	A3.2; A3.3	93	10.55
6	Термофилне четинарске шуме	<i>Pinion nigrae</i>	A5.1	96	4.51
7	Фригорифилне четинарске шуме	<i>Vaccinio-Piceetea</i>	A6.1; A6.2; A6.3	50	4.54
8	Шибљаци бора кривуља	<i>Pinion mugo</i>	B6.1	25	0.39

**SrbHab* ознака – ознака према националном правилнику о стаништима РС

Просторни распоред полигона типова шумске вегетације се не може окарактерисати као равномеран (**Слика 13.**). Због наведене процедуре креирања, локације полигона су у великој мери ограничене доступним подацима у Бази података о распрострањености потенцијално угрожених врста и станишта Србије. Специфично је и то што северни равничарски део подручја истраживања има само мали проценат под шумским покривачем и то са само неким типовима шумске вегетације. Додатно, неки типови шумске вегетације су присутни само у одређеним деловима подручја истраживања. Зато није било могуће креирати полигоне типова шумске вегетације са равномерном просторном дистрибуцијом, нити са распоредом који би посебно уважавао низијска и планинска подручја. Због тога је присутан већи број група (кластера) креираних

полигона типова шумске вегетације, при чему ипак не постоји класа која је представљена само једним кластером полигона.



Слика 13. Просторни распоред креираних полигона типова шумске вегетације

4.3.2 Сателитска опажања мисије *Sentinel-2*

Сателитска мисија *Sentinel-2* је релативно нова сателитска мисија Коперникус програма који је развила Европска свемирска агенција. Констелацију чине два идентична поларно-орбитална сателита, у заједничкој Сунцу-синхроној орбити, међусобно фазно померени за 180°. Први сателит, *Sentinel-2A* лансиран је јуна 2015. године, а констелација је комплетирана када је *Sentinel-2B* лансиран марта 2017. године. Оба сателита су опремљена мултиспектралним скенером који прикупља податке у 13 спектралних канала - 4 канала са просторном резолуцијом 10 m, 6 канала просторне резолуције 20 m и преостала 3 канала са просторном резолуцијом од 60 m (Табела 5.). Овакве карактеристике обезбеђују веома кратко време поновног прелета у условима без облака, које на екватору износи 10 дана за појединачни сателит, односно

5 дана у комбинацији оба сателита. За подручја средњих географских ширина просечно време поновног прелета је 2-3 дана (*Drusch* и др., 2012).

Табела 5. Карактеристике мултиспектралног скенера *Sentinel-2* сателитске мисије

Канал	Просторна резолуција [m]	Централна таласна дужина [nm]	Опис канала
B01	60	443	аеросоли (<i>aerosols</i>)
B02	10	490	плави (<i>blue</i>)
B03	10	560	зелени (<i>green</i>)
B04	10	665	црвени (<i>red</i>)
B05	20	705	ивични црвени 1 (<i>red edge 1</i>)
B06	20	740	ивични црвени 2 (<i>red edge 2</i>)
B07	20	783	ивични црвени 3 (<i>red edge 3</i>)
B08	10	842	блиски инфрацрвени (<i>near infrared</i>)
B08a	20	865	уски блиски инфрацрвени (<i>narrow near infrared</i>)
B09	60	945	испаривање воде (<i>water vapor</i>)
B10	60	1380	цирус (<i>cirrus</i>)
B11	20	1610	краткоталасни инфрацрвени 1 (<i>short-wave infrared 1</i>)
B12	20	2190	краткоталасни инфрацрвени 2 (<i>short-wave infrared 2</i>)

У оквиру овог истраживања коришћени су *Sentinel-2* производи нивоа обраде *Level-2A*, који представљају вредности рефлектансе који се односе на подножје атмосфере (енгл. *Bottom-Of-Atmosphere* – *BOA*) и које су кориговане за утицаје рељефа (*Müller-Wilm* и др., 2013). Подаци овог нивоа обраде су на располагању као грануле величине 100x100 km, као групе канала ресамповане на 10 m, 20 m и 60 m просторне резолуције. Поред тога, подаци се дистрибуирају са придруженом картом класификације сцене (енгл. *Scene Classification map* – *SCL*). *SCL* карта се израђује аутоматским процедурама и

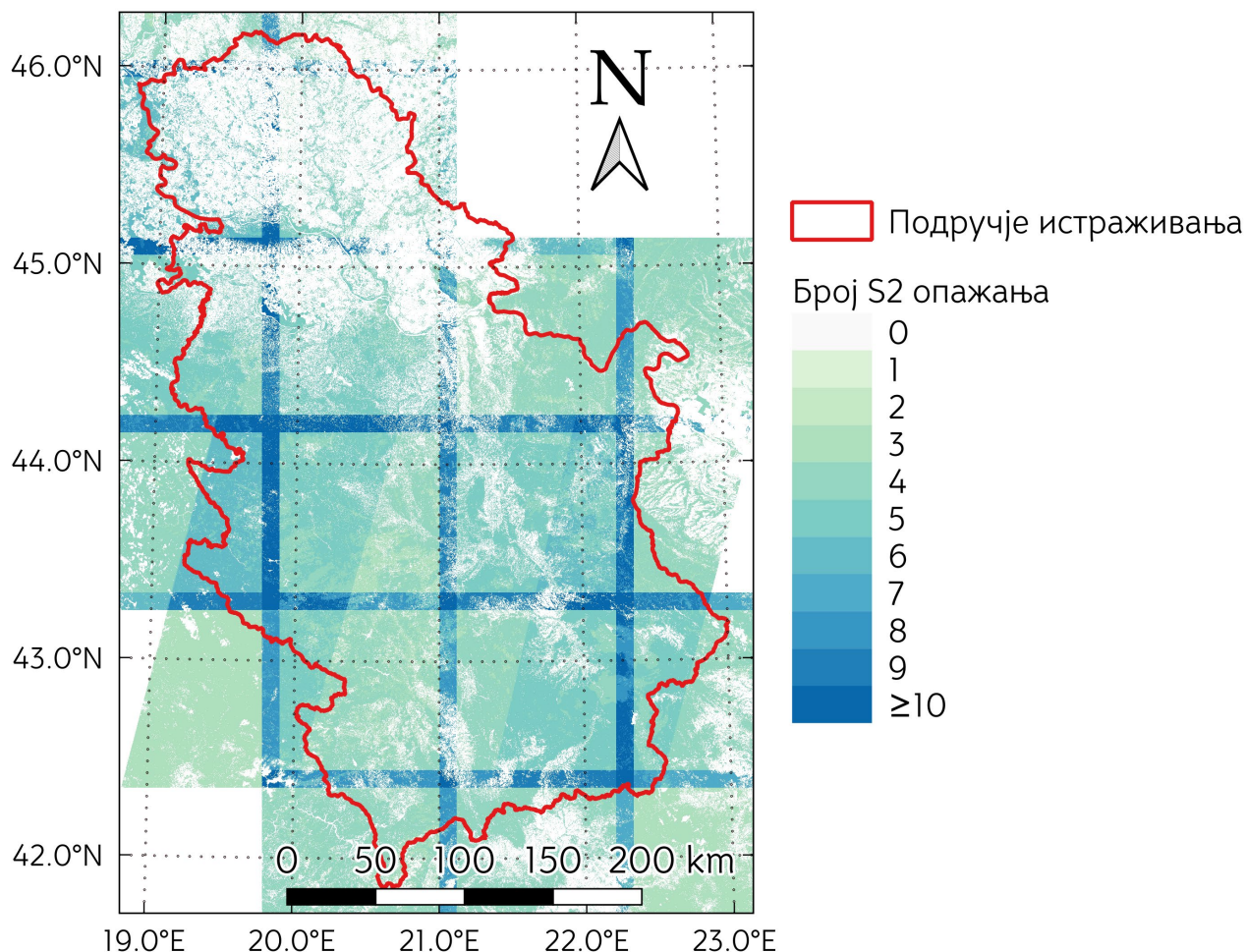
садржи неколико основних класа: облаци, сенке облака, вегетација, голо земљиште/пустиња, вода, снег итд. (Müller-Wilm и др., 2013). У оквиру овог истраживања коришћени су сви 20-метарски ресампловани канали (B02, B03, B04, B05, B06, B07, B08a, B11 и B12) заједно са припадајућом SCL картом.

Преузето је 89 Sentinel-2 гранула из 2019. године користећи Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). Грануле су одабране тако да имају мање од 20% присутне облачности и тако да је сваки део подручја истраживања покривен са најмање једном гранулом у сваком од следећих временских интервала: 1) јун – јул, 2) август, 3) септембар – октобар (Слика 14.). Месеци пре јуна и након октобра нису разматрани због велике количине облачности и појаве снежног покривача, посебно изражених на вишим надморским висинама. Сви типови шумске вегетације су у јуну већ у пуној разлистаности, која се затим различито мења како су се приближава јесен и зима. Претпоставка је да ова три временска интервала довољно наглашавају разлике у фенолошким обрасцима сваког типа шумске вегетације и да ће класификациони модел успети да препозна и искористи ове обрасце ради њиховог квалитетног разликовања.



Слика 14. Временске одреднице одабраних S2 гранула

Резултујући скуп *Sentinel-2* гранула покрива комплетно подручје истраживања са више опажања без облачности. Ако се филтрирају само опажања која су у класификацији сцене препозната као вегетација (*SCL* класа 4), може се видети да је на појединим деловима подручја на располагању преко 10 опажања (**Слика 15.**).



Слика 15. Просторни распоред расположивих S2 опажања вегетације на подручју истраживања

4.3.3 Допунски извори података

Као допунски извори података су посматрани скупови података који су коришћени у класификационом моделу, а који се не односе на сателитска опажања Земље. То обухвата *EU-DEM* податке висинске представе терена, карту основног земљишног покривача Републике Србије и Коперникус паневропске слојеве високе резолуције о шумама. Сваки од ових допунских скупова података је детаљније описан.

4.3.3.1 EU-DEM подаци висинске представе терена

EU-DEM представља дигитални модел површи терена (енгл. *Digital Surface Model – DSM*) који покрива подручја држава-чланица Европског економског простора (енгл. *European Economic Area – EEA*) и држава које сарађују са њима. Дигитални модел површи представља модел површи од које су се прво одбили електромагнетски таласи приликом опажања Земље. За разлику од дигиталног модела терена (енгл. *Digital Terrain Model*), тај модел површи не односи се само на „ голу“ површ терена, већ укључује и објекте и вегетацију који су присутни на обухваћеном подручју. То је хибридни производ заснован на *SRTM*, *ASTER GDEM* и руским топографским подацима обједињен аутоматским методама (*EU-DEM Statistical Validation, 2014*). Тренутно је актуелна верзија *EU-DEM v1.1*, која представља унапређење старије верзије *EU-DEM v1.0*. Унапређење подразумева да су уклоњене грешке геолокације, већу вертикалну тачност, мању количину артефаката и већу усаглашеност продукта са *EU-HYDRO* подацима (*EU-Hydro / EU-DEM Upgrade - User Manual, 2015*). Продукт *EU-DEM v1.1* је на располагању са просторном резолуцијом од 25 m и са вертикалном тачношћу ± 7 m *RMSE*. *EU-DEM v1.1* подаци су преузети са званичног сајта Коперникус програма (<https://land.copernicus.eu/>).

4.3.3.2 Карта основног земљишног покривача Републике Србије

Карта основног земљишног покривача за територију Републике Србије израђена је у оквиру *IGIS* пројекта коришћењем *SPOT-5* сателитских снимака (*Републички геодетски завод, 2021*). Карта се односи на епоху 2011. године и на располагању је у просторној резолуцији од 10 m. Производ је у складу са *INSPIRE* спецификацијом за тему земљишног покривача (енгл. *land cover*) и садржи десет основних класа: вештачке површине, гола земљишта, пољопривредна земљишта, травнате површине, жбуње, листопадне шуме, мешовите шуме, зимзелене шуме, влажна земљишта и водене површине (*Републички геодетски завод, 2021*). Подаци су уступљени од стране Републичког геодетског завода за потребе пројекта „Израда генерализоване карте станишта Србије“, а резултати ове дисертације ће бити коришћени за идентификацију шумских станишта у оквиру наведеног пројекта.

4.3.3 Коперникус паневропски слојеви високе резолуције о шумама

У оквиру Коперникус програма за праћење Земљине површи, на располагању су паневропски слојеви високе резолуције. Ради се о подацима који пружају информације о специфичним карактеристикама земљишног покривача и допуњују карте земљишног покривача. Конкретно је коришћен паневропски слој високе резолуције о шумама који се односи на 2018. годину. Овај скуп је креиран на основу серије података *Sentinel-2* сателитске мисије и обезбеђује информације у просторној резолуцији од 10 m о покривности крошњи шумског покривача (енгл. *tree cover density*), доминантном типу лишћа (енгл. *dominant leaf type*) и типу шуме (*Copernicus Land Monitoring Service: HRL Forest 2018 Product User Manual, 2020*). За потребе овог истраживања, коришћен је слој који се односи на покривност крошњи шумског покривача. Подаци су отвореног типа и доступни за преузимање у оквиру званичног сајта Коперникус програма (<https://land.copernicus.eu/pan-european/high-resolution-layers/forests>).

4.4 Припрема и иницијална анализа података

Како улазни подаци потичу из различитих извора и имају разнолике карактеристике, на почетку је било неопходно извршити њихово усаглашавање. За основу је усвојен *Sentinel-2* растерски гريد просторне резолуције 20 m у којем се дистрибуирају подаци ове сателитске мисије, а остали скупови података су прилагођени њему. Као координатни референтни систем је коришћена државна пројекција Републике Србије (UTM пројекција, зона 34N). Просторна резолуција карте основног земљишног покривача је преведена у 20 m тако што је извршена агрегација пиксела, при чему је вредност излазног пиксела одређена као најчешћа класа агрегираних пиксела. Подаци висинске представе терена су применом билинеарног ресамплинга препројектовани и преведени у растер са просторном резолуцијом од 20 m. Усаглашени растер висинске представе терена је затим коришћен и за креирање скупа изведених карактеристика терена коришћењем *terrain* функције из *raster R* пакета (*Hijmans и др., 2021*). Коперникус паневропски слој покривности крошњи шумског покривача је преведен у просторну резолуцију од 20 m применом билинеарног ресамплинга. Затим је на основу њега извршено креирање шумске маске тј. бинарне карте за диференцирање шумског од нешумског подручја. Шумским пикселима маске су проглашени сви пиксели са покривношћу крошњи већом од 10%. Поред усаглашавања просторне резолуције и

пројекције, припрема и иницијална обрада података обухватила је и поравнање (енгл. *alignment*) растера, како би се локације и центри кореспондентних пиксела поклапали.

Векторизовани полигони типова шумске вегетације су просторно преклопљени са претходно припремљеним растерским подацима. Сви пиксели који се налазе унутар полигона шумске вегетације се сматрају независним опажањима. Издвајање спектралних вредности је рађено независно за сваку *Sentinel-2* гранулу, и то само за пикселе који су препознати као шумски и који су у припадајућој карти класификације сцене означени као пиксели вегетације (вредност $SCL = 4$). Применом описане процедуре за свако опажање на нивоу пиксела су одређени следећи атрибути: датум аквизиције, вредности *Sentinel-2* спектралних канала, информација којој класи шумске вегетације припада, придружена класа из карте основног земљишног покривача, као и скуп карактеристика издвојених на основу висинске представе терена (**Табела 6.**).

Табела 6. Скуп карактеристика издвојених за сваки пиксел подручја истраживања

Ознака атрибута	Опис атрибута	Напомена
датум	Датум регистрације <i>Sentinel-2</i> опажања	
B02		плави
B03		зелени
B04		црвени
B05		ивични црвени 1
B06	Спектрални канали <i>Sentinel-2</i> сателитске мисије	ивични црвени 2
B07		ивични црвени 3
B08a		уски блиски инфрацрвени
B11		краткоталасни инфрацрвени 1
B12		краткоталасни инфрацрвени 2
LC	Класа земљишног покривача	Преузето из карте основног земљишног покривача Републике Србије

Ознака атрибута	Опис атрибута	Напомена
DEM	Висина терена	Преузето из <i>EU-DEM</i> -а
SLOPE	Нагиб терена	Срачунато на основу <i>EU-DEM</i> -а коришћењем <i>terrain</i> функције из <i>raster R</i> пакета
ASPECT	Експозиција (аспект) терена	
TPI	Индекс топографске позиције (енгл. <i>Topographic Position Index</i>)	
TRI	Индекс храпавости терена (енгл. <i>Terrain Ruggedness Index</i>)	
ROUGHNESS	Храпавост терена	
FLOWDIR	Правац отицања	

Овако креиран скуп података има експлицитно дефинисану временску компоненту и обухвата 1 231 783 опажања којима је представљено свих осам типова (класа) шумске вегетације. На основу њега је могуће формирати и скуп података са имплицитном временском дефиницијом. То је постигнуто увођењем три временска интервала: 1) јун – јул, 2) август и 3) септембар – октобар. За сваки од њих је на нивоу пиксела издвојен скуп расположивих *Sentinel-2* опажања, а интервали су у називима атрибута означени суфиксима „_06-07“, „_08“ и „_09-10“, респективно. Начин одабира *Sentinel-2* гранула значи да је бар једно опажање на располагању у сваком временском интервалу, а у случају вишеструких опажања је срачуната средња вредност рефлектансе сваког спектралног канала. Додатно, сваком пикселу су придружене и остале расположиве информације из допунских извора података које нису временски дефинисане. Резултујући скуп података са имплицитном временском дефиницијом има 196 454 опажања, што је шест пута мање опажања у односу на експлицитно дефинисану временску компоненту.

Пошто број полигона и њихове укупне површине варирају по класама типова шумске вегетације, ова диспропорција је присутна и у броју расположивих опажања по класама (**Табела 7.**). Посебно се истиче мали број опажања за класе шибљака бора кривуља (8 – *Pinion tugo*) и ксерофилних шума белобрабића и црног граба (4 – *Ostryo-Carpinion orientalis*). Поред тога, диспропорција између класа је још наглашенија приликом формирања имплицитног скупа података. То је пре свега последица великог броја вишеструких опажања у истим временским интервалима, која се тада представљају

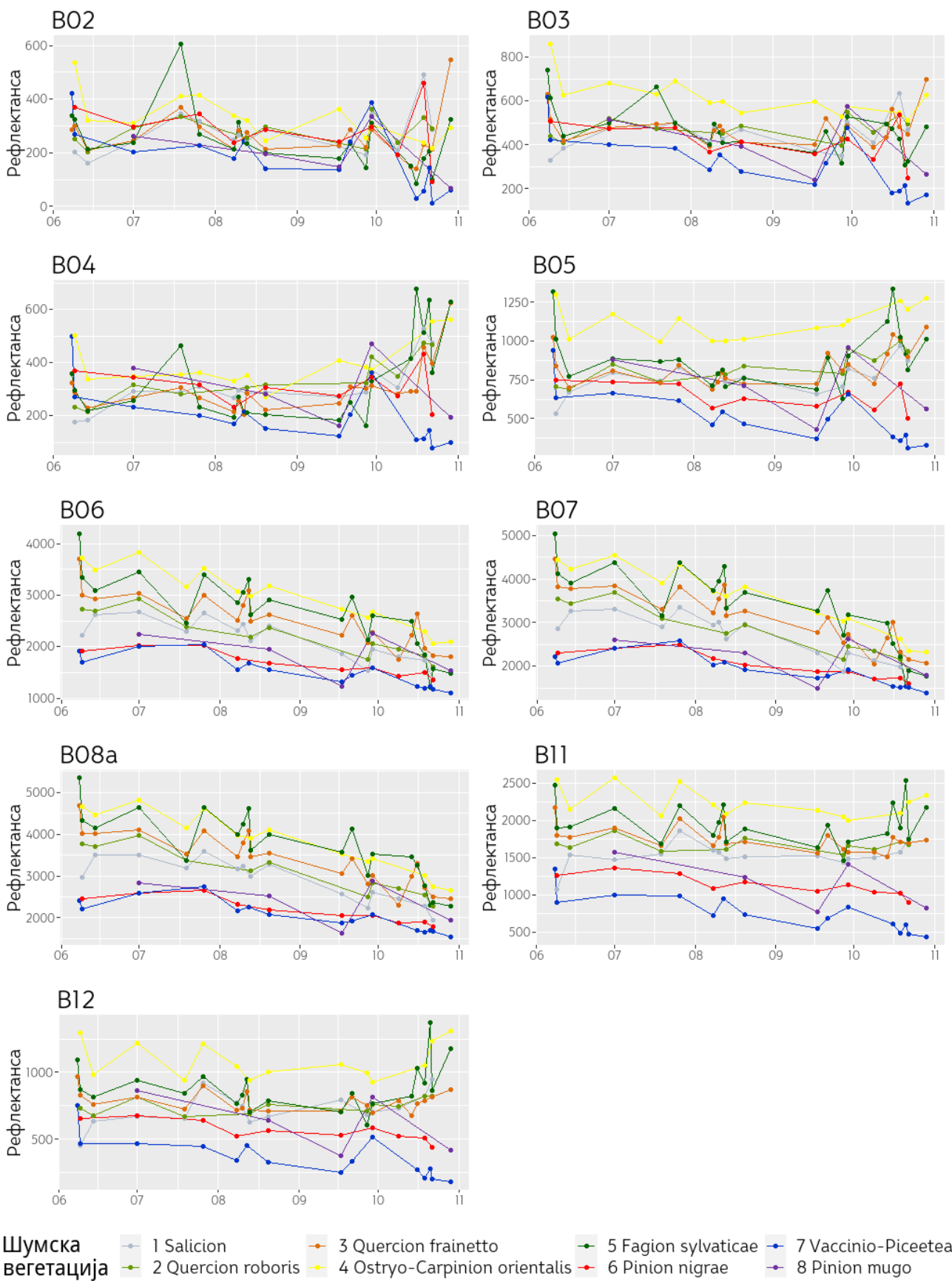
средњом вредношћу. Додатно, у случајевима где у неком интервалу није било ниједног *Sentinel-2* опажања, такво опажање је комплетно изостављено из скупа података.

Табела 7. Дефиниција класа и број расположивих опажања

ID класе	Тип шумске вегетације	Асоцијативно фитосоциолошко име	Имлицитна временска дефиниција	Експлицитна временска дефиниција
			Број пиксела	
1	Хигрофилне шуме врба и топола	Salicion	24 149	119 853
2	Хигрофилне шуме лужњака и пољског јасена	Quercion roboris	102 346	620 241
3	Ксерофилне шуме храстова	Quercion frainetto	29 910	178 807
4	Ксерофилне шуме белограбића и црног граба	Ostryo-Carpinion orientalis	3 101	20 460
5	Мезофилне шуме букве	Fagion sylvaticae	18 772	132 600
6	Термофилне четинарске шуме	Pinion nigrae	8 708	60 010
7	Фригорифилне четинарске шуме	Vaccinio-Piceetea	9 275	97 649
8	Шибљаци бора кривуља	Pinion mugo	193	2 163

Над узоркованим скуповима података је спроведена визуелна интерпретација ради прелиминарне процене значајности обезбеђених скупова података и репрезентативности усвојене класификационе шеме. То је постигнуто креирањем графика којима се илуструју везе између типова шумске вегетације и *Sentinel-2* спектралних канала, као и везе између типова шумске вегетације и висине терена.

Фенолошки спектрални обрасци су креирани и приказани графички. При томе је извршено осредњавање спектралних одзива сваке класе шумске вегетације за сваки расположиви датум. Ово је примењено над свим расположивим *Sentinel-2* спектралним каналима (**Слика 16.**), а затим је извршена њихова визуелна интерпретација.

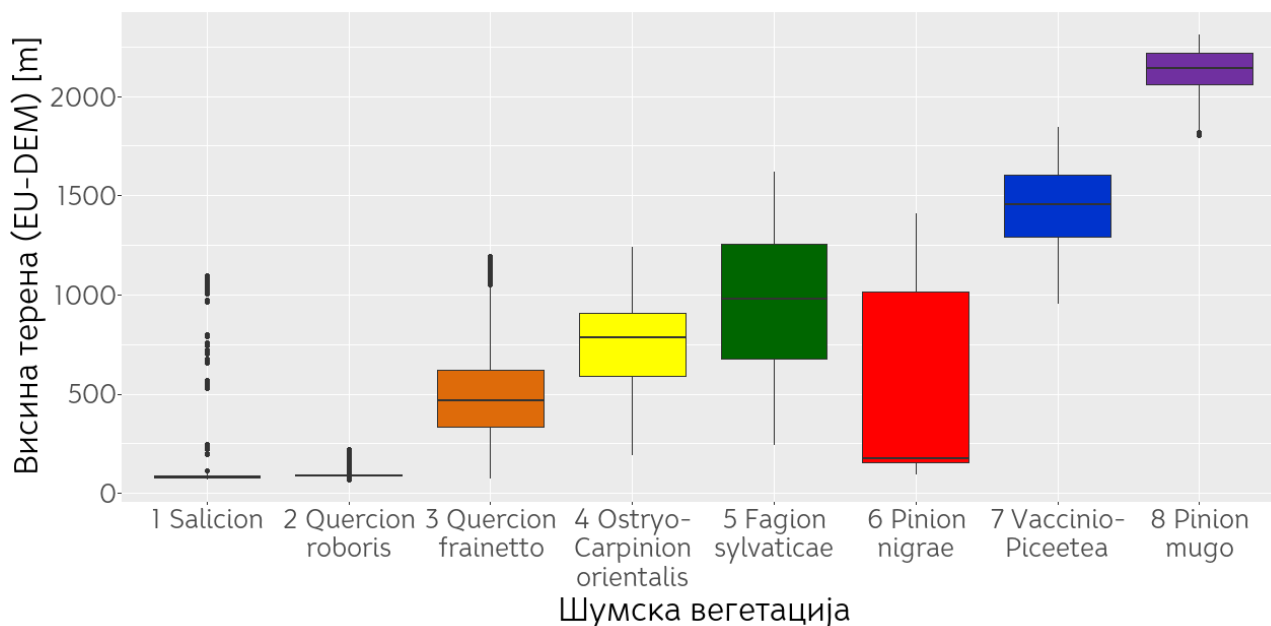


Слика 16. Графици фенолошких спектралних образаца различитих типова шумске вегетације

Графици спектралних образаца показују да спектрални одзиви варирају према типу шумске вегетације и према посматраном спектралном каналу. Ово се може протумачити као позитиван индикатор да је усвојена класификациона шема одговарајућа, тј. да постоји довољна разноликост између креираних група типова шумске вегетације које ће класификациони алгоритам моћи да препозна. Генерално се код већине спектралних канала може уочити опадајући тренд рефлектансе. Овај опадајући тренд је посебно приметан код листопадних типова шума, док је мање заступљен код зимзелених врста. Специфично се издваја и повећање вредности рефлектансе у каналима *B02*, *B03*, *B04*, *B05*, *B11* и *B12* током периода октобар – новембар, што је такође много израженије код листопадне шумске вегетације у поређењу са зимзеленом. Разликовање типова шумске вегетације је визуелно најизраженије у периоду јун – септембар, при чему и оно у великој мери зависи од посматраног канала и типа шумске вегетације. Појединачно је највећа могућност разликовања у блиским инфрацрвеним каналима (*B05*, *B06*, *B07* и *B08a*) и средњим инфрацрвеним каналима (*B11* и *B12*), где се лако могу уочити листопадне и зимзелене групе шумске вегетације. Појединачни типови шумске вегетације такође показују одређене специфичне спектралне обрасце на одређеним спектралним каналима. Карактеристичан пример су мезофилне шуме букве (5 – *Fagion sylvaticae*) у спектралним каналима *B02* и *B04* и ксерофилне шуме хрстова (3 – *Ostryo-Carpinion orientalis*) на у каналима *B03*, *B11* и *B12*.

Веза између типова шумске вегетације и висине терена је анализирана креирањем кутијастог дијаграма (енгл. *box-plot diagram*) (Слика 17.). Креирани кутијаста дијаграм истиче да постоји значајно висинско раслојавање различитих класа шумске вегетације. На дијаграму се уочава како се висински издвајају хигрофилне шуме врба и топола (1 – *Salicion*), хигрофилне шуме лужњака и пољског јасена (2 – *Quercion roboris*), фригорифилне четинарске шуме (7 – *Vaccinio-Piceetea*) и шибљаци бора кривуља (8 – *Pinion tugo*). Дијаграм приказује и да постоје релативно широки висински опсеги код већине класа шумске вегетације. Изузетак су класе шуме врба и топола (1 – *Salicion*) и хигрофилне шуме лужњака и пољског јасена (2 – *Quercion roboris*) које имају знатно уже опсеге од осталих класа, али који су праћени и већим бројем опажања која из њих значајно одскачу. Може се уочити и да постоје извесна преклапања између висинских опсега за поједине типове шумске вегетације. То се посебно истиче за термофилне четинарске шуме (6 – *Pinion nigrae*) које се преклапају са ксерофилним шумама

хрстова (3 – *Quercion frainetto*), ксерофилним шумама белограбића и црног граба (4 – *Ostryo-Carpinion orientalis*) и мезофилним шумама букве (5 – *Fagion sylvaticae*).



Слика 17. Кутијаста дијаграм типова шумске вегетације и вредности висине терена

Добијени изглед кутијастог дијаграма је у великој мери последица усвојене класификационе шеме типова шумске вегетације, чија станишта у великој мери висински нису ограничена. Ипак, треба нагласити и да је кутијаста дијаграм условљен и распоредом полигона шумске вегетације, где се већи број односи на северне равничарске делове подручја истраживања. Неретко је посредни планско пошумљавање, где су у појединим случајевима пошумљене врсте које примарно и нису карактеристичне за таква подручја. Типичан пример су четинарске шуме на подручју Војводине, где нпр. већи број полигона шумске вегетације потиче управо из Делиблатске пешчаре, што директно утиче на висински опсег овог типа шума.

4.5 Софтверска подршка

Комплетно прикупљање, обрада, анализа и презентација података је извршена коришћењем софтвера отвореног кода. Највећи део је урађен у програмском језику *R* (*R Core Team, 2020*) коришћењем већег броја различитих библиотека, а за картографску обраду је коришћен софтвер *QGIS* (*QGIS Development Team, 2021*). Коришћено је и *RStudio* интегрисано развојно окружење (*RStudio Team, 2021*) ради лакшег и ефикаснијег кодирања у *R*-у. Предложена методологија је имплементирана и

тестирана кроз серију *R* програмских скрипти. Најзначајније коришћене *R* библиотеке и њихова примена приказани су у наставку (**Табела 8.**).

Табела 8. Најзначајније *R* библиотеке коришћене за потребе овог истраживања

Назив библиотеке	Најзначајнија примена
<i>raster</i> (Hijmans и др., 2021), <i>rgdal</i> (Bivand и др., 2021), <i>rgeos</i> (Bivand и др., 2020), <i>sp</i> (Pebesma и др., 2021), <i>spex</i> (Sumner, 2020)	Учитавање, манипулација, обрада и анализа просторних података
<i>rpart</i> (Therneau и др., 2019)	Имплементација алгорита стабла одлучивања
<i>ranger</i> (Wright и др., 2020)	Имплементација алгорита случајне шуме
<i>xgboost</i> (Chen и др., 2021)	Имплементација алгорита градијентног појачавања
<i>FSelector</i> (Romanski и др., 2021)	Различите методе избора информативних атрибута
<i>GA</i> (Scrucca, 2021)	Имплементација генетских алгоритама (коришћена за селекцију информативних атрибута)
<i>smotefamily</i> (Siriseriwan, 2019)	Балансирање података применом <i>SMOTE</i> алгорита
<i>data.table</i> (Dowle и др., 2021), <i>dplyr</i> (Wickham и др., 2021b), <i>groupdata2</i> (Olsen, 2021), <i>stringr</i> (Wickham и RStudio, 2019)	Манипулација, обрада и анализа података
<i>ggplot2</i> (Wickham и др., 2021a), <i>ggpubr</i> (Kassambara, 2020)	Визуелизација података и креирање графика

У оквиру креираних *R* скрипти, библиотека *raster* (Hijmans и др., 2021) коришћена је приликом рада са растерским просторним подацима, односно библиотеке *rgdal* (Bivand и др., 2021), *rgeos* (Bivand и др., 2020), *sp* (Pebesma и др., 2021) и *spex* (Sumner, 2020) за манипулацију векторским просторним подацима. Додатно, ради смањења меморијских и процесорских захтева, помоћу ових библиотека је вршена сегментација подручја истраживања на мање целине. То обухвата креирање независних делова просторним исецањем, појединачна обрада сваког од њих и затим мозаиковање резултата у коначну карту.

За сваку технику машинског учења је коришћена посебна *R* библиотека, иако постоје и библиотеке у оквиру којих су све три технике машинског учења имплементирани на једном месту. Библиотеке су одабране на основу њихове популарности и присутности у литератури. Тако је коришћена библиотека *rpart* (Therneau и др., 2019) за

имплементацију стабла одлучивања, *ranger* (Wright и др., 2020) за имплементацију случајне шуме и *xgboost* (Chen и др., 2021) за имплементацију градијентног појачавања. Претпоставка је да су специјализоване библиотеке више оптимизоване, да обезбеђују ефикасније алгоритме са више могућности и параметара, као и да могу да се носе са великим количинама података. Мана овог приступа што је често неопходно прилагођавање и посебна припрема података и параметара, како би се одговорило специфичним захтевима сваке библиотеке.

FSelector R библиотека (Romanski и др., 2021) обезбеђује имплементацију алгоритма избора атрибута базираног на корелацији (CFS) и секвенцијалне претраге атрибута. Пошто ова библиотека не обезбеђује имплементацију генетских алгоритама, за те потребе је коришћена *R* библиотека *GA* (Scrucca, 2021). Обе библиотеке обезбеђују само механизам одабира/избацивања атрибута из скупа репрезентативних (код омотач-метода за избор репрезентативних атрибута), али се од корисника очекује да дефинише процедуру одређивања метрике која ће се користити за те потребе.

Балансирање података применом *SMOTE* алгоритма је извршено коришћењем основне верзије *SMOTE* алгоритма имплементиране у оквиру *smotefamily* (Siriseriwan, 2019) библиотеке. Пошто је алгоритам иницијално предвиђен за бинарне небалансиране скупове података, извршено је његово прилагођавање тако да може да се искористи и над скупом података са већим бројем циљних класа. То подразумева примену наведеног алгоритма кроз већи број сукцесивних пролаза (балансирања). Балансирање класа је реализовано само креирањем вештачких примерака мањинских класа без подузорковања примерака већинских класа. У сваком пролазу, скупу података се независно додају примерци само најмање заступљене класе, а ова процедура се итеративно понавља све док релативна разлика у заступљености свих класа није мања од дефинисаног прага. Као праг балансираности је одабрана вредност од 30%, а додатно је уведен и услов како би се очувао редослед заступљености класа из почетног скупа података. То практично значи да су релативни односи бројности класа очувани тако да и даље осликавају поредак из узорка, али су разлике у бројности мање заступљених и најзаступљеније класе сведене на максимално 30% броја примерака у најзаступљенијој класи. На овај начин су превазиђене екстремне разлике у заступљености класа које су присутне у почетном скупу података, али се не врши апсолутно изједначавање бројности свих класа. У комбинацији са нешто већом

вредности прага балансираности то значи да резултујући скуп података и даље донекле садржи небалансираност, али је претпоставка да се ради о небалансираности класа која постоји и у стварном свету.

Библиотеке *data.table* (Dowle и др., 2021), *dplyr* (Wickham и др., 2021b) и *groupdata2* (Olsen, 2021) су коришћене за ефикасну селекцију, придруживање и претраживање скупова података, агрегацију података по одређеном критеријуму итд. У случају када се радило о подацима/атрибутима текстуалног типа, ове библиотеке су се ослањале и на *stringr* (Wickham и RStudio, 2019) библиотеку. Сви графици у оквиру дисертације су креирани користећи *ggplot2* (Wickham и др., 2021a) и *ggpubr* (Kassambara, 2020) библиотеке. Поред тога, ове библиотеке су коришћене и као механизам визуелне контроле резултата појединих корака обраде.

5

РЕЗУЛТАТИ И ДИСКУСИЈА

Тестирање предложене методологије, дискусија и добијени резултати спроведених експеримената изложени су у петом поглављу дисертације. То обухвата четири експеримента, који се надовезују један на други, а сваки је разматран по један кључни аспект методологије. Закључци на основу спроведених експеримената су коришћени за утврђивање оптималних параметара методологије. Ови параметри су послужили за креирање класификационих модела којима су креиране коначне карте шумске вегетације за комплетну територију Републике Србије. Последњи сегмент петог поглавља се бави валидацијом креираних карата независним скупом података. На овај начин се тежило непристрасној оцени квалитета креираних карата, али су и из другог угла разматрани и коментарисани поједини аспекти предложене методологије.

5.1 Експериментално одређивање оптималних параметара методологије

Експериментално одређивање оптималних параметара предложене методологије је извршено кроз серију експеримената којима су анализирани различити аспекти методологије. Утврђено је да су најважније компоненте предложене методологије:

- 1) начин репрезентовања временске компоненте у класификационом моделу,
- 2) избор скупа репрезентативних атрибута,
- 3) решавање проблема небалансираности класа у подацима и
- 4) наглашавање просторне компоненте у класификационом моделу.

Због тога је осмишљено и реализовано четири експеримента, где је у сваком нагласак био на једној од главних компоненти предложене методологије. Тестирање је вршено итеративно, а експерименти се надовезују један на други. Сваки експеримент је подразумевао креирање већег броја класификационих модела, при чему су за сваки тестирани модел одређене квантитативне мере перформанси. Анализом и поређењем мера перформанси се тежило изналажењу оптималног решења, тј. оптималних параметара предложене методологије.

Сваки усвојени приступ је подразумевао валидацију применом све три претходно описане технике машинског учења (стабло одлучивања *[DT]*, случајна шума *[RF]* и градијентно појачавање *[XGB]*). Пошто се предложена методологија ослања и на просторно-временску агрегацију, разматране су и све три усвојене функције агрегације (најчешће присутна класа *[MC]*, класа са највећом вероватноћом срачунатом простом аритметичком средином *[SM]* или геометријском средином *[GM]*). Све технике машинског учења су коришћене са параметрима подешеним тако да се креирају стабла одлучивања истих карактеристика, док је остатак параметара остављен на препорученим вредностима. Варирање параметара технике машинског учења није вршено кроз експерименте, јер је акценат овог истраживања на развоју и анализи методологије за картирање шумске вегетације као целине. Пошто се све три коришћене технике машинског учења базирају на стаблу одлучивања, претпоставља се да промена параметара неће значајно утицати на одређивање оптималних параметара методологије. Утврђено је да, без обзира на фиксирање параметара техника машинског учења, број потенцијално оптималних параметара методологије прогресивно расте кроз експерименте. Како би се број тестирања свео на разумну меру, где год је то било могуће, вршено је одбацивање оних параметара методологије који очигледно не представљају оптимално решење.

У наредним поглављима приказани су резултати спроведених експеримената, уз опсежну дискусију утврђених оптималних параметара и карактеристика предложене методологије.

5.1.1 Експеримент 1:

Укључивање временске компоненте у класификациони модел

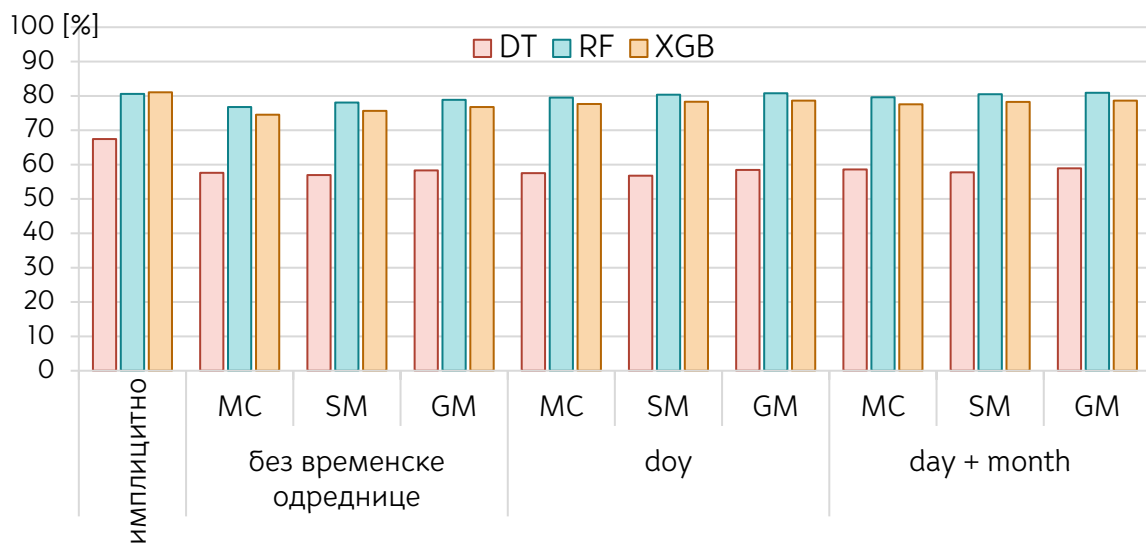
Први експеримент имао је за циљ анализу различитих приступа репрезентације временске компоненте у оквиру класификационог модела. Предложена методологија разликује два основна приступа, према томе да ли је временска компонента у класификационом моделу дефинисана имплицитно или експлицитно. Као што је то већ описано у **поглављу 3.2**, имплицитна репрезентација значи да се сваки временски тренутак/интервал представља скупом независних атрибута у класификационом моделу, а експлицитна репрезентација подразумева да се за свако опажање посебно разматра и води један или више временских атрибута (нпр. дан у години, кумулативни дан у години, датум итд.). Оба приступа користе само оригинална сателитска опажања без попуњавања рупа у подацима, што има различите последице.

Експлицитна временска дефиниција значи да је, како би се одредила класа шумске вегетације, потребно обезбедити макар једно употребљиво опажање (без облака, сенке итд.) за сваки део (пиксел) истраживаног подручја. Како се на подручју истраживања повећава број опажања без облака, за сваки део подручја истраживања на располагању је више излаза класификационог модела, а коначна класа се одређује применом просторно-временске агрегације. То омогућава да се целокупна методологија лако прилагоди било ком подручју истраживања, без обзира на њену величину. Исто тако, додавање опажања ван посматраног временског интервала додаје нове излазе и захтева понављање просторно-временске агрегације, али не подразумева обавезно креирање новог класификационог модела.

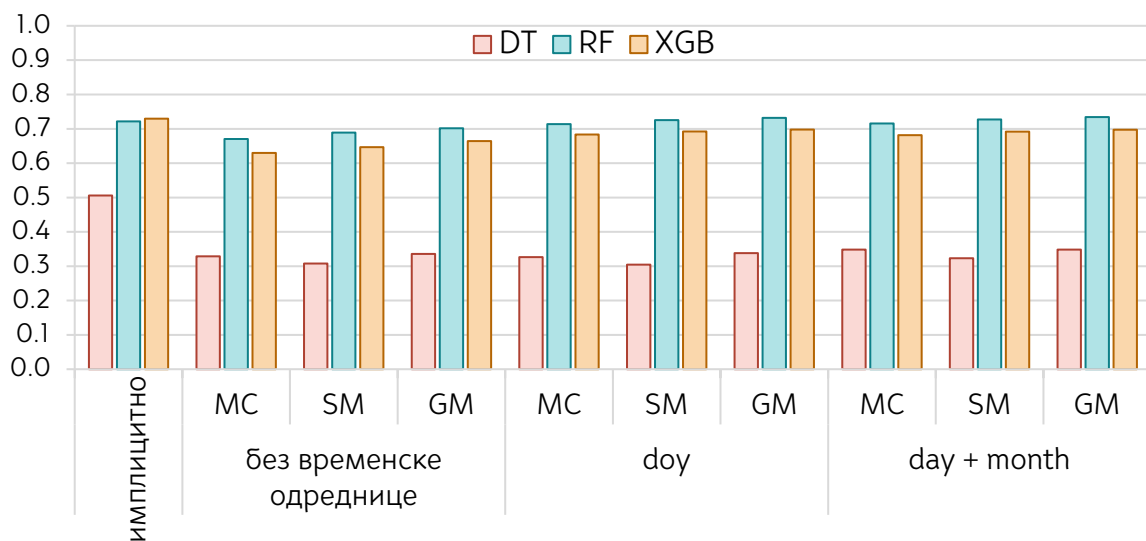
Насупрот томе, имплицитни приступ за сваки пиксел тражи по једно опажање у сваком дефинисаном временском интервалу, што се може показати као изузетно захтеван услов у појединим подручјима. Истовремено, промена (нарочито проширење) дефинисаних временских интервала утиче на атрибуте имплицитног класификационог модела, односно подразумева креирање новог класификационог модела за сваки нови скуп дефинисаних временских интервала.

Дакле, циљ овог експеримента је утврђивање оптималног начина репрезентације временске серије сателитских података у класификационом моделу. Разматран је имплицитни приступ са три временска интервала (јун – јул, август и септембар – октобар). Кроз три класификациона модела је представљена експлицитна временска

дефиниција и то: 1) без дефинисане временске одреднице, 2) увођењем атрибута за дан у години [*day*] и 3) увођењем два атрибута за дан и месец [*day + month*]. У оквиру овог експеримента коришћени су само атрибути којим се представљају спектрални канали *Sentinel-2* сателитске мисије, јер су једино они временски одређени. Претпоставка је да је изостављањем осталих атрибута из класификационог модела нагласио утицај временске компоненте. У наставку је приказано графичко поређење укупне тачности (Слика 18.) и капа коефицијената слагања свих тестираних модела (Слика 19.). Табеле са срачунатим вредностима за укупну тачност, капа коефицијент слагања, корисникову и произвођачеву тачност свих креираних модела се налазе у оквиру **прилога 1**.



Слика 18. Поређење укупне тачности различитих приступа укључивања временске компоненте



Слика 19. Поређење капа коефицијента слагања различитих приступа укључивања временске компоненте

Прегледом срачунатих метрика одмах се уочава да стабло одлучивања постиже знатно ниже мере перформанси у односу на остале две тестиране технике машинског учења. Ова разлика постоји код оба приступа респектовања временске компоненте, при чему је она израженија у случају експлицитно дефинисане временске компоненте. Ово је свакако и очекивано, јер је стабло одлучивања рудиментарнија техника машинског учења, док су преостале две тестиране технике њена надградња.

Резултати указују на то да, када се користе само подаци сателитских опажања, не постоји значајна разлика у мерама перформанси између имплицитно и експлицитно дефинисане временске компоненте. Постојеће разлике се могу сматрати занемарљивим, односно могу бити и последица коришћеног скупа података и мерне несигурности присутне у подацима. Због тога на основу тако малих разлика није могуће тврдити који приступ даје тачније резултате. Од свих тестираних варијанти највише метрике су постигнуте имплицитном временском репрезентацијом у комбинацији са техником градијентног појачавања ($OA = 81.06\%$, $Kappa = 0.73$). Експлицитно респектовање временске компоненте у незнатној мери заостаје за имплицитно дефинисаном временском компонентом. Модел са експлицитном дефиницијом који има највише метрике перформанси користи технику случајне шуме, са временском компонентом представљеном кроз дан и месец и применом агрегације геометријском средином ($OA = 80.93\%$, $Kappa = 0.73$).

Резултати показују да избор технике машинског учења нема значајан утицај у зависности од приступа за дефинисање временске компоненте. Резултати указују да се једино техника стабла одлучивања може искључити као оптимална, јер у свим тестираним комбинацијама обезбеђује знатно ниже мере перформанси у односу на преостале две технике машинског учења. Са друге стране, предност неке од преостале две технике машинског учења није једнозначна. Код имплицитне временске дефиниције су постигнуте веома сличне метрике перформанси и градијентним појачавањем и случајном шумом, где градијентно појачавање има незнатно више мере перформанси. У случају експлицитне временске дефиниције разлика између ове две технике машинског учења је нешто већа, при чему сада случајна шума надмашује градијентно појачавање. Овакви резултати значе да се не може дати коначан суд о предности ни случајне шуме, ни градијентног појачавања, већ да одговор на ово питање треба тражити кроз наредне експерименте.

Уколико се ради о експлицитном приступу, увођење временске одреднице опажања као посебних атрибута има допринос. То се посебно истиче када се користе технике случајне шуме или градијентног појачавања, где повећање мера перформанси износи и до 3% за укупну тачност и 0.05 за капа коефицијент слагања. Истовремено се не може уочити велика разлика према томе да ли се временска одредница представља кумулативним даном у години или даном и месецом, при чему модели са даном и месецом над коришћеним скупом података постижу незнатно више мере перформанси.

Што се тиче методе агрегације која се примењује код експлицитно дефинисане временске компоненте, постижу се веома сличне мере перформанси користећи све три функције агрегације. Разлика између прости и геометријске средине са временским одредницама не прелази 0.5% укупне тачности и вредности капа коефицијента изражене на две децимале су често једнаке. Агрегација најчешћом класом углавном заостаје око 1% по укупној тачности и до 0.03 по вредности капа коефицијента слагања. На основу оваквих разлика се не може дефинитивно тврдити предност неког приступа агрегације, при чему ипак делује да геометријска и проста аритметичка средина резултују вишим мерама перформанси од правила најчешће класе.

Наравно, за сагледавање потпуне слике рангирања различитих модела није довољно разматрати само укупну тачност и капа коефицијент слагања, већ је неопходно анализирати и вредности произвођачеве и корисникове тачности различитих класа шумске вегетације. Очекивано, упоређивањем табела укупне тачности, капа коефицијента слагања, произвођачеве и корисникове тачности (**Прилог 1**) може се уочити да се рангирање модела мења у зависности од тога која метрика се користи. Другим речима, не може се рећи да је неки приступ доминантан у односу на друге, већ да сваки подразумева одговарајуће компромисе. Ово сугерише да су све три тестиране технике машинског учења подложне да преферирају поједине класе у односу на остале. Специфично се издвајају изузетно ниске мере перформанси стабла одлучивања по питању мањински заступљених класа, где је класа шибљака бора кривуља (8 – *Pinion tugo*) потпуно занемарена, а ксерофилне шуме белограбића и црног граба (4 – *Ostryo-Carpinion orientalis*) су препознате само у комбинацији са имплицитном временском компонентом. Код преостале две технике машинског учења су резултати углавном слични и доминантно зависе од начина респектовања временске компоненте. Тако се

имплицитном временском дефиницијом постижу највише мере перформанси за термофилне четинарске шуме (6 – *Pinion nigrae*), док експлицитно дефинисана временска компонента са било којим временским атрибутима резултује вишим метрикама перформанси за класе ксерофилне шуме белограбића и црног граба (4 – *Ostryo-Carpinion orientalis*) и шибљаке бора кривуља (8 – *Pinion mugo*).

Табеле произвођачеве и корисникове тачности генерално указују на тренд да је висока вредност произвођачеве тачности праћена нижим вредностима корисникове тачности и обрнуто. У контексту рангирања различитих тестираних приступа, тренд се манифестује тако што нека метода често има највише вредности корисникове тачности, али истовремено је међу методама са најнижим вредностима произвођачеве тачности и обрнуто. Ово је посебно изражено за класе 1 – хигрофилне шуме врба и топола, 2 – хигрофилне шуме лужњака и пољског јасена и 3 – ксерофилне шуме храстова. У таквим ситуацијама је тешко одредити се за неки конкретни приступ, већ је избор потребно прилагодити конкретним захтевима и потребама да се преферира једна, односно друга метрика.

5.1.2 Експеримент 2:

Избор репрезентативних атрибута

У оквиру другог експеримента је тестирано више различитих техника за избор репрезентативних атрибута. На овај начин је извршено утврђивање да ли је нека од метода за избор атрибута доминантно подеснија у односу на остале. Истовремено се тежило одређивању таквог репрезентативног скупа атрибута који ће обезбедити највише мере перформанси класификационог модела, уз евентуално смањење меморијских и рачунских захтева. Анализиран је комплетан скуп расположивих атрибута, при чему је одвојено вршен избор скупа атрибута када је модел дефинисан са имплицитном и експлицитном временском компонентом. То значи да комплетан скуп атрибута садржи 35 атрибута у случају имплицитне временске дефиниције, односно 18 или 19 атрибута у случају експлицитне временске дефиниције. Примењена је метода за избор атрибута базирана на корелацији [*CFS*] из групе филтер-метода, а из групе омотач-метода је извршен избор атрибута применом генетских алгоритама [*GA*] и применом секвенцијалне методе избора атрибута (претрага унапред [*SFS*], претрага уназад [*SBS*] и претрага планинарењем [*HCS*]). Код филтер-метода није могуће узети у обзир хеуристику технике машинског учења, док су код сваке методе из

групе омотач-метода примењене све три технике машинског учења. Као мера перформанси за избор подскупа атрибута коришћена је укупна тачност класификације срачуната из двоструке крос-валидације са изостављањем комплетних локација. Скупови одабраних атрибута применом свих описаних приступа приказани су у одговарајућим табелама (Табела 9, Табела 10. и Табела 11.).

Табела 9. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – временска компонента дефинисана као дан у години (*doy*)

Атрибут	CFS	SFS			SBS			HCS			GA		
		DT	RF	XGB	DT	RF	XGB	DT	RF	XGB	DT	RF	XGB
doy					x	x	x	x	x	x	x	x	x
B02					x	x	x						
B03				x	x	x	x	x	x	x		x	
B04					x	x	x						x
B05		x			x	x	x				x	x	
B06					x	x	x	x	x	x	x	x	x
B07				x	x	x	x					x	x
B08a					x	x	x	x	x	x	x	x	
B11	x	x	x	x	x	x	x			x	x		x
B12					x	x	x		x	x		x	
LC	x		x	x	x	x	x	x	x	x	x	x	x
DEM	x	x	x	x	x	x	x	x	x	x	x	x	x
SLOPE				x	x	x	x				x		
ASPECT						x			x		x	x	
TPI	x				x								
TRI				x	x	x	x	x				x	x
ROUGHNESS					x	x	x	x	x	x		x	x
FLOWDIR					x			x	x	x			
Одабрано (од 18):	4	3	3	7	17	16	15	9	10	10	9	12	9

Табела 10. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – временска компонента дефинисана као дан и месец (*[day + month]*)

Атрибут	CFS	SFS			SBS			HCS			GA		
		DT	RF	XGB	DT	RF	XGB	DT	RF	XGB	DT	RF	XGB
day					x	x	x	x	x	x		x	x
month							x	x	x	x		x	x
B02					x	x	x						
B03					x	x	x	x	x	x		x	
B04					x		x				x		x
B05					x		x	x			x	x	x
B06					x	x		x	x	x		x	x
B07				x	x	x	x				x	x	
B08a					x	x	x	x	x	x		x	x
B11	x	x		x	x	x	x	x	x		x		x
B12			x		x	x	x		x	x		x	
LC	x	x	x	x	x	x	x	x	x	x		x	x
DEM	x	x	x	x	x	x	x	x	x	x	x	x	x
SLOPE					x	x	x						
ASPECT					x	x	x					x	x
TPI	x				x	x	x				x		
TRI		x			x		x			x	x	x	
ROUGHNESS					x	x	x	x	x	x	x	x	x
FLOWDIR					x	x	x	x	x				x
Одабрано (од 19):	4	4	3	4	18	15	18	11	11	11	8	13	12

Табела 11. Одабрани атрибути добијени различитим методама избора репрезентативних атрибута – имплицитно дефинисана временска компонента

Атрибут	CFS	SFS			SBS			HCS			GA		
		DT	RF	XGB	DT	RF	XGB	DT	RF	XGB	DT	RF	XGB
B02_06-07					x	x		x	x	x	x		
B02_08					x	x	x	x	x	x	x		x
B02_09-10					x	x	x					x	
B03_06-07				x	x	x	x					x	
B03_08					x	x	x					x	
B03_09-10					x	x	x				x	x	
B04_06-07		x			x	x	x				x	x	x
B04_08	x					x	x	x	x	x		x	x
B04_09-10					x	x	x	x	x	x	x		
B05_06-07				x	x	x	x	x	x	x	x		x
B05_08				x	x	x	x	x	x	x		x	
B05_09-10					x	x	x				x		
B06_06-07		x			x	x	x				x		
B06_08					x	x	x				x		x
B06_09-10		x			x	x	x						
B07_06-07			x		x	x	x						
B07_08	x				x	x	x			x	x		x
B07_09-10					x	x	x						
B08a_06-07	x			x	x	x	x				x	x	x
B08a_08					x	x	x	x	x	x	x	x	
B08a_09-10					x	x	x		x	x		x	
B11_06-07	x				x	x	x	x	x	x	x	x	x
B11_08					x		x	x			x		x
B11_09-10		x			x	x	x						
B12_06-07					x	x	x	x	x	x	x		x
B12_08					x	x	x	x		x		x	x
B12_09-10	x				x	x	x	x	x	x		x	
LC	x		x		x	x	x					x	
DEM	x	x	x	x	x	x	x	x	x	x	x	x	x
SLOPE	x				x	x	x			x	x	x	x
ASPECT					x	x	x				x	x	x
TPI	x				x	x	x	x	x	x		x	
TRI					x	x	x						x
ROUGHNESS					x	x	x						
FLOWDIR					x	x	x				x	x	x
Одабрано (од 35):	9	5	3	5	34	34	34	14	13	16	19	19	16

Прегледом табела може се уочити велика разноликост у скуповима одабраних репрезентативних атрибута. Може се рећи да је разноликост доминантно условљена примењеном методом избора репрезентативних атрибута, док је утицај примењене технике машинског учења (код омотач-метода) знатно мања. Иницијални преглед скупова атрибута добијених секвенцијалним претрагама сугеришу да је код свих приступа ове групе дошло до проналажења скупа атрибута који одговара локалном минимуму, који највероватније није и глобални. То се може образложити изузетно великом разликом у коначним скуповима атрибута, као и тиме што су све претраге прекидане после релативно малог броја итерација. Приступ базиран на корелацији је резултирао веома малим бројем одабраних атрибута, где у случају експлицитне дефиниције недостају временске одреднице. У односу на остале тестиране приступе, генетски алгоритми приликом креирања наредних генерација решења врше анализу знатно већег броја потенцијалних скупова атрибута, што даје веће шансе изналажења глобалног минимума. Резултати овог приступа су најсличнији резултатима секвенцијалне претраге планинарењем.

Појединачно, податак висинске представе терена (*DEM*) се показао као најрепрезентативнији, односно једино се тај атрибут нашао у сваком скупу репрезентативних атрибута. Класа основног земљишног покривача (*LC*) је такође присутна у 12/13 тестираних случајева код експлицитне временске дефиниције, али у само 6/13 случајева са имплицитном временском дефиницијом. Карактеристике изведене из висинске представе терена (*SLOPE, ASPECT, TPI, TRI, ROUGHNESS, FLOWDIR*) су разнолико одабирани, где се по правилу један или два атрибута много чешће бирају него остали. Што се *Sentinel-2* спектралних канала тиче, *B11* (један од краткоталасних инфрацрвених канала) се најчешће одабира као репрезентативан. У случају имплицитне временске дефиниције, такође се може уочити да рефлектанса из једног временског интервала најчешће предњачи у односу на остале. Најчешће је реч о опажањима из интервала јун – јул или августа, док се опажања из периода септембар – октобар знатно ређе одабирају као репрезентативни атрибути. Овакви резултати су сагласни са иницијалном анализом података, где се из фенолошких спектралних образаца могла уочити велика диференцијабилност класа преко канала *B11* (**Слика 16.**), односно висинско раслојавање типова шумске вегетације уочено са кутијастог дијаграма (**Слика 17.**).

Резултати показују да је избор атрибута секвенцијалном претрагом унапред [*SFS*] најагресивнији када је реч о одбацавању атрибута. Коначни скупови атрибута добијени овим приступом броје до трећине комплетног скупа атрибута, а у појединим случајевима одабира се свега десетина атрибута комплетног скупа. Примењена техника машинског учења не утиче значајно на бројност атрибута, при чему случајна шума над овим скупом података одабира мањи број атрибута у односу на друге две технике. У случају имплицитне временске репрезентације не може се уочити правилност између примењене технике машинског учења и одабраних атрибута. Сваком техником су одабирани у великој мери различити скупови атрибута, где је једино *DEM* заједнички атрибут, а остали атрибути су спектрални канали најчешће из интервала јун – јул. Насупрот томе, у случају експлицитне временске дефиниције су сваком примењеном техником машинског учења одабрани слични атрибути. Додатно је интересантно што се ни у једном случају временски атрибути нису нашли у коначном скупу репрезентативних атрибута, иако је претходни експеримент показао да имају утицај на излазну тачност класификације.

Секвенцијална претрага уназад [*SBS*] за одабир репрезентативних атрибута, супротно претходном приступу, показује „заштитничко“ понашање према комплетном скупу атрибута. Репрезентативни скупови атрибута одређени овим приступом се често разликују од комплетног скупа за свега неколико атрибута, а често и за само један. Не може се уочити нека посебна правилност око тога који атрибути су препознати као сувишни, нити да ли је то у некој вези са примењеном техником машинског учења.

Секвенцијална претрага планинарењем [*HCS*] је последњи тестирани приступ из групе секвенцијалних метода. Резултати показују да је она по бројности негде између претраге унапред и претраге уназад, где коначан скуп атрибута броји око половине атрибута из комплетног скупа. Ова метода, за разлику од секвенцијалне претраге унапред, препознаје временске атрибуте као значајне у свакој тестираној варијанти. Исто тако, не може се рећи ни да постоји нека значајна разлика између скупова одабраних атрибута по питању примењене технике машинског учења, где је преклапање репрезентативних скупова преко 70%.

Метода за избор атрибута базирана на корелацији *[CFS]* се такође може окарактерисати као агресивна по питању одбацавања атрибута. У случају имплицитне временске дефиниције одређено је девет репрезентативних атрибута, што је оквирно дупло више у односу број атрибута добијен секвенцијалном претрагом унапред, али ипак мање од броја атрибута добијеног применом приступа планинарењем или коришћењем генетских алгоритама. Насупрот томе, у случају експлицитне временске репрезентације, метода одабира исти број атрибута као и претрага унапред, при чему је већина атрибута идентична. Такође, ни ова метода није препознала временске одреднице опажања као значајне атрибуте.

Последњи тестирани приступ користи генетске алгоритме *[GA]* за избор скупа репрезентативних атрибута. Резултујући скупови атрибута су нешто бројнији у односу на скупове атрибута добијених секвенцијалном претрагом планинарењем, при чему су обе методе препознале већи број истих репрезентативних атрибута. Генерално делује да случајна шума у комбинацији са овим приступом резултира нешто већим бројем репрезентативних атрибута у односу на остале две технике, али није могуће проценити да ли је реч о правилу или последици коришћеног скупа података.

Ради одређивања коначне оцене перформанси сваког креираног скупа репрезентативних атрибута примењена је усвојена процедура за оцену квалитета класификације (пет пута насумично поновљена петострука крос-валидација са изостављањем комплетних локација). Ради смањења броја комбинација за валидацију, валидација је рађена оном техником машинског учења којом је репрезентативни скуп атрибута и настао (у случају омотач-метода), а код филтер-метода су примењене све три технике. Додатно, валидација је урађена и за комплетне скупове атрибута ради утврђивања референтне вредности. У наставку су приказане највише постигнуте метрике перформанси за све три технике машинског учења и за сваки приступ репрезентације временске компоненте (**Табела 12.**). Резултати постигнутих квантитативних мера перформанси за све тестиране приступе налазе се у оквиру **прилога 2.**

Табела 12. Највише постигнуте мере перформанси и скупови атрибута који су коришћени за сваки приступ репрезентације временске компоненте

Техника машинског учења	Временска одредница	Правило агрегације	Скуп атрибута	ОА [%]	Карра
		имплицитно	SFS	76.08	0.65
DT	doy	MC	сви атрибути	72.32	0.60
	day + month	MC	HCS	72.52	0.61
		имплицитно	сви атрибути	83.33	0.76
RF	doy	GM	SBS	81.30	0.74
	day + month	GM	HCS	81.33	0.74
		имплицитно	сви атрибути	83.22	0.76
XGB	doy	GM	SBS	81.43	0.74
	day + month	GM	сви атрибути	81.20	0.74

Постигнуте вредности укупне тачности и капа коефицијента слагања указују на несумњив допринос укључивања додатних атрибута у класификациони модел. Постигнуте су више мере перформанси у већини тестираних скупова атрибута него у случају коришћења само *Sentinel-2* опажања. Резултат са највишим квантитативним метрикама код имплицитне временске дефиниције користи случајну шуму (ОА = 83.33%, *Карра* = 0.76) са комплетним скупом атрибута. Приликом експлицитне временске дефиниције, највише мере перформанси се постижу са подскупом изабраних репрезентативних атрибута и то применом технике градијентног појачавања са [*doy*] временском одредницом и геометријском средином као правилом агрегације (ОА = 81.43%, *Карра* = 0.74).

Срачунате вредности метрика перформанси указују на то да већина трендова из првог експеримента и даље важи. Случајна шума и градијентно појачавање и даље показују очигледну предност у односу на стабло одлучивања, али је та предност ипак нешто умањена укључивањем додатних атрибута. Све три функције агрегације и даље дају врло сличне резултате и још увек не постоји јасна дистинкција која експлицитна временска одредница даје више мере перформанси. Ипак, укључивање додатних атрибута резултира и неким новим трендовима. Најзначајнији је да се може уочити

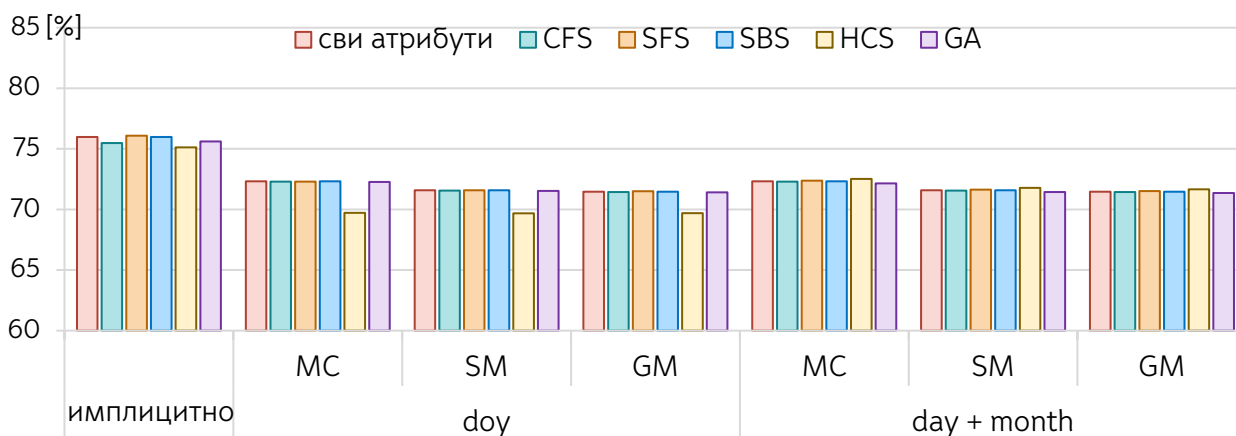
предност имплицитног приступа у односу на експлицитну временску репрезентацију, коју у случају коришћења само сателитских опажања није било. Нумерички, предност код стабла одлучивања износи скоро 4% укупне тачности и 0.05 по капа коефицијенту слагања, а за случајну шуму и градијентно појачавање достиже 2% укупне тачности и 0.02 капа коефицијента слагања.

Као што је и очекивано, постоје разлике у оствареним мерама перформанси коришћењем различитих скупова атрибута. Међутим, релативно поларизован утицај избора репрезентативних атрибута у односу на комплетан скуп атрибута, када је реч о мерама перформанси класификационог модела, представља изненађење. Показало се да коришћење комплетног скупа атрибута у великом броју случајева резултује највишим квантитативним метрикама перформанси, а ако то није случај, онда подскуп атрибута или даје незнатно више или значајно ниже метрике. Када се посматра приступ за избор репрезентативних атрибута, секвенцијална претрага уназад [*SBS*] је најчешћи приступ који надмашује комплетан скуп атрибута, а у појединим случајевима се добро показала и секвенцијална претрага планинарењем [*HCS*].

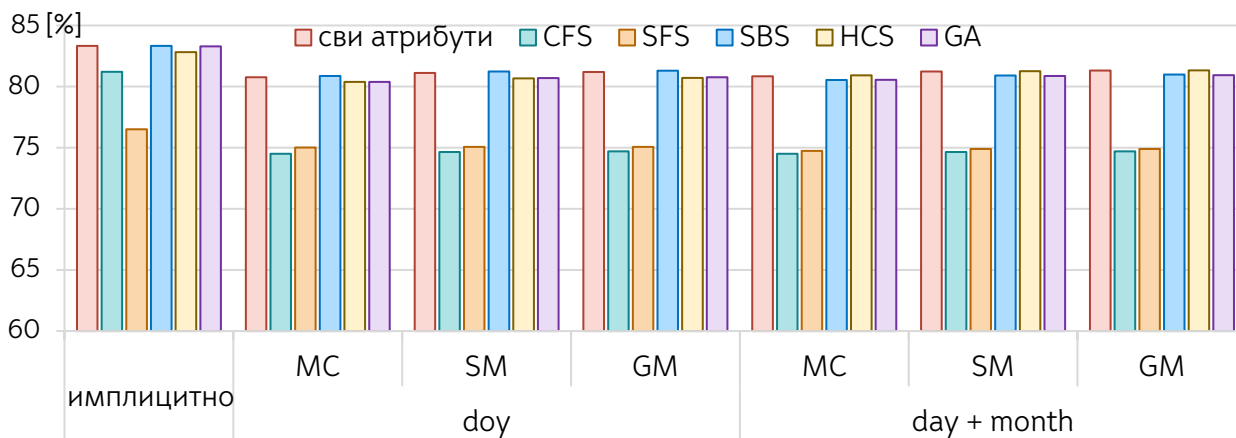
Објашњење за овако велике разлике у постигнутим мерама перформанси приступа за избор атрибута треба тражити у особинама примењених техника машинског учења. Приликом изградње стабла одлучивања, избор најподеснијег атрибута за чворове стабла подразумева и њихово интерно рангирање. Практично се на тај начин врши интерна селекција репрезентативних атрибута, где ће кориснији атрибути бити при врху стабла и имаће велики утицај на излазно решење, односно мање корисни атрибути ће бити близу листова стабла и са мањим утицајем. Пошто све три коришћене технике машинског учења користе стабло одлучивања, очигледно је да им расположивост комплетног скупа атрибута и интерно рангирање највише погодују. Насупрот томе, избор репрезентативних атрибута коришћењем било ког приступа може само незнатно побољшати или, што је чешћи случај, покварити излазне резултате.

Појединачно, избор репрезентативних атрибута има највећи ефекат код стабла одлучивања (**Слика 20.**). Може се рећи да избор атрибута у комбинацији са овом техником машинског учења значајно смањује меморијске захтеве уз повећање перформанси модела. У појединим случајевима су постигнуте највише мере перформанси коришћењем 3/18 скупа атрибута. Коришћење комплетног скупа атрибута у овом случају се може окарактерисати као непотребно, јер резултује мерама

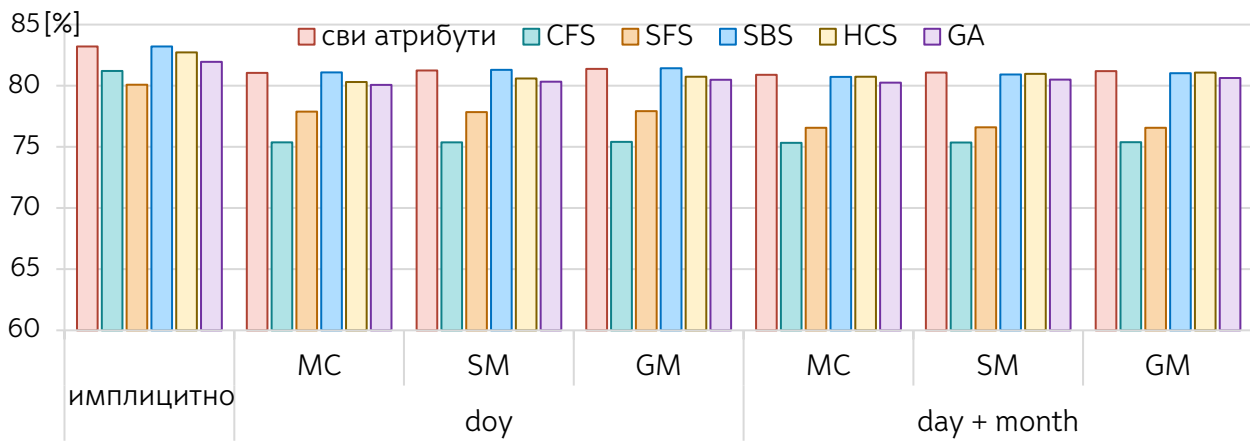
перформанси које су идентичне или заостају у односу на оне добијене са коришћењем других скупова атрибута. Насупрот томе, избор репрезентативних атрибута има помешане ефекте код случајне шуме и градијентног појачавања (Слика 21. и Слика 22.). Комплетан скуп атрибута је у сваком тестираном приступу један од модела у врху по мерама перформанси. Очигледно је да изузетно агресивно избацавање атрибута карактеристично за секвенцијалну претрагу унапред и избор базиран на корелацији не погодују овим техникама, јер резултира значајним умањењем метрика перформанси модела. Са друге стране, секвенцијална претрага уназад користи свега неколико атрибута мање у односу на комплетан скуп атрибута, уз идентичне или минимално више метрике. Делује да једино секвенцијална претрага планинарењем и генетским алгоритмом теже да изврше оптимално смањење скупа атрибута, али је то најчешће по цену умањених мера перформанси модела. Међутим, то умањење је знатно мање у односу на претрагу унапред и на избор базиран на корелацији.



Слика 20. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом стабла одлучивања



Слика 21. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом случајне шуме



Слика 22. Графичко поређење укупне тачности различитих приступа избора репрезентативних атрибута применом градијентног појачавања

Другим речима, може се рећи да избор атрибута није уродио плодом код случајне шуме и градијентног појачавања. То се може образложити тиме да је избор атрибута „упао“ у локалне минимуме, али и ефектом који обезбеђује коришћење ансамбла стабла одлучивања. Очигледно је да се ове технике машинског учења добро носе са великим скуповима атрибута, односно да њихови интерни механизми довољно добро рангирају важност атрибута. Поставља се питање да ли уопште има смисла спроводити избор репрезентативних атрибута ако ће резултујући репрезентативни скуп бројати свега неколико атрибута мање, или ће имати знатно ниже метрике перформанси. Уместо тога, може се рећи да коришћење комплетног скупа атрибута представља сигурни приступ који, чак и ако не даје максималне мере перформанси, оне су онда само незнатно ниже.

5.1.3 Експеримент 3: Утицај балансирања класа

Расположиви скуп података се може сматрати небалансираним због неравномерне заступљености циљних класа. Циљ овог експеримента јесте анализа утицаја неравномерне заступљености класа, као и да ли је могуће превазилажење овог проблема применом *SMOTE* алгоритма за балансирање класа. Експеримент је тако реализован да је балансирање класа вршено само над скуповима за тренирање класификационог модела, независно за сваку итерацију из пет пута поновљене петоструке крос-валидације са изостављањем. Ради смањења броја тестирања, утицај балансирања класа је анализиран само за моделе и скупове атрибута који су давали највише мере перформанси у претходном експерименту (**Табела 12.**).

Описана процедура балансирања класа (**поглавље 4.5**) је примењена одвојено за скупове података са имплицитном и експлицитном временском дефиницијом. Комбинације параметара и техника машинског учења којима су постигнуте највише квантитативне мере перформанси су приказане у наставку (**Табела 13**), а комплетне табеле са метрикама свих тестираних приступа се налазе у оквиру **прилога 3**.

Табела 13. Комбинације параметара чијим коришћењем су постигнуте највише мере перформанси (скуп података је балансиран применом *SMOTE* алгоритма)

Техника машинског учења	Скуп атрибута	Временска одредница	Правило агрегације	Балансиран скуп података	
				ОА [%]	Карра
	SFS	имплицитно		74.51	0.63
DT	сви атрибути	doy	GM	69.28	0.58
	HCS	day + month	GM	69.26	0.57
RF	сви атрибути	имплицитно		83.36	0.76
	SBS	doy	GM	81.76	0.75
	HCS	day + month	GM	81.82	0.75
XGB	сви атрибути	имплицитно		83.20	0.76
	SBS	doy	GM	81.57	0.75
	сви атрибути	day + month	GM	81.27	0.74

Највећа предност коришћења балансираног скупа података јесте значајно повећање мера перформанси у смислу препознавања мањински заступљених класа (4 – ксерофилних шума белобрабића и 8 – црног граба и шибљака бора кривуља), што се огледа у већим вредностима корисникове и произвођачеве тачности. То се посебно истиче код стабла одлучивања, где су мањинске класе потпуно игнорисане ако се користи небалансирани скуп података, док то није случај код балансираног скупа података. Ни код преостале две технике машинског учења овај тренд није занемарљив, али је мање изражен. Неповољан ефекат је што су претходно описани трендови најчешће пропраћени смањењем мера перформанси код већински заступљених класа. Додатно, балансирање класа је значајно повећало меморијске и процесорске захтева неопходне за креирање модела, што је разумљива последица повећања броја опажања коришћених за креирање модела.

Последица претходно уочених трендова код мањинских и већинских класа се осликава и на срачунате вредности укупне тачности, код којих доминантан утицај имају већински заступљене класе. Последично, у највећем броју случајева је постигнута веома слична укупна тачност као и код небалансираног скупа података, при чему највећи напредак износи 0.5%. Са друге стране, стабло одлучивања је показало знатно ниже метрике перформанси које се у неким случајевима огледају у паду укупне тачности и већем од 3%. Уопштено говорећи, квалитетнији је онај класификациони модел који све класе препознаје (макар и са нижим мерама перформанси) него модел у коме постоје екстремне разлике у препознавању појединих класа. Зато се може закључити да добити балансирања класа преовлађују побројане негативне карактеристике, тј. да балансирање класа свакако доприноси поузданијим класификационим моделима.

5.1.4 Експеримент 4:

Наглашавање просторних аспеката у класификационом моделу

Сви претходни експерименти су вршени тако да је класификација рађена на нивоу пиксела, тј. коришћене су само информације везане за пиксел који се тренутно обрађује. Ово отвара и питање да ли је то оптималан приступ и шта би се укључивањем суседних пиксела у процедуру класификације постигло. Са том намером је спроведен четврти експеримент, где је за претходно дефинисане оптималне скупове репрезентативних атрибута и уз коришћење балансираног скупа података извршено наглашавање просторне компоненте у класификационом моделу. То је постигнуто агрегацијом оцењених вредности вероватноћа унутар Муровог суседства посматраног пиксела, при чему су примењена иста правила агрегације као и код респектовања временске компоненте (најчешћа класа *[MC]*, класа са највећом вероватноћом срачуната простом *[SM]* и геометријском средином *[GM]*). Разматрана су Мурова суседства величине 3×3 и 5×5 пиксела, што значи да је у случају експлицитне временске компоненте агрегација вероватноћа вршена и у просторном и временском домену, а код имплицитне временске компоненте само у просторном домену. Скуп параметара са највишим постигнутим мерама перформанси приказан је у наредној табели (**Табела 14.**), а комплетне табеле са метрикама квантитативних мера перформанси свих тестираних приступа се налазе у оквиру **прилога 4.**

Табела 14. Комбинација параметара којима су постигнуте највише мере перформанси уз наглашавање просторних компоненти у класификационом моделу

Техника машинског учења	Скуп атрибута	Величина суседства	Временска одредница	Правило агрегације	ОА [%]	Карра
DT	SFS	5×5	имплицитно	GM	77.36	0.67
	сви атрибути	5×5	doy	GM	71.08	0.60
	HCS	5×5	day + month	GM	70.87	0.60
RF	сви атрибути	5×5	имплицитно	SM	85.43	0.79
	SBS	5×5	doy	SM	83.32	0.77
	HCS	5×5	day + month	SM	83.62	0.78
XGB	сви атрибути	5×5	имплицитно	GM	85.29	0.79
	SBS	5×5	doy	GM	82.68	0.76
	сви атрибути	5×5	day + month	GM	82.29	0.76

Ако се добијене метрике из **табеле 14.** упореде са метрикама перформанси класификације приказаним у **табели 13,** може се закључити да је ефекат наглашавања просторних компоненти значајан. То се огледа у повећању мера перформанси код свих тестираних приступа. Позитивни ефекти постоје већ код суседства величине 3×3, а мало су већи код суседства 5×5. У просеку се укупна тачност повећава за 2%, како за имплицитну тако и за експлицитну временску дефиницију. Повећање постоји и када је реч о произвођачевој и корисниковој тачности, и то код свих циљних класа. Генерално, већина правила на која су претходни експерименти указивали важе и након укључивања суседства. Имплицитна временска репрезентација и даље даје више мере перформанси од експлицитне репрезентације, али треба нагласити да експлицитна временска репрезентација са суседством може да парира пикселској имплицитној временској репрезентацији. Такође све три функције агрегације и даље постижу сличне мере перформанси, при чему овога пута агрегација простом аритметичком средином незнатно надмашује преостале две функције агрегације. Исто тако, случајна шума и градијентно појачавање још више надмашују стабло одлучивања, при чему и даље не постоји дефинитивна предност било које од ове две технике машинског учења.

Овакви резултати показују да укључивање суседства помаже да се изађе на крај са шумом у подацима који се код класификације на нивоу пиксела јавља у виду усамљених

пиксела различитих класа у хомогеним шумским подручјима. Ипак, треба напоменути да увођење просторно-временске или само просторне агрегације доноси и неке нове проблеме, а главни је осетљивост методологије с обзиром на временске и/или просторне хетерогености. Поуздани резултати се могу очекивати за подручја у којима не долази до промена на терену током временског периода покривеног сателитским опажањима и у оквиру дефинисаног суседства. У осталим случајевима се може очекивати да предложена методологија до неке мере може да се избори са променама насталим на терену. Свако од коришћених *Sentinel-2* опажања утиче на одлуку о коначној класи шумске вегетације и било каква временска или просторна промена потенцијално може да потисне друга опажања. Временски гледано, то значи да ће модел вероватно препознати извесне промене које у посматраном временском интервалу трају дужи. Аналогно томе, биће препознате и само просторне промене које имају већи просторни опсег у односу на дефинисану величину суседства. То практично значи да просторна резолуција картирања постаје све грубља како се величина суседства повећава. У исто време, коришћење суседства има и ефекте глачања, при чему се изоловани пиксели класификују у класу пиксела која доминира у њеном суседству. Због тога се не препоручује повећавање временског интервала посматрања и величине суседства више него што је неопходно. Краћи временски интервал посматрања смањује број промена на терену које се дешавају, а одговарајућа величина суседства ограничава степен глачања који се јавља.

5.2 Оптимални параметри методологије

Кроз реализоване експерименте су анализирани главне компоненте предложене методологије, односно начини репрезентације временске компоненте у подацима, наглашавање просторне компоненте укључивањем суседства у процедуру класификације, креирање и избор репрезентативних атрибута, као и балансирање класа. Коришћена техника машинског учења такође прожима сваку од ових фаза и не сме бити занемарена. Ипак, важно је напоменути да проглашени модели са оптималним параметрима често постижу веома сличне и само незнатно више мере перформанси у односу на остале креиране моделе. Због тога се проглашени оптимални параметри методологије морају посматрати само за дефинисано подручје истраживања и за обезбеђен скуп података. То значи да је у случају значајне промене било које од ове две карактеристике процедуру одређивања оптималних параметара

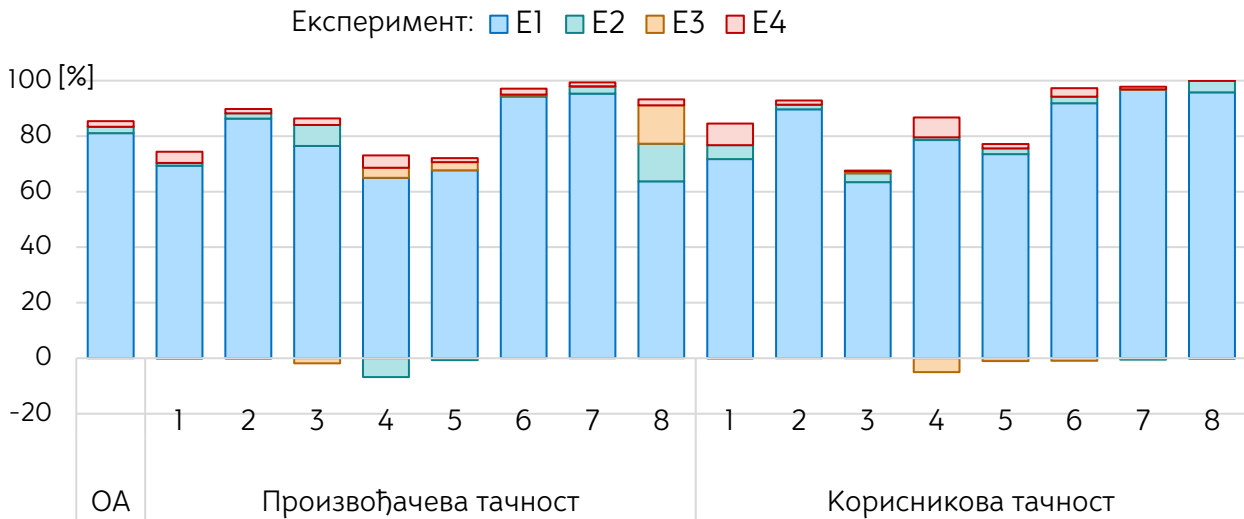
методологије потребно поновити. Коначни оптимални параметри за два приступа респектовања временске компоненте који одговарају класификационим моделима са највишим мерама перформанси приказани су у наредној табели (**Табела 15.**).

Табела 15. Комбинација параметара методологије којима су постигнуте највише мере перформанси

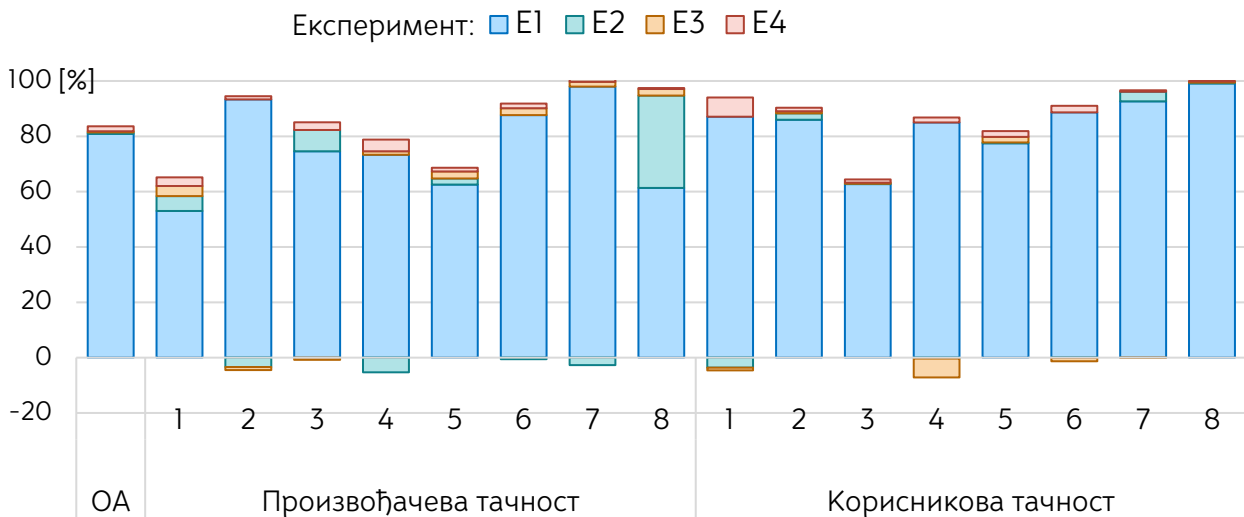
	Имплицитна временска репрезентација	Експлицитна временска репрезентација
Техника машинског учења	Случајна шума (RF)	Случајна шума (RF)
Одабир репрезентативних атрибута	сви атрибути	SBS метода за избор атрибута
Величина суседства	5×5	5×5
Правило агрегације	SM	SM

Уочљиво је да за већину параметара не постоје разлике када се упореде два приступа временске репрезентације. Показало се да су највише мере перформанси добијене код класификационих модела креираних коришћењем случајне шуме и да треба користити суседство величине 5×5 у комбинацији са простом аритметичком средином као правилом агрегације. Разлика једино постоји код одабраних скупова атрибута, где имплицитна репрезентација временске компоненте подразумева комплетан скуп атрибута, док је код експлиците временске репрезентације комплетан скуп умањен за четири атрибута који су оцењени као мање информативни.

Наравно, јасно је да компоненте методологије неједнако утичу на коначне мере перформанси класификационог модела. Очигледан је тренд повећања мера перформанси како се методологије усложњава, а појединачни допринос сваке компоненте се може сагледати са наредних графика (**Слика 23.** и **Слика 24.**).



Слика 23. Допринос корака методологије постигнутим мерама перформанси оптималних класификационих модела са имплицитном временском репрезентацијом



Слика 24. Допринос корака методологије постигнутим мерама перформанси оптималних класификационих модела са експлицитном временском репрезентацијом

Појединачно, укључивање додатних и избор репрезентативних атрибута највише доприноси повећању мера перформанси. На другом месту по доприносу се налази наглашавање просторних компоненти, док најмање доприноси балансирање класа. Највећи број метрика показује тренд повећања, међутим, вредности корисникове и произвођачеве тачности појединих класа су чак и снижене у појединим случајевима. Негативни ефекти су најизраженији приликом укључивања додатних и избора репрезентативних атрибута, као и приликом балансирања класа. Ипак, треба напоменути да у тим случајевима корист код осталих класа вишеструко надмашује ове губитке, тако да би избацивање било ког корака методологије било погрешно.

Вредности срачунатих метрика које одговарају одређеним оптималним параметрима методологије су приказане у наредној табели (Табела 16.).

Табела 16. Вредности метрика оптималних класификационих модела

Метрика		Имплицитна временска репрезентација	Експлицитна временска репрезентација
ОА [%]		85.43	83.62
Каппа		0.79	0.78
Произвођачева тачност класа шумске вегетације [%]	1	74.30	65.17
	2	89.72	90.01
	3	84.59	84.29
	4	66.23	73.51
	5	71.38	68.61
	6	97.12	91.29
	7	99.37	98.38
	8	93.23	97.45
Корисникова тачност класа шумске вегетације [%]	1	84.54	89.46
	2	92.86	90.36
	3	67.62	64.40
	4	81.72	79.67
	5	76.14	81.87
	6	96.35	89.72
	7	97.30	96.58
	8	100.00	100.00

Укупна тачности и капа коефицијент слагања указују на предност приступа са имплицитном временском дефиницијом у односу на експлицитну временску дефиницију. Делује да експлицитна временска компонента резултира вишим мерама перформанси код мањински заступљених класа, али да је у осталим случајевима подеснија имплицитна временска репрезентација. Ипак, вредности корисникове и произвођачеве тачности оптималних модела за ова два приступа указују да је предност није доминантна, већ је напротив, у великом броју случајева само незнатна. Због тога одабир експлицитног или имплицитног приступа треба вршити узимајући у обзир и

остале предности и мане сваког приступа. Када су разлике у мерама перформанси ова два приступа веома мале, попут у овом истраживању, те остале карактеристике сваког приступа се могу показати као доминантан фактор приликом избора.

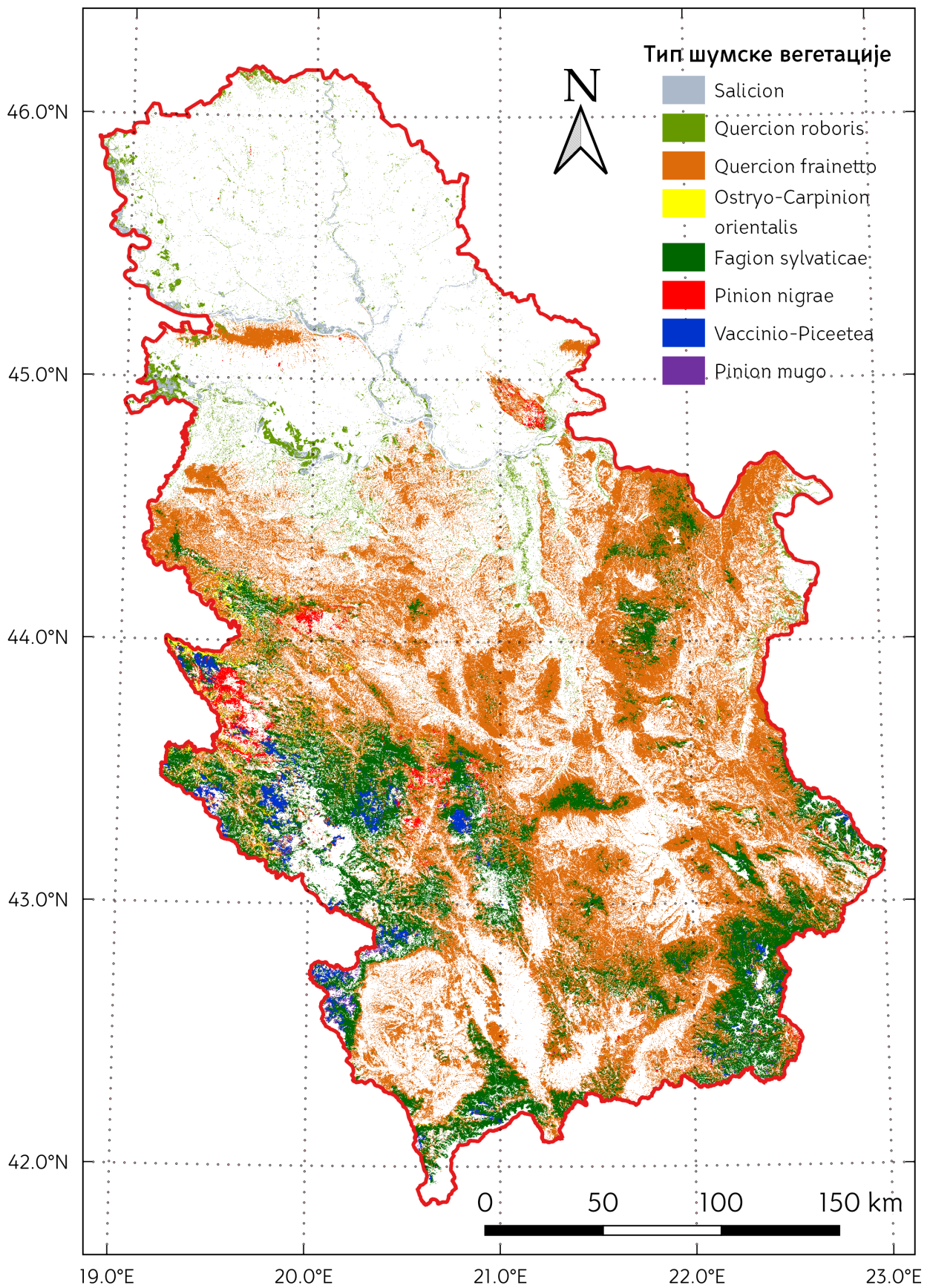
Резултати овог истраживања се подударају са резултатима других сличних студија заснованих на оптичким сателитским опажањима високе просторне резолуције. *Grabska* и др. (2019) су постигли укупну тачност већу од 90% на истраживаној површини од 305 km² за девет типова шумске вегетације користећи 18 *Sentinel-2* снимака без облака. *Liu* и др. (2018) су идентификовали типове шума користећи *Sentinel-1A*, *Sentinel-2A*, *Landsat-8* и податке висинске представе податке са укупном тачношћу од 82.78%. Они су издвојили четири типа шумске вегетације и четири мешовите шумске класе на површини од 2261 km². *Lu* и др. (2017) су имплементирали просторно-спектрално-временску интегрисану фузију података (енгл. *Spatial-Spectral-Temporal Integrated Fusion*) ради идентификације шест хомогених и два мешовита типа шума постижући ОА од 83.6% на површини од 1610 km². Поред тога што покрива много веће подручје истраживања, ово истраживање и његово предложено решење се ослањају примарно на оптичке сателитске изворе и неколико додатних извора података. Стога, постоји потенцијал за даље побољшање модела коришћењем других извора података као што су радарски сателитски подаци, подаци ласерског скенирања из ваздуха (енгл. *airborne laser scanning*), различити оптички извори података или други помоћни подаци.

Оптимални параметри методологије су тестирани и одређени користећи 20-метарске канале *Sentinel-2* сателитске мисије, чијој су просторној резолуцији сви остали извори података прилагођени. Како су неки спектрални канали расположиви и у просторној резолуцији од 10 m, поставља се питање како би њихово коришћење утицало на ефекте мешовитих пиксела и на постигнуте мере перформанси класификације. Осим тога, очекивано је да виша просторна резолуција може да помогне и када је реч о претходно разматраним питањима хетерогености, где би се потенцијално могле моделирати мање промене. За веће просторне резолуције се такође може претпоставити снажнији ефекат суседства, где би величина суседства могла постати још важнији параметар методологије. Међутим, постоји неколико аргумената који оправдавају одабир грубље просторне резолуције. Прво, то је усклађено са дефинисаном класификационом шемом која је дизајнирана за картирање типова шумске вегетације на нивоу целе земље. Коришћена шема дефинисана је према Бази података о распрострањености

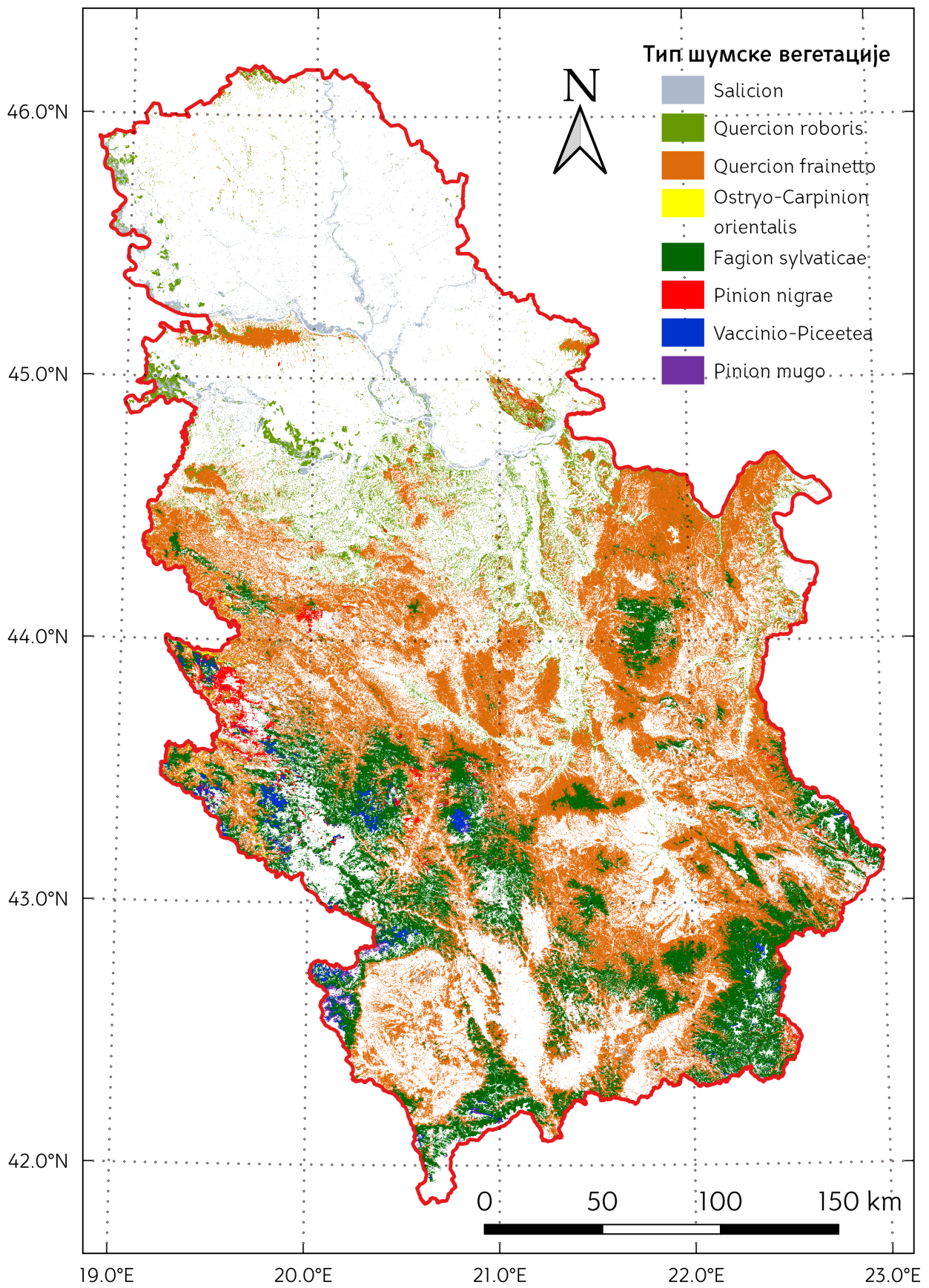
потенцијално угрожених врста и станишта Србије, а повећање просторне резолуције би значило увођење нових класа шумске вегетације за које недостају одговарајући теренски подаци. Друго и још важније, један од главних циљева предложене методологије јесте да се она користи на великим подручјима, па се грубљом просторном резолуцијом смањују меморијски и процесорски захтеви на разумну меру. Ипак, ако би се претходни проблеми предупредили обезбеђивањем детаљније класификационе шеме и пратећих теренских података, заједно са рачунаром одговарајућих перформанси, може се очекивати да се методологија добро покаже и за податке више просторне резолуције.

5.3 Коначна карта шумске вегетације

Коначна карта типова шумске вегетације је креирана за територију Републике Србије користећи описану методологију и утврђене оптималне параметре. Како оба приступа временске репрезентације показују одређене предности и мане, креиране су две карте (**Слика 25.** и **Слика 26.**). Генерално, визуелном упоређивањем се не може уочити значајна разлика у распореду класа на резултујућим картама. Може се рећи да распоред типова шумске вегетације одговара постојећим општим картама вегетације: Карти природне вегетације Европе (размера 1: 2 500 000) (*Bohn* и др., 2000) и Карти природне потенцијалне вегетације СФР Југославије (размера 1: 1 000 000) (*Fukarek* и *Jovanović*, 1983). Ипак, локалне разлике постоје и претежно се односе на различит распоред класа унутар група листопадне или четинарске шумске вегетације. Најизраженије разлике су по питању ксерофилних шума храстова (3 - *Quercion frainetto*) и мезофилних шуме букве (5 - *Fagion sylvaticae*), односно између термофилних (6 - *Pinion nigrae*) и фригорифилних четинарских шума (7 - *Vaccinio-Piceetea*).



Слика 25. Коначна карте типа шумске вегетације применом модела са имплицитном временском компонентом



Слика 26. Коначна карте типа шумске вегетације применом модела са експлицитном временском компонентом

Експлицитна временска дефиниција визуелно има израженије ефекте глачања, тј. резултујуће површине ове карте су хомогеније у односу на карту имплицитне временске дефиниције. Ово је очигледно последица примењене просторно-временске агрегације. Како се у случају експлицитне временске дефиниције агрегација врши на основу већег броја опажања без тежина, утицај појединачног опажања је мањи и разлике бивају потиснуте. Увођењем тежина према просторној и временској удаљености опажања у посматраном суседству ефекат глачања може бити контролисан. Алтернативно, глачање може бити умањено и дефинисањем неке нове агрегационе функције. Након креирања коначних карата, за сваку класу шумске вегетације је срачуната површина и заступљеност које заузимају на њима (Табела 17.).

Табела 17. Површине и удео заступљености класа шумске вегетације у коначним картама

ID класе	Тип шумске вегетације	Асоцијативно фитосоциолошко име	Имплицитна временска дефиниција		Експлицитна временска дефиниција	
			Површина [km ²]	Удео [%]	Површина [km ²]	Удео [%]
-9999	недостајући подаци		330.06	0.87	24.38	0.06
1	Хигрофилне шуме врба и топола	Salicion	887.19	2.34	1 880.65	4.96
2	Хигрофилне шуме лужњака и пољског јасена	Quercion roboris	1 643.30	4.34	2 991.77	7.89
3	Ксерофилне шуме храстова	Quercion frainetto	24 245.62	63.97	22 063.13	58.22
4	Ксерофилне шуме белограбића и црног граба	Ostryo-Carpinion orientalis	388.28	1.02	139.32	0.37
5	Мезофилне шуме букве	Fagion sylvaticae	8 740.38	23.06	9 427.00	24.87
6	Термофилне четинарске шуме	Pinion nigrae	710.50	1.87	683.15	1.80
7	Фригорифилне четинарске шуме	Vaccinio-Piceetea	892.47	2.35	583.78	1.54
8	Шибљаци бора кривуља	Pinion mugo	61.57	0.16	106.18	0.28

Може се одмах уочити да недостајућих пиксела нема много. Очекивано, експлицитна временска дефиниција се по том питању боље показала, јер се добија занемарљиво мали број недостајућих пиксела. Са друге стране, количина недостајућих података и у случају имплицитне временске дефиниције је испод 1%, што се такође може сматрати прихватљивим. Разлика у недостајућим пикселима ће свакако бити много израженија када се користи скуп *Sentinel-2* снимака са присутнијим проблемом облачности, па се ту очекује и значајнија предност експлицитне временске дефиниције. Срачунате површине и удели заступљености анализираних типова шумске вегетације се доста добро слажу са резултатима добијеним првом Националном инвентуром шума Србије у периоду 2004 – 2006. која је урађена коришћењем традиционалних метода теренског прикупљања података (Банковић и др., 2009а). Значајна одступања су добијена у уделима мезофилних шума букве (5 – *Fagion sylvaticae*) и ксерофилних шума храстова (3 – *Quercion frainetto*). Удео шума букве је у резултујућим картама потцењен за 7%, што одговара површини од око 2 000 km². Ксерофилне шуме храстова у инвентури шума имају удео од око 30%, што је дупло мање од удела на картама. Посматрано у површинама, ова разлика у уделу износи око 17 000 km². Један део овог одступања се свакако може приписати релативно великом временском распону између креираних карата (2019.) и инвентуре шума (2006.). Ипак, одступање удела и површине ксерофилних шума храстова је превелико и захтева додатно објашњење. Потенцијални извор проблема може бити коришћена шумска маска која процењује шумску покривеност Србије на око 43%, у поређењу са 29.1% из националне инвентуре шума. У комбинацији са карактеристиком методологије да се сваки пиксел подручја увек разврстава у неку шумску класу, овај „вишак“ пиксела је очигледно претежно завршио управо у тој класи, чиме је нереално повећана њена површина.

Обавезно сврставање пиксела у неку од циљних класа очигледно представља недостатак методологије. У ситуацији у којој су сви типови шумске вегетације добро репрезентовани теренским подацима, то не би представљало проблем. Међутим, када је подручје истраживања веома сложено или недовољно истражено у смислу шумске вегетације, тада није могуће представити све типове на истраживаном подручју. Последично, методологија ће пикселе незаступљених класа погрешно распоредити у неку од дефинисаних класа, што ће вештачки повећавати њихове површине и заступљености. Потенцијално решење овог проблема може бити анализа оцењених

вредности вероватноће, са циљем идентификације пиксела са малом вероватноћом, што може указивати на то да они вероватно не припадају ниједној од усвојених класа.

Други идентификовани недостатак креираних карата, самим тим и методологије, је што се она ослања на шумску маску у кораку дефинисања подручја у којем се врши картирање типова шумске вегетације. У овом истраживању шумска маска је креирана користећи Коперникус паневропски слој покривности крошње шумског покривача за епоху 2018, селекцијом пиксела са покривношћу крошњи већом од 10%. Важно је користити шумску маску која је временски најближа посматраном временском интервалу покривеном сателитским опажањима, посебно у подручјима брзе промене шума. Шумска маска коришћена у овом истраживању се односи на 2018. годину, што ствара једногодишњи јаз са коришћеним *Sentinel-2* опажањима. Ипак, већи проблем код шумске маске је то што је она креирана само на основу покривности крошњи, без узимања у обзир висине стабала. То за последицу има да маска обухвата и подручја са шибљем и нижим деградираним облицима шумске вегетације. Пошто овакви типови вегетације у већини инвентура шума нису препознати као шумско подручје, тешко је вршити директно упоређивање ова два скупа података.

5.4 Валидација коришћењем независног скупа података

Коришћена крос-валидација са изостављањем комплетних локација може дати информације о тачности модела, али су срачунате метрике тачности генерално пристрасне када је реч о оцени квалитета карте на комплетном подручју истраживања (*Olofsson* и др., 2014). Стога је валидација коначних класификационих карата типова шумске вегетације извршена и поређењем са независним скупом података.

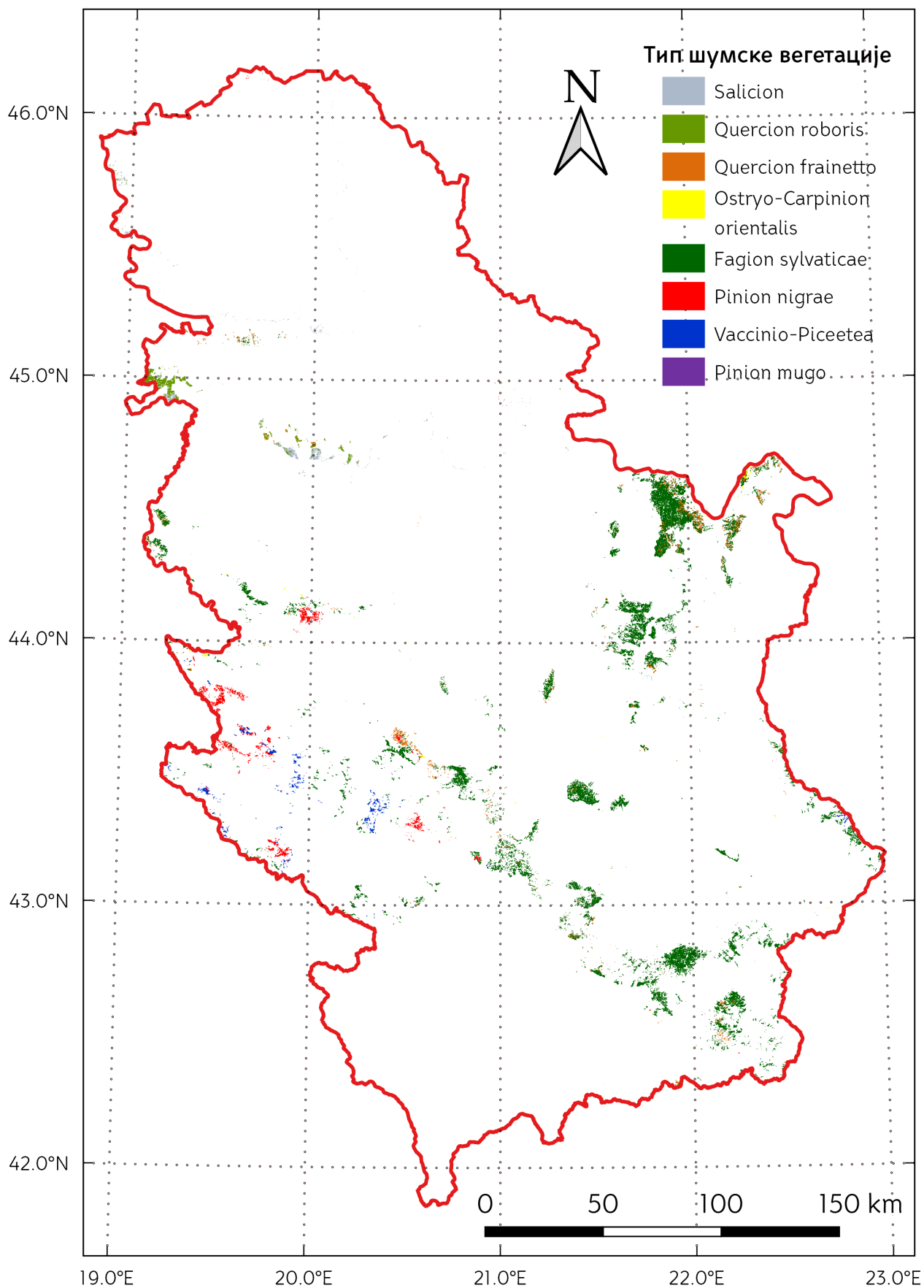
Независни скуп података за валидацију је уступио Биолошки факултет Универзитета у Београду. У питању је скуп података који се за потребе Завода за заштиту природе Републике Србије користи за израду и валидацију карте типова станишта у Србији, а у оквиру пројеката за успостављање еколошке мреже, укључујући и *Natura 2000* мрежу (*Лакушић* и др., 2021). Подаци представљају извод из базе података државних шума који обухвата 25 235 полигона шумских одсека и о којима се води информација о идентитету репрезентативних шумских састојина. Користећи информацију о састојини шумског одсека, извршена је конверзија у класификациону шему примењену за израду карата типова шумске вегетације. Скуп полигона шумских одсека је затим преведен у

растерски формат (Слика 27.). Растеризација је спроведена тако да се резултујући референтни растер у потпуности просторно поклапа са излазним картама типова шумске вегетације, односно да имају исти просторни обухват, просторну резолуцију, број и поравнање пиксела. Основне информације о референтном растерском скупу података налазе се у наредној табели (Табела 18).

Табела 18. Информације о референтном скупу података

ID класе	Тип шумске вегетације	Асоцијативно фитосоциолошко име	Број полигона	Површина [km ²]	Број пиксела
1	Хигрофилне шуме врба и топола	<i>Salicion</i>	2 310	68.39	162 937
2	Хигрофилне шуме лужњака и пољског јасена	<i>Quercion roboris</i>	1 419	100.83	235 054
3	Ксерофилне шуме храстова	<i>Quercion frainetto</i>	2 914	184.08	436 373
4	Ксерофилне шуме белограбића и црног граба	<i>Ostryo-Carpinion orientalis</i>	135	14.46	32 850
5	Мезофилне шуме букве	<i>Fagion sylvaticae</i>	15 046	1 555.79	3 789 761
6	Термофилне четинарске шуме	<i>Pinion nigrae</i>	1 647	149.01	328 007
7	Фригорифилне четинарске шуме	<i>Vaccinio-Piceetea</i>	806	82.52	195 184
8	Шибљаци бора кривуља	<i>Pinion mugo</i>	0	0.00	0

Одмах се може уочити неравномерна заступљеност класа у референтном скупу података. То је последица стварне заступљености класа на терену, али и тога што се ради само о шумама које су у државном власништву. Такође, у референтном скупу података нема класе која одговара шибљацима бора кривуља (8 - *Pinion mugo*), па је ова класа изузета из валидације.



Слика 27. Референтни растер креиран за валидацију карата типова шумске вегетације

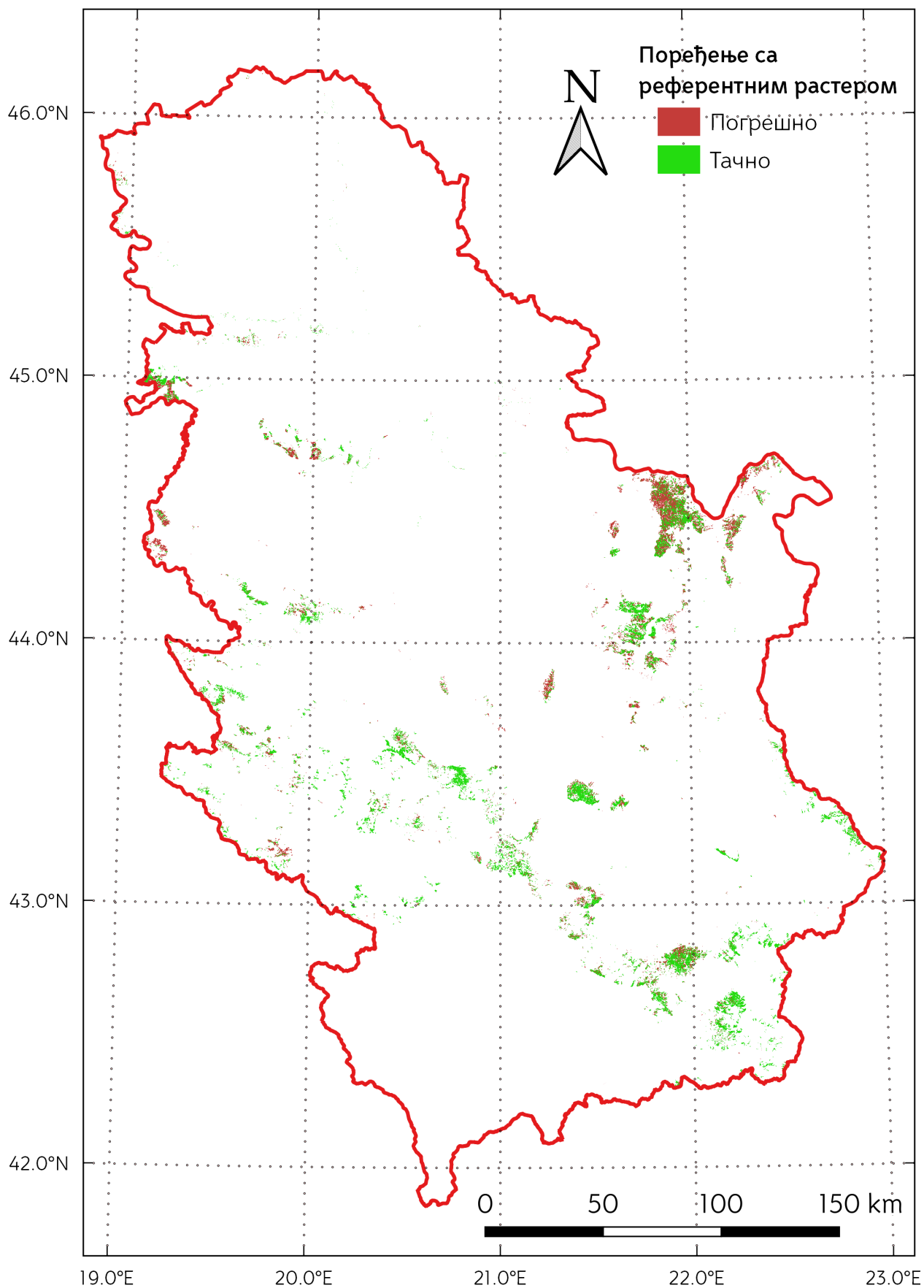
Поређење креираних карата и референтног растера је рађено на нивоу пиксела, а резултати валидације су приказани користећи раније коришћене квантитативне мере перформанси класификационих модела. Ради сагледавања просторног размештаја валидације у односу на референтни растер, креиране су и посебне карте тачно и погрешно класификованих пиксела (Слика 28. и Слика 29.). Карта имплицитне временске дефиниције постиже укупну тачност $OA = 64.02\%$ и капа коефицијент слагања $Kappa = 0.41$, док карта с експлицитном временском дефиницијом има укупну тачност $OA = 60.59\%$ и капа коефицијент слагања $Kappa = 0.37$. Резултујуће матрице конфузије са срачунатим вредностима произвођачеве и корисникове тачности су приказане у наставку (Табела 19. и Табела 20.).

Табела 19. Матрица конфузије карте са имплицитном временском дефиницијом

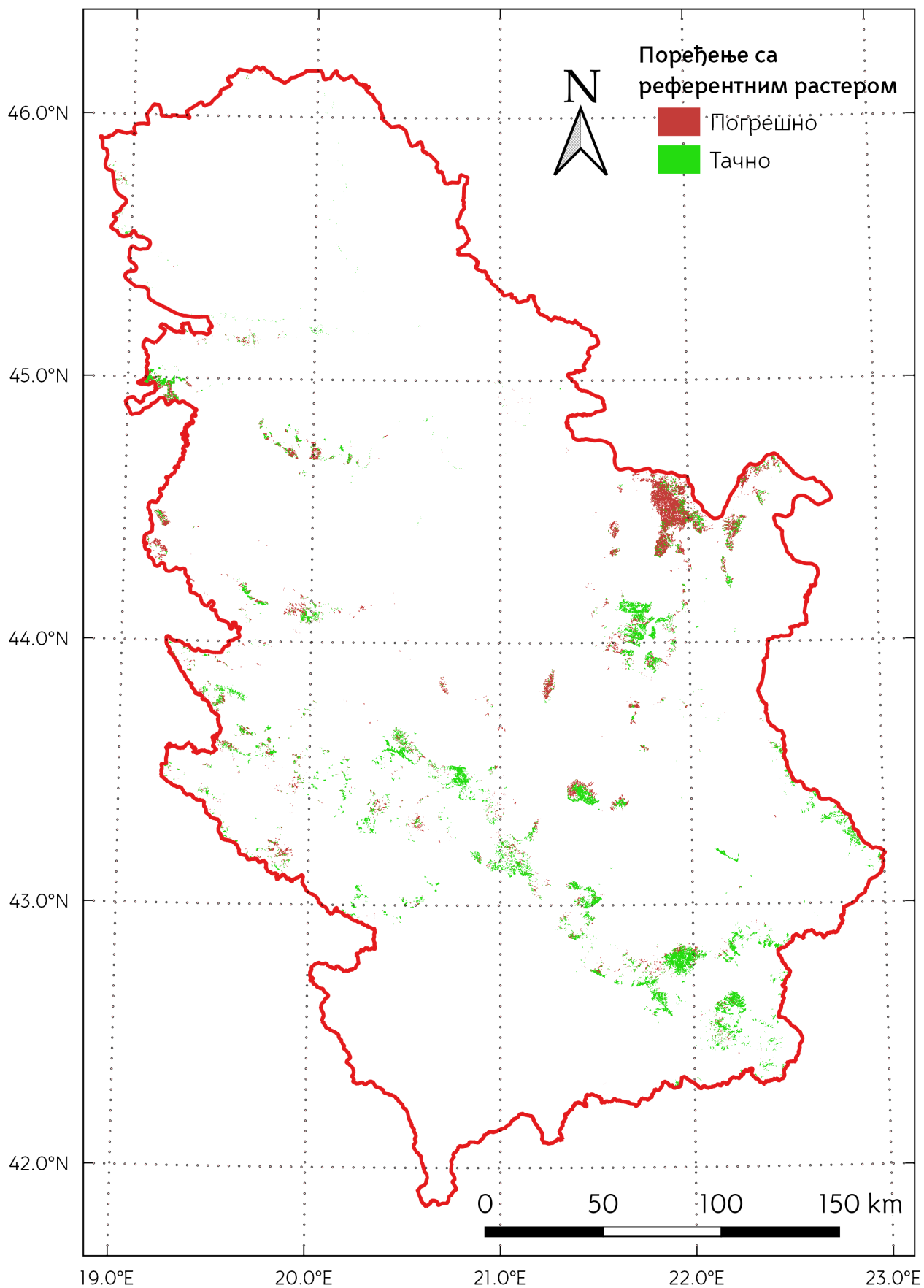
		Референтне вредности							
	Класа	1	2	3	4	5	6	7	Сума
Класификација	1	69 502	75 331	9	11	515	46	57	145 471
	2	92 561	158 073	19 191	50	317	494	0	270 686
	3	539	1 167	360 339	22 862	1 382 651	47 571	9 284	1 824 413
	4	0	0	3 126	6 095	20 226	13 004	872	43 323
	5	139	0	51 984	3 780	2 372 293	12 429	23 731	2 464 356
	6	196	483	1 617	39	2 195	202 295	13 277	220 102
	7	0	0	107	13	11 553	52 168	146 887	210 728
	Сума	162 937	235 054	436 373	32 850	3 789 750	328 007	194 108	5 179 079
	РА [%]	42.66	67.25	82.58	18.55	62.60	61.67	75.67	
	UA [%]	47.78	58.40	19.75	14.07	96.26	91.91	69.70	

Табела 20. Матрица конфузије карте са експлицитном временском дефиницијом

		Референтне вредности							
	Класа	1	2	3	4	5	6	7	Сума
Класификација	1	74 069	60 145	2 726	443	10 991	9 443	478	158 295
	2	88 466	173 645	23 889	1 339	5 793	670	41	293 843
	3	299	1 076	349 796	24 227	1 528 043	71 955	6 851	1 982 247
	4	0	0	1 884	4 412	7 383	4 135	2	17 816
	5	3	0	56 689	2 389	2 233 156	22 967	38 751	2 353 955
	6	100	188	1 357	40	2 954	183 546	27 704	215 889
	7	0	0	32	0	1 430	35 291	120 281	157 034
	Сума	162 937	235 054	436 373	32 850	3 789 750	328 007	194 108	5 179 079
	РА [%]	45.46	73.87	80.16	13.43	58.93	55.96	61.97	
	UA [%]	46.79	59.09	17.65	24.76	94.87	85.02	76.60	



Слика 28. Просторни распоред тачно и погрешно класификованих пиксела карте са имплицитном временском дефиницијом у односу на референтни растер



Слика 29. Просторни распоред тачно и погрешно класификованих пиксела карте са експлицитном временском дефиницијом у односу на референтни растер

Срачунате метрике значајно одступају од метрика добијених крос-валидацијом, при чему су све мере перформанси умањене. Код већине класа шумске вегетације су постигнуте осредње мере перформанси, уз две класе са изразито ниским мерама перформанси. У питању су ксерофилне шуме храстова (3 – *Quercion frainetto*) и ксерофилне шуме белограбића и црног граба (4 – *Ostryo-Carpinion orientalis*). Ниске метрике перформанси код ксерофилних шума белограбића и црног граба су вероватно последица мале количине података ове класе коришћене приликом креирања модела, која очигледно није могла бити превазиђена употребом *SMOTE* алгоритма. Изузетно висока произвођачева и истовремено ниска корисникова тачност код ксерофилних шума храстова значи да модел преферира ову класу. То се може образложити бројем и просторним распоредом полигона шумске вегетације, с обзиром на то да ксерофилне шуме храстова имају највише полигона, уз погоднији просторни распоред у поређењу с осталим листопадним шумским класама. Тиме ова класа има и шири опсег вредности атрибута за класификацију, због којег велики број пиксела завршава у њој. Карактеристичан пример ове ситуације је велики број погрешно класификованих пиксела у подручју Источне Србије, где су мезофилне шуме букве (5 – *Fagion sylvaticae*) погрешно представљене као ксерофилне шуме храстова (3 – *Quercion frainetto*).

Овако велике разлике у срачунатим мерама перформанси на основу крос-валидације и коришћењем независног скупа података отвара питање упоредивости креираних карата шумске вегетације и независног референтног скупа података. Карте шумске вегетације користе типологију доминантно дефинисану узимајући у обзир аспект биодиверзитета. Са друге стране, независни скуп података је прикупљен техникама за инвентуру шума и користи шумарску типологију. Дакле, постоји оправдана сумња да је значајан део одступања последица спроведене конверзије информације о идентитету шумских састојина у усвојене класе типова шумске вегетације. Да постоји озбиљан проблем са конверзијом информација између ова два скупа података сугерише и то што највећа одступања претежно постоје код класа које се јављају као комбинација већег броја шумских врста, као и што се ниске вредности срачунатих метрика поклапају са раније уоченим одступањима површина у односу на инвентуру шума. Очигледно је конверзија шумарске типологије врло осетљиво питање, која мора да обухвати велики број специфичности. Ту се мисли на методолошке и суштинске разлике када је реч о посматрању шумске вегетације из угла управљања шумским ресурсима и шумске вегетације из угла биодиверзитета. То практично значи да се, када је реч о праћењу и

истраживању биодиверзитета, постојећи шумарски подаци морају користити са посебном пажњом.

Ипак, сигурно да ниске мере перформанси нису само последица спроведене конверзије, већ и да одређени део одступања потиче и због недостатака у скупу података коришћеном за креирање класификационих модела. Ту се пре свега мисли на то да поједине класе нису довољно заступљене да би била обухваћена разноликост дефинисаног типа шумске вегетације. Такође, просторна дистрибуција полигона коришћених за креирање класификационих модела се не може сматрати идеалном. Због тога се може очекивати да додатни полигони типова шумске вегетације у подручјима која су слабо покривена (попут јужног дела истраживаног подручја) могу значајно допринети квалитетнијем разликовању класа и тачнијој процени површине и удела заступљености.

Све ово сугерише да коначне карте типова шумске вегетације захтевају додатну валидацију пре него што могу бити пуштене у ширу употребу. Такође, додатна унапређења и повећања мера перформанси су могућа применом различитих поступака. За квалитетнији продукт свакако треба тежити обезбеђивању репрезентативнијег скупа података за креирање класификационог модела. Како је коначни продукт ограничен квалитетом коришћене шумске маске, пажњу треба усмерити и ка изналажењу што квалитетнијег продукта за диференцирање шумског од нешумског подручја. Такође, укључивање додатних извора података и проширење скупа сателитских снимака може бити од користи помоћи, при чему ово има смисла радити тек након што су претходне препоруке испоштоване.

6

ЗАКЉУЧАК

Базе података сателитских опажања Земље су сваки дан богатије захваљујући великом броју активних сензора у оквиру сателитских мисија. Широк дијапазон сателитских сензора обезбеђује различите скупове података, што пружа велике могућности за анализу појава и феномена на Земљиној површини. Велики истраживачки напор је усмерен ка употреби ових извора података као алтернативи теренским методама, са циљем смањења потребне количине времена и рада, али уз очување тачности и богатства информација. Праћење и анализа шумског вегетационог покривача представља област истраживања где су се подаци сателитског осматрања Земље показали посебно корисним. До сада је развијен велики број различитих метода за одређивање обухвата и типова шумске вегетације и различитих биофизичких шумских карактеристика. Ипак, простора за напредак и даље има, а истраживање у оквиру ове дисертације обухвата развој методологије за картирање типова шумске вегетације. Предложена методологија се у основи ослања на издвајање тематских информација применом надзираних непараметарских метода класификације које спадају у групу техника машинског учења. Истраживање је реализовано кроз више корака и обухватало је дефиницију проблема и циљева, прикупљање и припрему података, одређивање оптималних параметара методологије креирањем и валидацијом скупа класификационих модела и, на крају, изналажење и примену оптималног модела.

Сprovedена су четири експеримента за потребе испитивања предложене методологије. Циљеви експеримента обухватају утврђивање применљивости различитих техника машинског учења на картирање типова шумске вегетације, анализу различитих

репрезентација података ради наглашавања просторних и временских компоненти у класификационом моделу, као и примену различитих техника за креирање оптималног скупа података и атрибута од значаја за квалитетно картирање шумске вегетације. Коришћене су три технике машинског учења (стабло одлучивања, случајна шума и градијентно појачавање) за креирање класификационих модела, чије су перформансе изражене и анализирани скупом метрика (укупна тачност, капа коефицијент слагања, произвођачева и корисникова тачност сваке класе) срачунате на основу петоструке крос-валидације са изостављањем комплетних локација. На овај начин је извршено одређивање оптималних параметара предложене методологије, која је затим примењена за израду карте са типовима шумске вегетације за комплетну територију Републике Србије. Усвојена је класификациона шема која обухвата осам карактеристичних типова шумске вегетације овог подручја, а за сваки од њих су обезбеђени репрезентативни полигони. За потребе реализације експеримената и израде карте типова шумске вегетације, формирана је просторна база података коју чини 89 мултиспектралних снимака сателитске мисије *Sentinel-2* из периода јун – октобар 2019. године, подаци висинске представе терена, карта основног земљишног покривача Републике Србије и Коперникус слојеви високе резолуције о шумама. Као координатни референтни систем је коришћена државна пројекција Републике Србије (VTM пројекција, зона 34N), а као просторни оквир је усвојен 20-метарски *Sentinel-2* растерски грид у којем се дистрибуирају подаци ове сателитске мисије.

На основу добијених резултата, све три коришћене технике машинског учења могу бити примењене за картирање типова шумске вегетације. Случајна шума и градијентно појачавање су се подједнако добро показале, а приметно је да стабло одлучивања постиже знатно ниже мере перформанси у односу на њих. Ово је и очекивано, пошто је стабло одлучивања рудиментарнија техника машинског учења, док су преостале две тестиране технике практично њена надградња.

Предложена методологија разликује два основна приступа према томе да ли је временска компонента у класификационом моделу дефинисана имплицитно или експлицитно. Имплицитна репрезентација значи да се сваки временски тренутак/интервал представља скупом независних атрибута у моделу класификације, док експлицитна репрезентација подразумева да се за свако опажање посебно разматра и води један или више временских атрибута (нпр. дан у години, кумулативни

дан у години, датум итд.). Главна последица приступа са експлицитном временском дефиницијом је што на овај начин дефинисан класификациони модел за сваки пиксел подручја истраживања производи скуп излаза, у зависности од броја расположивих сателитских опажања. Последично, приликом одређивања коначне класе шумске вегетације неопходно је извршити агрегацију вишеструких излаза модела на нивоу сваког пиксела подручја од истраживања. Разматрано је више агрегационих функција (правило најчешће присутне класе, проста аритметичка средина и геометријска средина), при чему су се све три функције агрегације показале подједнако успешне. Резултати експеримената указују на више перформансе приступа са имплицитном временском дефиницијом у односу на експлицитну временску дефиницију. При томе ова предност не постоји када се користе само подаци сателитских опажања, већ постају видљиве тек укључивањем додатних извора података. Нумерички, на основу петоструке крос-валидације са изостављањем, имплицитном временском компонентом су постигнуте мере перформанси за око 2% више по питању укупне тачности. Ипак, вредности корисникове и произвођачеве тачности указују да предност није доминантна на нивоу свих класа. Додатно, треба имати на уму и да је имплицитно дефинисан модел осетљивији на недостајуће податке услед облачности, чиме се његова употребљивост у неким подручјима умањује.

Истраживање је обухватило примену више различитих приступа за избор репрезентативних атрибута, ради одређивања оног скупа атрибута који ће обезбедити највише мере перформанси класификационог модела, уз евентуално смањење меморијских и рачунских захтева. Примењена је метода за избор атрибута базирана на корелацији из групе филтер-метода. Вршен је и избор атрибута применом генетских алгоритама и применом метода секвенцијалне претраге (претрага унапред, претрага уназад и претрага планинарењем) из групе омотач-метода. Експерименти су показали да коришћење комплетног скупа атрибута у великом броју случајева резултује највишим метрикама перформанси, а ако то и није случај, онда подскуп атрибута или даје незнатно више или значајно ниже метрике. Очигледно је да се коришћене технике машинског учења добро носе са великим скуповима атрибута, односно да њихови интерни механизми довољно добро рангирају атрибуте по значајности. Ово је последица тога што све три тестиране технике машинског учења користе стабло одлучивања. Приликом изградње стабла одлучивања врши се интерно рангирање репрезентативности атрибута, где важнији атрибути остају при врху стабла, а мање

важни атрибути завршавају близу листова стабла. Другим речима, може се рећи да избор атрибута у комбинацији са ове три технике учења не доноси велика побољшања и да коришћење комплетног скупа атрибута представља „сигурни приступ“ који, чак и ако не даје максималне мере перформанси, оне су онда само незнатно ниже.

Предложена методологија обухвата и анализира и утицај неравномерне заступљености циљних класа у подацима на резултате класификације. Ради превазилажења овог проблема, коришћен је *SMOTE* алгоритам за балансирање класа, креирањем вештачких примерака мањинских класа без подузорковања примерака већинских класа. Очекивано, предност коришћења балансираног скупа података јесте значајно побољшање квалитета препознавања мањински заступљених класа. Истовремено, балансирање класа је значајно повећало меморијске и рачунске захтева неопходне за креирање модела, што је последица повећања броја опажања коришћених за креирање модела.

Наглашавање просторне компоненте у класификационом моделу се такође показало као важан фактор предложене методологије. Ово наглашавање подразумева агрегацију оцењених вредности вероватноћа унутар Муровог суседства величине 3×3 и 5×5 пиксела, при чему су коришћена иста правила агрегације као и код експлицитне временске компоненте. Резултати експеримената показују да укључивање суседства у процедуру класификације помаже да се изађе на крај са шумом у подацима који се при класификацији на нивоу пиксела јавља у виду усамљених пиксела различитих класа у хомогеним шумским подручјима. Ипак, треба напоменути да увођење просторно-временске или просторне агрегације значи да методологија постаје осетљива на промене настале на терену током временског периода покривеног сателитским опажањима и у оквиру дефинисаног суседства. Због тога се не препоручује повећавање временског интервала посматрања и величине суседства више него што је неопходно.

Узевши у обзир све спроведене експерименте, оптимални модели подразумевају коришћење случајне шуме у комбинацији са суседством величине 5×5 и простом аритметичком средином као правилом агрегације. Треба користити балансирани скуп података, при чему се у случају имплицитне временске репрезентација користи комплетан скуп атрибута, док је код експлиците временске репрезентације комплетан скуп умањен за четири атрибута који су процењени мање информативним. Метрике оптималних модела су срачунате петоструком крос-валидацијом са изостављањем

комплетних локација. Оне износе за укупну тачност $OA_{\text{имплицитно}} = 85.43\%$ и $OA_{\text{експлицитно}} = 83.62\%$, односно за капа коефицијенте слагања $Kappa_{\text{имплицитно}} = 0.79$ и $Kappa_{\text{експлицитно}} = 0.78$. Вредности произвођачеве тачности оба приступа се налазе у распону између 65% и 99%, односно вредности корисникове тачности су у распону од 64% до 100%. Добијене метрике се подударају са резултатима других сличних студија заснованих на оптичким сателитским опажањима високе просторне резолуције.

Извршена је валидација креираних карата типова шумске вегетације Републике Србије независним скупом података. Подаци представљају извод из базе података државних шума који обухвата 25 235 полигона шумских одсека о којима се води информација о идентитету репрезентативних шумских састојина. Валидација коришћењем независног скупа података је показала значајна одступања срачунатих метрика од резултата добијених крос-валидацијом. Карта имплицитне временске дефиниције постиже укупну тачност $OA_{\text{имплицитно}} = 64.02\%$ и капа коефицијент слагања $Kappa_{\text{имплицитно}} = 0.41$, док карта с експлицитном временском дефиницијом има укупну тачност $OA_{\text{експлицитно}} = 60.59\%$ и капа коефицијент слагања $Kappa_{\text{експлицитно}} = 0.37$. Срачунате вредности произвођачеве тачности оба приступа се налазе у распону између 13% и 82%, а вредности корисникове тачности су у распону од 14% до 96%. Постоји оправдана сумња да је узрок великог дела одступања последица спроведене конверзије информације о идентитету шумских састојина у усвојене класе типова шумске вегетације. Очигледно је конверзија шумарске типологије врло осетљиво питање, која мора да обухвати методолошке и суштинске разлике када је реч о посматрању шумске вегетације из угла управљања шумским ресурсима и шумске вегетације из угла биодиверзитета. Поред тога, ниже вредности метрика су и последица тога што је просторна расподела репрезентативних полигона коришћених за креирање класификационих модела далеко од идеалне, а показало се и да поједине класе нису заступљене довољно да би била обухваћена њихова разноликост на подручју истраживања. Због тога се може очекивати да додатни полигони типова шумске вегетације у подручјима која су слабо покривена могу значајно допринети квалитетнијем разликовању класа и тачнијој процени површине и удела заступљености. Све ово сугерише да коначне карте типова шумске вегетације захтевају додатну валидацију пре него што могу бити пуштене у ширу употребу.

Због специфичности шумске вегетације, проглашени оптимални параметри методологије се могу сматрати локално прилагођеним, применљивим само за дефинисано подручје истраживања и за обезбеђен скуп података. То значи да је у случају значајне промене било које од ове две карактеристике процедуру одређивања оптималних параметара методологије потребно поновити. Исто тако, важно је напоменути да проглашени модели са оптималним параметрима често постижу веома сличне или само незнатно више мере перформанси у односу на остале креиране моделе. Разлике по питању постигнуте укупне тачности и капа коефицијента слагања су у највећем броју случајева веома мале – незнатне. Са практичног аспекта то значи да за бројне примене избор конкретних параметара методологије нема велики значај, докле год је свака компонента методологије укључена. Могу се очекивати слични коначни резултати без обзира на: примењену технику машинског учења (са изузетком стабла одлучивања), одабран начин временске репрезентације података, наглашавање просторних аспеката и/или избор агрегационе функције. За такве примене, пажњу треба посветити квалитету расположивог скупа података за креирање класификационог модела, сложености подручја истраживања и коришћеној класификационој шеми, јер ће то у крајњој мери и највише утицати на квалитет излазних резултата.

На крају се може закључити да је предложена методологија у сваком случају применљива за картирање типова шумске вегетације у Србији. Предуслови за њену успешну примену обухватају дефинисање класификационе шеме и обезбеђивање репрезентативних полигона за сваку класе шумске вегетације. Једна од идентификованих слабих тачака методологије јесте да поузданост излазних карата зависи од коришћене шумске маске. Поред тога, ограничење које у великој мери снижава перформансе модела је што се сваки пиксел обавезно разврстава у неку од циљних класа. Методологија може бити проширена тако да се у неком кораку врши и квалитетно разврставање шумских и нешумских пиксела. Други поменути недостатак методологије је знатно сложенији. У том случају потенцијално решење може бити анализа вредности оцењених вероватноћа са циљем идентификације пиксела са малом вероватноћом, што може указивати на то да они вероватно не припадају ниједној од усвојених класа. Методологија је тестирана на територији Србије, али има значајан потенцијал да се примени и на другим подручјима уз одговарајућа прилагођавања класификационе шеме и додатне репрезентативне полигоне за креирање модела.

Такође, предложена методологија може бити релативно једноставно прилагођена и неким другим класификационим проблемима, попут идентификације других типова вегетације, картирање земљишног покривача и начина употребе земљишта, односно, структуре предела и елемената зелене инфраструктуре итд. Да би се ове претпоставке потврдиле, неопходна су додатна истраживања.

ЛИТЕРАТУРА

- 1) Aabeyir, R., Adu-Bredu, S., Agyare, W.A., Weir, M.J.C., 2020. Allometric models for estimating aboveground biomass in the tropical woodlands of Ghana, West Africa. *Forest Ecosystems* 7, 41. <https://doi.org/10.1186/s40663-020-00250-3>
- 2) Ahamed, T., Tian, L., Zhang, Y., Ting, K.C., 2011. A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy* 35, 2455–2469. <https://doi.org/10.1016/j.biombioe.2011.02.028>
- 3) Alappat, V.O., Joshi, A.K., Krishnamurthy, Y.V.N., 2011. Tropical Dry Deciduous Forest Stand Variable Estimation Using SAR Data. *J Indian Soc Remote Sens* 39, 583–589. <https://doi.org/10.1007/s12524-011-0118-y>
- 4) Ananto, W.H.G., Putri, A.F.S., Hadi, H.A., Hanum, D.N., Wiryawan, B.K.P., Prabaswara, R.R., Arjasakusuma, S., 2019. Performance of various speckle filter methods in modelling forest aboveground biomass using Sentinel-1 data: case study of Barru Regency, South Sulawesi, in: Sixth Geoinformation Science Symposium. Presented at the Sixth Geoinformation Science Symposium, International Society for Optics and Photonics, p. 113110P. <https://doi.org/10.1117/12.2549036>
- 5) Anderson, M.C., Zolin, C.A., Hain, C.R., Semmens, K., Tugrul Yilmaz, M., Gao, F., 2015. Comparison of satellite-derived LAI and precipitation anomalies over Brazil with a thermal infrared-based Evaporative Stress Index for 2003–2013. *Journal of Hydrology, Drought processes, modeling, and mitigation* 526, 287–302. <https://doi.org/10.1016/j.jhydrol.2015.01.005>
- 6) Andries, A., Morse, S., Murphy, R., Lynch, J., Woolliams, E., Fonweban, J., 2019. Translation of Earth observation data into sustainable development indicators: An analytical framework. *Sustainable Development* 27, 366–376. <https://doi.org/10.1002/sd.1908>
- 7) Aonpong, P., Kasetkasem, T., Rakwatin, P., Kumazawa, I., Chanwimaluang, T., 2016. The Random forest approach for land cover mapping, in: 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES). Presented at the 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), IEEE, Bangkok, Thailand, pp. 1–6. <https://doi.org/10.1109/ICTEmSys.2016.7467112>
- 8) Artiola, J.F., Pepper, I.L., Brusseau, M.L. (Eds.), 2004. Environmental monitoring and characterization. Elsevier Academic Press, Amsterdam ; Boston.
- 9) Askar, Nuthammachot, N., Phairuang, W., Wicaksono, P., Sayektiningsih, T., 2018. Estimating Aboveground Biomass on Private Forest Using Sentinel-2 Imagery. *Journal of Sensors* 2018, e6745629. <https://doi.org/10.1155/2018/6745629>
- 10) Asrat, Z., Taddese, H., Ørka, H.O., Gobakken, T., Burud, I., Næsset, E., 2018. Estimation of Forest Area and Canopy Cover Based on Visual Interpretation of Satellite Images in Ethiopia. *Land* 7, 92. <https://doi.org/10.3390/land7030092>
- 11) Bae, S., Levick, S.R., Heidrich, L., Magdon, P., Leutner, B.F., Wöllauer, S., Serebryanyk, A., Nauss, T., Krzystek, P., Gossner, M.M., Schall, P., Heibl, C., Bäessler, C., Doerfler, I., Schulze, E.-D., Krah, F.-S., Culmsee, H., Jung, K., Heurich, M., Fischer, M., Seibold, S., Thorn, S., Gerlach, T., Hothorn, T., Weisser, W.W., Müller, J., 2019. Radar vision in the mapping of forest biodiversity from space. *Nature Communications* 10, 4757. <https://doi.org/10.1038/s41467-019-12737-x>
- 12) Bayr, C., Gallaun, H., Kleb, U., Kornberger, B., Steinegger, M., Winter, M., 2016. Satellite-based forest monitoring: spatial and temporal forecast of growing index and short-wave infrared band. *Geospatial Health*. <https://doi.org/10.4081/gh.2016.310>

- 13) Beasley, D., Bull, D.R., Martin, R., 1993. An overview of Genetic Algorithms: Pt1, Fundamentals. *University computing* 15, 58–69.
- 14) Beeri, O., Netzer, Y., Munitz, S., Mintz, D.F., Pelta, R., Shilo, T., Horesh, A., Mey-tal, S., 2020. Kc and LAI Estimations Using Optical and SAR Remote Sensing Imagery for Vineyards Plots. *Remote Sensing* 12, 3478. <https://doi.org/10.3390/rs12213478>
- 15) Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., Radeloff, V.C., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences* 114. <https://doi.org/10.1029/2008JG000883>
- 16) Bériaux, E., Waldner, F., Collienne, F., Bogaert, P., Defourny, P., 2015. Maize Leaf Area Index Retrieval from Synthetic Quad Pol SAR Time Series Using the Water Cloud Model. *Remote Sensing* 7, 16204–16225. <https://doi.org/10.3390/rs71215818>
- 17) Berninger, A., Lohberger, S., Stängel, M., Siegert, F., 2018. SAR-Based Estimation of Above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan Using L- and C-Band. *Remote Sensing* 10, 831. <https://doi.org/10.3390/rs10060831>
- 18) Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Baston, D., Rouault, E., Warmerdam, F., Ooms, J., Rundel, C., 2021. rgdal: Bindings for the "Geospatial" Data Abstraction Library.
- 19) Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K.O., Giraudoux, P., Davis, M., Santilli, S., 2020. rgeos: Interface to Geometry Engine - Open Source ('GEOS').
- 20) Bohn, U., Gollub, G., Hettwer, C., Neuhäuslová, Z., Raus, T., Schlüter, H., Weber, H., 2000. Karte der natürlichen Vegetation Europas / Map of the Natural Vegetation of Europe. Maßstab / Scale 1 : 2 500 000. Münster (Landwirtschaftsverlag).
- 21) Borlaf-Mena, I., Santoro, M., Villard, L., Badea, O., Tanase, M.A., 2020. Investigating the Impact of Digital Elevation Models on Sentinel-1 Backscatter and Coherence Observations. *Remote Sensing* 12, 3016. <https://doi.org/10.3390/rs12183016>
- 22) Boyd, D.S., Danson, F.M., 2005. Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography: Earth and Environment* 29, 1–26. <https://doi.org/10.1191/0309133305pp432ra>
- 23) Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- 24) Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification And Regression Trees. Routledge. <https://doi.org/10.1201/9781315139470>
- 25) Caudullo, G., San Miguel, J., Estreguil, C., Rigo, D. de, 2013. Forest landscape in Europe: pattern, fragmentation and connectivity. Publications Office of the European Union, LU.
- 26) Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 321–357. <https://doi.org/10.1613/jair.953>
- 27) Chen, J.M., Chen, X., Ju, W., Geng, X., 2005. Distributed hydrological model for mapping evapotranspiration using remote sensing inputs. *Journal of Hydrology* 305, 15–39. <https://doi.org/10.1016/j.jhydrol.2004.08.029>
- 28) Chen, L., Ren, C., Zhang, B., Wang, Z., 2020. Multi-Sensor Prediction of Stand Volume by a Hybrid Model of Support Vector Machine for Regression Kriging. *Forests* 11, 296. <https://doi.org/10.3390/f11030296>

- 29) Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., implementation), Xgb. contributors (base Xgb., 2021. xgboost: Extreme Gradient Boosting.
- 30) Cheng, K., Wang, J., 2019. Forest-Type Classification Using Time-Weighted Dynamic Time Warping Analysis in Mountain Areas: A Case Study in Southern China. *Forests* 10, 1040. <https://doi.org/10.3390/f10111040>
- 31) Clerici, N., Weissteiner, C.J., Gerard, F., 2012. Exploring the Use of MODIS NDVI-Based Phenology Indicators for Classifying Forest General Habitat Categories. *Remote Sensing* 4, 1781–1803. <https://doi.org/10.3390/rs4061781>
- 32) Cleugh, H.A., Leuning, R., Mu, Q., Running, S.W., 2007. Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment* 106, 285–304. <https://doi.org/10.1016/j.rse.2006.07.007>
- 33) Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 37–46. <https://doi.org/10.1177/001316446002000104>
- 34) Copernicus Land Monitoring Service: HRL Forest 2018 Product User Manual, 2020. . European Environment Agency.
- 35) Council of Europe, Berne, 1979. Convention on the Conservation of European Wildlife and Natural Habitats, CETS 104.
- 36) Council of the European Union, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora, 92/43/EEC.
- 37) Daba, D.E., Soromessa, T., 2019. The accuracy of species-specific allometric equations for estimating aboveground biomass in tropical moist montane forests: case study of *Albizia grandibracteata* and *Trichilia dregeana*. *Carbon Balance and Management* 14, 18. <https://doi.org/10.1186/s13021-019-0134-8>
- 38) Darvishzadeh, R., Atzberger, C., Skidmore, A.K., Abkar, A.A., 2009. Leaf Area Index derivation from hyperspectral vegetation indices and the red edge position. *International Journal of Remote Sensing* 30, 6199–6218. <https://doi.org/10.1080/01431160902842342>
- 39) Davies, C.E., Moss, D., 2002. EUNIS Habitat Classification, February 2002.
- 40) De Jong, K.A., 1975. An analysis of the behavior of a class of genetic adaptive systems. (phd). University of Michigan, USA.
- 41) Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., Ritchie, S., Ren, K., Tan, X., Saporta, R., Seiskari, O., Dong, X., Lang, M., Iwasaki, W., Wenchel, S., Broman, K., Schmidt, T., Arenburg, D., Smith, E., Cocquemas, F., Gomez, M., Chataignon, P., Blaser, N., Selivanov, D., Riabushenko, A., Lee, C., Groves, D., Possenriede, D., Parages, F., Toth, D., Yaramaz-David, M., Perumal, A., Sams, J., Morgan, M., Quinn, M., @javrucebo, @marc-outins, Storey, R., Saraswat, M., Jacob, M., Schubmehl, M., Vaughan, D., Hocking, T., Silvestri, L., Barrett, T., Hester, J., Damico, A., Freundt, S., Simons, D., Andrade, E.S. de, Miller, C., Meldgaard, J.P., Tlapak, V., Ushey, K., Eddelbuettel, D., Schwen, B., 2021. data.table: Extension of "data.frame."
- 42) Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment, The Sentinel Missions - New Opportunities for Science* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- 43) Duchemin, B., Hadria, R., Erraki, S., Boulet, G., Maisongrande, P., Chehbouni, A., Escadafal, R., Ezzahar, J., Hoedjes, J.C.B., Kharrou, M.H., Khabba, S., Mougnot, B., Olioso, A., Rodriguez, J.-C., Simonneaux, V., 2006. Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index

- and remotely-sensed vegetation indices. *Agricultural Water Management* 79, 1–27. <https://doi.org/10.1016/j.agwat.2005.02.013>
- 44) El Aboudi, N., Benhlima, L., 2016. Review on wrapper feature selection approaches, in: 2016 International Conference on Engineering & MIS (ICEMIS). Presented at the 2016 International Conference on Engineering & MIS (ICEMIS), IEEE, Agadir, Morocco, pp. 1–5. <https://doi.org/10.1109/ICEMIS.2016.7745366>
 - 45) EU-DEM Statistical Validation, 2014.
 - 46) EU-Hydro / EU-DEM Upgrade - User Manual (No. C4EO17), 2015.
 - 47) European Environmental Agency, 2007. European Forest Types: Categories and Types for Sustainable Forest Management Reporting and Policy. (No. EEA Technical Report No 9/2006).
 - 48) Fagan, M., DeFries, R., 2009. Measurement and Monitoring of the World's Forests: A Review and Summary of Remote Sensing Technical Capability, 2009–2015. Resources for the Future.
 - 49) Fassnacht, F.E., Poblete-Olivares, J., Rivero, L., Lopatin, J., Ceballos-Comisso, A., Galleguillos, M., 2021. Using Sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types. *International Journal of Applied Earth Observation and Geoinformation* 94, 102236. <https://doi.org/10.1016/j.jag.2020.102236>
 - 50) Ferreira, B., Iten, M., Silva, R.G., 2020. Monitoring sustainable development by means of earth observation data and machine learning: a review. *Environ Sci Eur* 32, 120. <https://doi.org/10.1186/s12302-020-00397-4>
 - 51) Ferreira, M.P., Wagner, F.H., Aragão, L.E.O.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2019. Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 149, 119–131. <https://doi.org/10.1016/j.isprsjprs.2019.01.019>
 - 52) Franklin, S.E., Blodgett, C.F., 1993. An example of satellite multisensor data fusion. *Computers & Geosciences* 19, 577–583. [https://doi.org/10.1016/0098-3004\(93\)90083-H](https://doi.org/10.1016/0098-3004(93)90083-H)
 - 53) Fukarek, P., Jovanović, B., 1983. Karta prirodne potencijalne vegetacije SFR Jugoslavije 1:1.000.000.. - Naučno veće Vegetacijske karte Jugoslavije. Naučno veće Vegetacijske karte Jugoslavije, Šumarski fakultet Univerziteta Kiril i Metodije u Skopju.
 - 54) Galiatsatos, N., Donoghue, D.N.M., Watt, P., Bholanath, P., Pickering, J., Hansen, M.C., Mahmood, A.R.J., 2020. An Assessment of Global Forest Change Datasets for National Forest Monitoring and Reporting. *Remote Sensing* 12, 1790. <https://doi.org/10.3390/rs12111790>
 - 55) Ganz, S., Adler, P., Kändler, G., 2020. Forest Cover Mapping Based on a Combination of Aerial Images and Sentinel-2 Satellite Data Compared to National Forest Inventory Data. *Forests* 11, 1322. <https://doi.org/10.3390/f11121322>
 - 56) Gao, Y., Zhou, X., Wang, Q., Wang, C., Zhan, Z., Chen, L., Yan, J., Qu, R., 2013. Vegetation net primary productivity and its response to climate change during 2001–2008 in the Tibetan Plateau. *Science of The Total Environment* 444, 356–362. <https://doi.org/10.1016/j.scitotenv.2012.12.014>
 - 57) García-Llamas, P., Calvo, L., Álvarez-Martínez, J.M., Suárez-Seoane, S., 2016. Using remote sensing products to classify landscape. A multi-spatial resolution approach. *International Journal of Applied Earth Observation and Geoinformation* 50, 95–105. <https://doi.org/10.1016/j.jag.2016.03.010>
 - 58) Gargiulo, M., Dell'Aglia, D.A.G., Iodice, A., Riccio, D., Ruello, G., 2020. Integration of Sentinel-1 and Sentinel-2 Data for Land Cover Mapping Using W-Net. *Sensors (Basel)* 20. <https://doi.org/10.3390/s20102969>

- 59) Ghosh, S.M., Behera, M.D., Paramanik, S., 2020. Canopy Height Estimation Using Sentinel Series Images through Machine Learning Models in a Mangrove Forest. *Remote Sensing* 12, 1519. <https://doi.org/10.3390/rs12091519>
- 60) Global Forest Resources Assessment 2020, 2020. . FAO. <https://doi.org/10.4060/ca8753en>
- 61) Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. Addison-Wesley Longman Publishing Co., Inc., USA.
- 62) Gourlet-Fleury, S., Mortier, F., Fayolle, A., Baya, F., Ouédraogo, D., Bénédet, F., Picard, N., 2013. Tropical forest recovery from logging: a 24 year silvicultural experiment from Central Africa. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120302. <https://doi.org/10.1098/rstb.2012.0302>
- 63) Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment* 251, 112103. <https://doi.org/10.1016/j.rse.2020.112103>
- 64) Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing* 11, 1197. <https://doi.org/10.3390/rs11101197>
- 65) Gschwantner, T., Lanz, A., Vidal, C., Bosela, M., Di Cosmo, L., Fridman, J., Gasparini, P., Kuliešis, A., Tomter, S., Schadauer, K., 2016. Comparison of methods used in European National Forest Inventories for the estimation of volume increment: towards harmonisation. *Annals of Forest Science* 73, 807–821. <https://doi.org/10.1007/s13595-016-0554-5>
- 66) Gu, H., Dai, L., Wu, G., Xu, D., Wang, S., Wang, H., 2006. Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data. *SCI CHINA SER E* 49, 54–62. <https://doi.org/10.1007/s11431-006-8107-z>
- 67) Hall, M.A., 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, in: *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 359–366.
- 68) Hall, M.A., Smith, L.A., 1999. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper, in: *Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference*. AAAI Press, pp. 235–239.
- 69) Hansen, E.H., Ene, L.T., Mauya, E.W., Patočka, Z., Mikita, T., Gobakken, T., Næsset, E., 2017. Comparing Empirical and Semi-Empirical Approaches to Forest Biomass Modelling in Different Biomes Using Airborne Laser Scanner Data. *Forests* 8, 170. <https://doi.org/10.3390/f8050170>
- 70) Hastie, T., Tibshirani, R., Friedman, J., 2009. Boosting and Additive Trees, in: Hastie, T., Tibshirani, R., Friedman, J. (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics. Springer, New York, NY, pp. 337–387. https://doi.org/10.1007/978-0-387-84858-7_10
- 71) Hastings, D.A., Emery, W.J., 1992. The advanced very high resolution radiometer (AVHRR) - A brief reference guide. *Photogrammetric Engineering and Remote Sensing* 58, 1183–1188.
- 72) He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. *Water Resources Research* 52, 8217–8237. <https://doi.org/10.1002/2016WR019034>
- 73) Hijmans, R.J., Etten, J. van, Sumner, M., Cheng, J., Baston, D., Bevan, A., Bivand, R., Busetto, L., Canty, M., Fasoli, B., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J.A., Hiemstra, P., Hingee, K., Geosciences, I. for M.A., Karney, C., Mattiuzzi, M., Mosher, S., Naimi, B., Nowosad,

- J., Pebesma, E., Lamigueiro, O.P., Racine, E.B., Rowlingson, B., Shortridge, A., Venables, B., Wueest, R., 2021. raster: Geographic Data Analysis and Modeling.
- 74) Hirschmugl, M., Deutscher, J., Sobe, C., Bouvet, A., Mermoz, S., Schardt, M., 2020. Use of SAR and Optical Time Series for Tropical Forest Disturbance Mapping. *Remote Sensing* 12, 727. <https://doi.org/10.3390/rs12040727>
 - 75) Hirschmugl, M., Ofner, M., Raggam, J., Schardt, M., 2007. Single tree detection in very high resolution remote sensing data. *Remote Sensing of Environment, ForestSAT Special Issue* 110, 533-544. <https://doi.org/10.1016/j.rse.2007.02.029>
 - 76) Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA.
 - 77) Huang, X., Ziniti, B., Torbick, N., Ducey, M.J., 2018. Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. *Remote Sensing* 10, 1424. <https://doi.org/10.3390/rs10091424>
 - 78) Huete, A.R., 2012. Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass* 6, 513-532. <https://doi.org/10.1111/j.1749-8198.2012.00507.x>
 - 79) Huete, A.R., 2004. 11 - REMOTE SENSING FOR ENVIRONMENTAL MONITORING, in: Artiola, J.F., Pepper, I.L., Brusseau, M.L. (Eds.), *Environmental Monitoring and Characterization*. Academic Press, Burlington, pp. 183-206. <https://doi.org/10.1016/B978-012064477-3/50013-8>
 - 80) Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing* 4, 2661-2693. <https://doi.org/10.3390/rs4092661>
 - 81) Iverson, L.R., Graham, R.L., Cook, E.A., 1989. Applications of satellite remote sensing to forested ecosystems. *Landscape Ecol* 3, 131-143. <https://doi.org/10.1007/BF00131175>
 - 82) Jensen, J., 2006. *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd edition. ed. Pearson, Upper Saddle River, NJ.
 - 83) Jensen, J.R., Qiu, F., Ji, M., 1999. Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *Int. J. Remote* 2805-2822.
 - 84) Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M.R., Kuemmerle, T., Meyfroidt, P., Mitchard, E.T.A., Reiche, J., Ryan, C.M., Waske, B., 2016. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing* 8, 70. <https://doi.org/10.3390/rs8010070>
 - 85) Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Zhengming Wan, Huete, A.R., Leeuwen, W. van, Wolfe, R.E., Giglio, L., Muller, J., Lewis, P., Barnsley, M.J., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing* 36, 1228-1249. <https://doi.org/10.1109/36.701075>
 - 86) Kassambara, A., 2020. ggpubr: "ggplot2" Based Publication Ready Plots.
 - 87) Kim, S.-R., Lee, W.-K., Kwak, D.-A., Biging, G.S., Gong, P., Lee, J.-H., Cho, H.-K., 2011. Forest Cover Classification by Optimal Segmentation of High Resolution Satellite Imagery. *Sensors* 11, 1943-1958. <https://doi.org/10.3390/s110201943>
 - 88) Kirchhoefer, M., Schumacher, J., Adler, P., Kändler, G., 2017. Considerations towards a Novel Approach for Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories. *Forests* 8, 239. <https://doi.org/10.3390/f8070239>

- 89) Kivinen, J., Warmuth, M.K., 1999. Averaging Expert Predictions, in: Fischer, P., Simon, H.U. (Eds.), *Computational Learning Theory, Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 153–167. https://doi.org/10.1007/3-540-49097-3_13
- 90) Kivinen, S., Koivisto, E., Keski-Saari, S., Poikolainen, L., Tanhuanpää, T., Kuzmin, A., Viinikka, A., Heikkinen, R.K., Pykälä, J., Virkkala, R., Vihervaara, P., Kumpula, T., 2020. A keystone species, European aspen (*Populus tremula* L.), in boreal forests: Ecological role, knowledge needs and mapping using remote sensing. *Forest Ecology and Management* 462, 118008. <https://doi.org/10.1016/j.foreco.2020.118008>
- 91) Koju, U., Zhang, J., Gilani, H., 2017. Exploring multi-scale forest above ground biomass estimation with optical remote sensing imageries. *IOP Conf. Ser.: Earth Environ. Sci.* 57, 012011. <https://doi.org/10.1088/1755-1315/57/1/012011>
- 92) Kovačević, J., Cvijetinović, Ž., Lakušić, D., Kuzmanović, N., Šinžar-Sekulić, J., Mitrović, M., Stančić, N., Brodić, N., Mihajlović, D., 2020. Spatio-Temporal Classification Framework for Mapping Woody Vegetation from Multi-Temporal Sentinel-2 Imagery. *Remote Sensing* 12, 2845. <https://doi.org/10.3390/rs12172845>
- 93) Kuenzer, C., Ottinger, M., Wegmann, M., Guo, H., Wang, C., Zhang, J., Dech, S., Wikelski, M., 2014. Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks. *International Journal of Remote Sensing* 35, 6599–6647. <https://doi.org/10.1080/01431161.2014.964349>
- 94) Laporte, N.T., Lin, T.S., 2003. Monitoring logging in the tropical forest of Republic of Congo with Landsat imagery, in: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. Presented at the IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), pp. 2565–2567 vol.4. <https://doi.org/10.1109/IGARSS.2003.1294510>
- 95) Lechner, A.M., Foody, G.M., Boyd, D.S., 2020. Applications in Remote Sensing to Forest Ecology and Management. *One Earth* 2, 405–412. <https://doi.org/10.1016/j.oneear.2020.05.001>
- 96) Li, J., Heap, A.D., Potter, A., Daniell, J.J., 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software* 26, 1647–1659. <https://doi.org/10.1016/j.envsoft.2011.07.004>
- 97) Li, L., Wang, Y.-P., Beringer, J., Shi, H., Cleverly, J., Cheng, L., Eamus, D., Huete, A., Hutley, L., Lu, X., Piao, S., Zhang, L., Zhang, Y., Yu, Q., 2017. Responses of LAI to rainfall explain contrasting sensitivities to carbon uptake between forest and non-forest ecosystems in Australia. *Scientific Reports* 7, 11720. <https://doi.org/10.1038/s41598-017-11063-w>
- 98) Li, Y., Kang, W., Han, Y., Song, Y., 2018. Spatial and temporal patterns of microclimates at an urban forest edge and their management implications. *Environ Monit Assess* 190, 93. <https://doi.org/10.1007/s10661-017-6430-4>
- 99) Liang, S., Wang, J., 2020. Advanced remote sensing: terrestrial information extraction and applications.
- 100) Liu, H., 2010. Feature Selection, in: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning*. Springer US, Boston, MA, pp. 402–406. https://doi.org/10.1007/978-0-387-30164-8_306
- 101) Liu, K., Zhou, Q., Wu, W., Xia, T., Tang, H., 2016. Estimating the crop leaf area index using hyperspectral remote sensing. *Journal of Integrative Agriculture* 15, 475–491. [https://doi.org/10.1016/S2095-3119\(15\)61073-5](https://doi.org/10.1016/S2095-3119(15)61073-5)
- 102) Liu, Y., Gong, W., Hu, X., Gong, J., 2018. Forest Type Identification with Random Forest Using Sentinel-1A, Sentinel-2A, Multi-Temporal Landsat-8 and DEM Data. *Remote Sensing* 10, 946. <https://doi.org/10.3390/rs10060946>

- 103) Lucas, R., Blonda, P., Bunting, P., Jones, G., Inglada, J., Arias, M., Kosmidou, V., Petrou, Z.I., Manakos, I., Adamo, M., Charnock, R., Tarantino, C., Múcher, C.A., Jongman, R.H.G., Kramer, H., Arvor, D., Honrado, J.P., Mairota, P., 2015. The Earth Observation Data for Habitat Monitoring (EODHaM) system. *International Journal of Applied Earth Observation and Geoinformation* 37, 17–28. <https://doi.org/10.1016/j.jag.2014.10.011>
- 104) Lutz, D.A., Washington-Allen, R.A., Shugart, H.H., 2008. Remote sensing of boreal forest biophysical and inventory parameters: a review. *Canadian Journal of Remote Sensing* 34, 28.
- 105) Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 277–293. <https://doi.org/10.1016/j.isprsjprs.2017.06.001>
- 106) Ma, W., Gong, C., Hu, Y., Meng, P., Xu, F., 2013. The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake, in: *International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications*. Presented at the International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications, International Society for Optics and Photonics, p. 89101G. <https://doi.org/10.1117/12.2034457>
- 107) Madonsela, S., Cho, M.A., Ramoelo, A., Mutanga, O., Naidoo, L., 2018. Estimating tree species diversity in the savannah using NDVI and woody canopy cover. *International Journal of Applied Earth Observation and Geoinformation* 66, 106–115. <https://doi.org/10.1016/j.jag.2017.11.005>
- 108) Manninen, T., Stenberg, P., Rautiainen, M., 2003. Estimation of boreal forest LAI using C-band SAR, in: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. Presented at the IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. (IEEE Cat. No.03CH37477), IEEE, Toulouse, France, pp. 1631–1633. <https://doi.org/10.1109/IGARSS.2003.1294199>
- 109) Marrs, J., Ni-Meister, W., 2019. Machine Learning Techniques for Tree Species Classification Using Co-Registered LiDAR and Hyperspectral Data. *Remote Sensing* 11, 819. <https://doi.org/10.3390/rs11070819>
- 110) Martone, M., Rizzoli, P., Wecklich, C., González, C., Bueso-Bello, J.-L., Valdo, P., Schulze, D., Zink, M., Krieger, G., Moreira, A., 2018. The global forest/non-forest map from TanDEM-X interferometric SAR data. *Remote Sensing of Environment* 205, 352–373. <https://doi.org/10.1016/j.rse.2017.12.002>
- 111) Matasci, G., Coops, N.C., Williams, D.A.R., Page, N., 2018. Mapping tree canopies in urban environments using airborne laser scanning (ALS): a Vancouver case study. *Forest Ecosystems* 5, 31. <https://doi.org/10.1186/s40663-018-0146-y>
- 112) Mathieu, P.-P., Aubrecht, C. (Eds.), 2018. *Earth Observation Open Science and Innovation*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-65633-5>
- 113) Matlack, G.R., 1994. Vegetation Dynamics of the Forest Edge -- Trends in Space and Successional Time. *Journal of Ecology* 82, 113–123. <https://doi.org/10.2307/2261391>
- 114) Mazza, A., Sica, F., Rizzoli, P., Scarpa, G., 2019. TanDEM-X Forest Mapping Using Convolutional Neural Networks. *Remote Sensing* 11, 2980. <https://doi.org/10.3390/rs11242980>
- 115) Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing* 166, 333–346. <https://doi.org/10.1016/j.isprsjprs.2020.05.013>

- 116) Miah, Md.D., Islam, K.N., Kabir, Md.H., Koike, M., 2020. Allometric models for estimating aboveground biomass of selected homestead tree species in the plain land Narsingdi district of Bangladesh. *Trees, Forests and People* 2, 100035. <https://doi.org/10.1016/j.tfp.2020.100035>
- 117) Mohammadi, J., Joibary, S.S., Yaghmaee, F., Mahiny, A.S., 2010. Modelling forest stand volume and tree density using Landsat ETM+ data. *International Journal of Remote Sensing* 31, 2959–2975. <https://doi.org/10.1080/01431160903140811>
- 118) Müller-Wilm, U., Louis, J., Richter, R., Gascon, F., Niezette, M., 2013. Sentinel-2 Level-2A Prototype Processor: Architecture, Algorithms and First Results.
- 119) Nelson, R., Horning, N., 1993. Cover Forest/non-forest classification of Madagascar from AVHRR data. *International Journal of Remote Sensing* 14, 1445–1446. <https://doi.org/10.1080/01431169308953979>
- 120) Nguyen Trong, H., Nguyen, T.D., Kappas, M., 2020. Land Cover and Forest Type Classification by Values of Vegetation Indices and Forest Structure of Tropical Lowland Forests in Central Vietnam. *International Journal of Forestry Research* 2020, e8896310. <https://doi.org/10.1155/2020/8896310>
- 121) Novo-Fernández, A., Barrio-Anta, M., Recondo, C., Cámara-Obregón, A., López-Sánchez, C.A., 2019. Integration of National Forest Inventory and Nationwide Airborne Laser Scanning Data to Improve Forest Yield Predictions in North-Western Spain. *Remote Sensing* 11, 1693. <https://doi.org/10.3390/rs11141693>
- 122) Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148, 42–57. <https://doi.org/10.1016/j.rse.2014.02.015>
- 123) Olsen, L.R., 2021. *groupdata2: Creating Groups from Data*.
- 124) Palubinskas, G., Makarau, A., Tao, J., 2011. Fusion of optical and radar data for the extraction of higher quality information. Presented at the 4th TerraSAR-X Science Team Meeting, DLR, Oberpfaffenhofen, Germany, pp. 1–9.
- 125) Pause, M., Schweitzer, C., Rosenthal, M., Keuck, V., Bumberger, J., Dietrich, P., Heurich, M., Jung, A., Lausch, A., 2016. In Situ/Remote Sensing Integration to Assess Forest Health—A Review. *Remote Sensing* 8, 471. <https://doi.org/10.3390/rs8060471>
- 126) Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J., Lindgren, F., O'Brien, J., O'Rourke, J., 2021. *sp: Classes and Methods for Spatial Data*.
- 127) Pekkarinen, A., Reithmaier, L., Strobl, P., 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64, 171–183. <https://doi.org/10.1016/j.isprsjprs.2008.09.004>
- 128) Pinto, S.R.R., Mendes, G., Santos, A.M.M., Dantas, M., Tabarelli, M., Melo, F.P.L., 2010. Landscape Attributes Drive Complex Spatial Microclimate Configuration of Brazilian Atlantic Forest Fragments. *Tropical Conservation Science* 3, 389–402. <https://doi.org/10.1177/194008291000300404>
- 129) Pipia, L., Muñoz-Marí, J., Amin, E., Belda, S., Camps-Valls, G., Verrelst, J., 2019. Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. *Remote Sensing of Environment* 235, 111452. <https://doi.org/10.1016/j.rse.2019.111452>
- 130) Proulx, R., Fahrig, L., 2010. Detecting human-driven deviations from trajectories in landscape composition and configuration. *Landscape Ecol* 25, 1479–1487. <https://doi.org/10.1007/s10980-010-9523-9>

- 131) Pulella, A., Aragão Santos, R., Sica, F., Posovszky, P., Rizzoli, P., 2020. Multi-Temporal Sentinel-1 Backscatter and Coherence for Rainforest Mapping. *Remote Sensing* 12, 847. <https://doi.org/10.3390/rs12050847>
- 132) QGIS Development Team, 2021. QGIS Geographic Information System. QGIS Association.
- 133) Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Elsevier. <https://doi.org/10.1016/C2009-0-27846-9>
- 134) Quinlan, J.R., 1986. Induction of decision trees. *Mach Learn* 1, 81-106. <https://doi.org/10.1007/BF00116251>
- 135) R Core Team, 2020. R: A Language and Environment for Statistical Computing.
- 136) Ranson, K.J., Kovacs, K., Sun, G., Kharuk, V.I., 2003. Disturbance recognition in the boreal forest using radar and Landsat-7. *Canadian Journal of Remote Sensing* 29, 271-285. <https://doi.org/10.5589/m02-096>
- 137) Richard, D., Spyropoulou, R., Poncet, L., Evans, D., Ichter, J., Pereira Martins, I., 2014. Terrestrial habitat mapping in Europe: an overview. European Environment Agency, LU.
- 138) Rignot, E.J.M., Williams, C.L., Way, J., Viereck, L.A., 1994. Mapping of forest types in Alaskan boreal forests using SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 32, 1051-1059. <https://doi.org/10.1109/36.312893>
- 139) Romanski, P., Kotthoff, L., Schratz, P., 2021. FSelector: Selecting Attributes.
- 140) RStudio Team, 2021. RStudio: Integrated Development Environment for R. RStudio, PBC.
- 141) Ruiz-Ramos, J., Marino, A., Boardman, C., Suarez, J., 2020. Continuous Forest Monitoring Using Cumulative Sums of Sentinel-1 Timeseries. *Remote Sensing* 12, 3061. <https://doi.org/10.3390/rs12183061>
- 142) Saleska, S.R., Wu, J., Guan, K., Araujo, A.C., Huete, A., Nobre, A.D., Restrepo-Coupe, N., 2016. Dry-season greening of Amazon forests. *Nature* 531, E4-E5. <https://doi.org/10.1038/nature16457>
- 143) Samardžić-Petrović, M., 2014. Predicting land use change with data-driven models.
- 144) Samardžić-Petrović, M., Dragičević, S., Kovačević, M., Bajat, B., 2016. Modeling Urban Land Use Changes Using Support Vector Machines. *Transactions in GIS* 20, 718-734. <https://doi.org/10.1111/tgis.12174>
- 145) Samardžić-Petrović, M., Kovačević, M., Bajat, B., Dragičević, S., 2017. Machine Learning Techniques for Modelling Short Term Land-Use Change. *ISPRS International Journal of Geo-Information* 6, 387. <https://doi.org/10.3390/ijgi6120387>
- 146) Schepaschenko, D., See, L., Lesiv, M., Bastin, J.-F., Mollicone, D., Tsendbazar, N.-E., Bastin, L., McCallum, I., Laso Bayas, J.C., Baklanov, A., Perger, C., Dürauer, M., Fritz, S., 2019. Recent Advances in Forest Observation with Visual Interpretation of Very High-Resolution Imagery. *Surv Geophys* 40, 839-862. <https://doi.org/10.1007/s10712-019-09533-z>
- 147) Scrucca, L., 2021. GA: Genetic Algorithms.
- 148) Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random Forest Spatial Interpolation. *Remote Sensing* 12, 1687. <https://doi.org/10.3390/rs12101687>
- 149) Shackleton, S., Chinyimba, A., Hebinck, P., Shackleton, C., Kaoma, H., 2015. Multiple benefits and values of trees in urban landscapes in two towns in northern South Africa. *Landscape and Urban Planning* 136, 76-86. <https://doi.org/10.1016/j.landurbplan.2014.12.004>
- 150) Sheeren, D., Fauvel, M., Josipović, V., Lopes, M., Planque, C., Willm, J., Dejoux, J.-F., 2016. Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing* 8, 734. <https://doi.org/10.3390/rs8090734>

- 151) Siriseriwan, W., 2019. smotefamily: A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE.
- 152) Soegaard, H., 1999. Fluxes of carbon dioxide, water vapour and sensible heat in a boreal agricultural area of Sweden – scaled from canopy to landscape level. *Agricultural and Forest Meteorology* 98–99, 463–478. [https://doi.org/10.1016/S0168-1923\(99\)00116-1](https://doi.org/10.1016/S0168-1923(99)00116-1)
- 153) Stankevich, S.A., Kozlova, A.A., Piestova, I.O., Lubskyi, M.S., 2017. Leaf area index estimation of forest using sentinel-1 C-band SAR data, in: 2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS). Presented at the 2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS), pp. 253–256. <https://doi.org/10.1109/MRRS.2017.8075075>
- 154) Straub, C.L., Koontz, S.R., Loomis, J.B., 2019. Economic valuation of landsat imagery (USGS Numbered Series No. 2019-1112), Economic valuation of landsat imagery, Open-File Report. U.S. Geological Survey, Reston, VA. <https://doi.org/10.3133/ofr20191112>
- 155) Sumner, M.D., 2020. spex: Spatial Extent Tools.
- 156) Sun, A.Y., Scanlon, B.R., 2019. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environ. Res. Lett.* 14, 073001. <https://doi.org/10.1088/1748-9326/ab1b7d>
- 157) Tanaka, S., Takahashi, T., Nishizono, T., Kitahara, F., Saito, H., Iehara, T., Kodani, E., Awaya, Y., 2015. Stand Volume Estimation Using the k-NN Technique Combined with Forest Inventory Data, Satellite Image Data and Additional Feature Variables. *Remote Sensing* 7, 378–394. <https://doi.org/10.3390/rs70100378>
- 158) Tesemma, Z.K., Wei, Y., Peel, M.C., Western, A.W., 2015. The effect of year-to-year variability of leaf area index on Variable Infiltration Capacity model performance and simulation of runoff. *Advances in Water Resources* 83, 310–322. <https://doi.org/10.1016/j.advwatres.2015.07.002>
- 159) The United Nations, Rio de Janeiro, 1992. Convention on Biological Diversity.
- 160) Therneau, T., Atkinson, B., port, B.R. (producer of the initial R., maintainer 1999-2017), 2019. rpart: Recursive Partitioning and Regression Trees.
- 161) Tianyang, D., Jian, Z., Sibin, G., Ying, S., Jing, F., 2018. Single-Tree Detection in High-Resolution Remote-Sensing Images Based on a Cascade Neural Network. *ISPRS International Journal of Geo-Information* 7, 367. <https://doi.org/10.3390/ijgi7090367>
- 162) Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46, 234–240. <https://doi.org/10.2307/143141>
- 163) Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (Eds.), 2010. National Forest Inventories: Pathways for Common Reporting. Springer Netherlands. <https://doi.org/10.1007/978-90-481-3233-1>
- 164) Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment* 112, 1982–1999. <https://doi.org/10.1016/j.rse.2007.03.032>
- 165) Torgo, L., 2010. Data Mining with R, learning with case studies. Chapman and Hall/CRC.
- 166) Tsendbazar, N.-E., Tarko, A., Li, L., Herold, M., Lesiv, M., Fritz, S., Maus, V., 2020. Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Validation Report. Zenodo. <https://doi.org/10.5281/zenodo.3938974>
- 167) Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8, 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)

- 168) Vasiljević, N., Radić, B., Šljukić, B., Ristić, R., 2016. Landscape Planning and Green Infrastructure in Serbia: From National to Belgrade City Planning. Proceedings of the Fábos Conference on Landscape and Greenway Planning 5.
- 169) Vastaranta, M., Niemi, M., Karjalainen, M., Peuhkurinen, J., Kankare, V., Hyyppä, J., Holopainen, M., 2014. Prediction of Forest Stand Attributes Using TerraSAR-X Stereo Imagery. *Remote Sensing* 6, 3227–3246. <https://doi.org/10.3390/rs6043227>
- 170) Wang, H., Ma, C., Zhou, L., 2009. A Brief Review of Machine Learning and Its Application, in: 2009 International Conference on Information Engineering and Computer Science. Presented at the 2009 International Conference on Information Engineering and Computer Science, pp. 1–4. <https://doi.org/10.1109/ICIECS.2009.5362936>
- 171) Waser, L.T., Fischer, C., Wang, Z., Ginzler, C., 2015. Wall-to-Wall Forest Mapping Based on Digital Surface Models from Image-Based Point Clouds and a NFI Forest Definition. *Forests* 6, 4510–4528. <https://doi.org/10.3390/f6124386>
- 172) Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., RStudio, 2021a. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.
- 173) Wickham, H., François, R., Henry, L., Müller, K., RStudio, 2021b. dplyr: A Grammar of Data Manipulation.
- 174) Wickham, H., RStudio, 2019. stringr: Simple, Consistent Wrappers for Common String Operations.
- 175) Wright, M.N., Wager, S., Probst, P., 2020. ranger: A Fast Implementation of Random Forests.
- 176) Wu, J., Jelinski, D.E., Luck, M., Tueller, P.T., 2000. Multiscale Analysis of Landscape Heterogeneity: Scale Variance and Pattern Metrics. *Geographic Information Sciences* 6, 6–19. <https://doi.org/10.1080/10824000009480529>
- 177) Wulder, M., 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography: Earth and Environment* 22, 449–476. <https://doi.org/10.1177/030913339802200402>
- 178) Wulder, M.A., Hermosilla, T., Stinson, G., Gougeon, F.A., White, J.C., Hill, D.A., Smiley, B.P., 2020. Satellite-based time series land cover and change information to map forest area consistent with national and international reporting requirements. *Forestry: An International Journal of Forest Research* 93, 331–343. <https://doi.org/10.1093/forestry/cpaa006>
- 179) Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., Sun, G., Yuan, W., Zhang, L., Zhang, X., 2019. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment* 233, 111383. <https://doi.org/10.1016/j.rse.2019.111383>
- 180) Xiao, J., Moody, A., 2005. Geographical distribution of global greening trends and their climatic correlates: 1982–1998. *International Journal of Remote Sensing* 26, 2371–2390. <https://doi.org/10.1080/01431160500033682>
- 181) Xiao-chen, L., Wen-jie, F., Qing-jiu, T., Xi-ru, X., 2008. The LAI Inversion based on Directional Second Derivative using Hyperspectral Data, in: IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, p. III-740-III-743. <https://doi.org/10.1109/IGARSS.2008.4779454>
- 182) Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., Casa, R., Pignatti, S., Laneve, G., Pascucci, S., Dong, Y., Ye, H., 2018. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, 1482–1493. <https://doi.org/10.1109/JSTARS.2018.2813281>

- 183) Xinfang Yui, Dafang Zhuang, Hua Chen, Xiyong Hou, 2004. Forest classification based on MODIS time series and vegetation phenology, in: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, pp. 2369–2372 vol.4. <https://doi.org/10.1109/IGARSS.2004.1369764>
- 184) Xue, J., Su, B., 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*. <https://doi.org/10.1155/2017/1353691>
- 185) Yan, L., Roy, D.P., 2018. Large-Area Gap Filling of Landsat Reflectance Time Series by Spectral-Angle-Mapper Based Spatio-Temporal Similarity (SAMSTS). *Remote Sensing* 10, 609. <https://doi.org/10.3390/rs10040609>
- 186) Yu, D., Shi, P., Shao, H., Zhu, W., Pan, Y., 2009. Modelling net primary productivity of terrestrial ecosystems in East Asia based on an improved CASA ecosystem model. *International Journal of Remote Sensing* 30, 4851–4866. <https://doi.org/10.1080/01431160802680552>
- 187) Yuan, J., Lv, X., Li, R., 2018. A Speckle Filtering Method Based on Hypothesis Testing for Time-Series SAR Images. *Remote Sensing* 10, 1383. <https://doi.org/10.3390/rs10091383>
- 188) Zeng, W., Tomppo, E., Healey, S.P., Gadov, K.V., 2015. The national forest inventory in China: history - results - international context. *Forest Ecosystems* 2, 23. <https://doi.org/10.1186/s40663-015-0047-2>
- 189) Zhang, C., Ju, W., Chen, J.M., Li, D., Wang, X., Fan, W., Li, M., Zan, M., 2014. Mapping forest stand age in China using remotely sensed forest height and observation data. *Journal of Geophysical Research: Biogeosciences* 119, 1163–1179. <https://doi.org/10.1002/2013JG002515>
- 190) Zhang, C., Marzougui, A., Sankaran, S., 2020. High-resolution satellite imagery applications in crop phenotyping: An overview. *Computers and Electronics in Agriculture* 175, 105584. <https://doi.org/10.1016/j.compag.2020.105584>
- 191) Zhang, Y., Zhang, C., Wang, Z., Chen, Y., Gang, C., An, R., Li, J., 2016. Vegetation dynamics and its driving forces from climate change and human activities in the Three-River Source Region, China from 1982 to 2012. *Science of The Total Environment* 563–564, 210–220. <https://doi.org/10.1016/j.scitotenv.2016.03.223>
- 192) Zheng, D., Wang, Y., Shao, Y., Wang, L., 2019. The Vegetation Dynamics and Climate Change Responses by Leaf Area Index in the Mu Us Desert. *Sustainability* 11, 3151. <https://doi.org/10.3390/su11113151>
- 193) Zheng, G., Moskal, L.M., 2009. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors* 9, 2719–2745. <https://doi.org/10.3390/s90402719>
- 194) Zhu, X., Cai, F., Tian, J., Williams, T.K.-A., 2018. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sensing* 10, 527. <https://doi.org/10.3390/rs10040527>
- 195) Zlinszky, A., Heilmeyer, H., Balzter, H., Czúcz, B., Pfeifer, N., 2015. Remote Sensing and GIS for Habitat Quality Monitoring: New Approaches and Future Research. *Remote Sensing* 7, 7987–7994. <https://doi.org/10.3390/rs70607987>
- 196) Банковић, С., Медаревић, М., Пантић, Д., Петровић, Н., 2009а. Национална инвентура шума Републике Србије - шумски фонд Републике Србије. Министарство пољопривреде шумарства и водопривреде Републике Србије - Управа за шуме.
- 197) Банковић, С., Медаревић, М., Пантић, Д., Петровић, Н., Шљукић, Б., Обрадовић, С., 2009б. Шумски фонд Републике Србије - стање и проблеми. *Glas Suma Fakul* 7–29. <https://doi.org/10.2298/GSF0900007B>

- 198) Лакушић, Д., 2019. База података о распрострањености потенцијално угрожених врста и станишта Србије - Резултати пројекта „Успостављање еколошке мреже Европске уније NATURA 2000 као дела еколошке мреже Републике Србије“. Центар за информације о биодиверзитету, Биолошки факултет, Универзитет у Београд.
- 199) Лакушић, Д., Кузмановић, Н., Ковачевић, Ј., 2021. Генерализована карта станишта Србије - Верзија 02.
- 200) Пантић, Д., Борота, Д., 2015. Национална инвентура шума Србије – стање и могући правци даљег развоја. Гласник Шумарског факултета 9-32. <https://doi.org/10.2298/GSF1512009P>
- 201) Републички геодетски завод, 2021. Карта основног земљишног покривача.
- 202) Службени гласник РС, 35/2010-29, 2010. Правилник о критеријумима за издвајање типова станишта, о типовима станишта, осетљивим, угроженим, ретким и за заштиту приоритетним типовима станишта и о мерама заштите за њихово очување, 35/2010-29.
- 203) Службени гласник РС, 102/2010-10, 2010. Уредба о еколошкој мрежи, 102/2010-10.

ПРИЛОЗИ

Коришћене скраћенице у табелама прилога

Ознака класе шумске вегетације:

- **1** – Salicion;
- **2** – Quercion roboris;
- **3** – Quercion frainetto;
- **4** – Ostryo-Carpinion orientalis;
- **5** – Fagion sylvaticae;
- **6** – Pinion nigrae;
- **7** – Vaccinio-Piceetea;
- **8** – Pinion mugo.

Примењена техника машинског учења:

- **DT** – стабло одлучивања
- **RF** – случајна шума
- **XGB** – градијентно појачавање

Временска одредница:

- **doy** – кумулативни дан у години сателитског опажања
- **day + month** – дан и месец сателитског опажања

Правило за агрегацију вишеструких вредности оцењених вероватноћа:

- **MC** – најчешће присутна класа
- **SM** – класа са највећом вероватноћом срачунатом простом аритметичком средином
- **GM** – класа са највећом вероватноћом срачунатом геометријском средином

Коришћен скуп атрибута

- **сви атрибути** – комплетан скуп атрибута
- **CFS** – скуп атрибута одређен избором атрибута заснованим на корелацији
- **SFS** – скуп атрибута одређен секвенцијалном претрагом унапред
- **SBS** – скуп атрибута одређен секвенцијалном претрагом уназад
- **HCS** – скуп атрибута одређен претрагом планинарењем
- **GA** – скуп атрибута одређен применом генетских алгоритама

Прилог 1 – Квантитативне мере перформанси различитих приступа респектовања временске компоненте

Табела П1. Произвођачева тачност различитих приступа респектовања временске
компоненте

Временска одредница		Правило агрегације	1	2	3	4	5	6	7	8
DT	имплицитно		29.33	88.90	43.76	39.83	44.66	71.89	83.82	0.00
	без временске одреднице	MC	3.12	98.35	0.69	0.00	41.62	42.58	96.53	0.00
		SM	1.95	99.32	0.27	0.00	34.35	41.91	97.77	0.00
		GM	7.41	99.33	5.40	0.00	33.36	42.71	96.48	0.00
	doy	MC	2.99	98.31	0.69	0.00	41.63	41.55	96.41	0.00
		SM	1.24	99.39	0.27	0.00	34.36	39.73	97.59	0.00
		GM	8.52	99.40	5.51	0.00	33.36	41.89	96.13	0.00
		day + month	MC	6.25	98.25	4.61	0.00	42.01	43.07	95.31
	SM		3.57	99.30	3.02	0.00	35.51	42.09	97.49	0.00
	GM		7.17	99.39	8.32	0.00	34.47	42.46	95.96	0.00
RF	имплицитно		66.51	87.36	74.82	60.53	65.74	93.14	95.40	62.20
	без временске одреднице	MC	42.54	94.16	56.77	60.97	65.44	85.06	97.12	18.01
		SM	43.20	94.57	60.37	63.95	67.57	85.74	97.77	15.88
		GM	44.27	94.30	67.33	61.19	64.94	85.26	97.41	17.74
	doy	MC	49.78	92.57	70.74	70.14	63.22	87.48	97.62	53.15
		SM	50.44	93.29	72.16	72.53	63.82	87.97	98.09	56.99
		GM	52.86	93.30	73.97	71.24	62.99	87.61	97.98	54.07
		day + month	MC	50.02	92.57	71.39	72.20	62.65	87.38	97.68
	SM		50.67	93.32	72.86	74.25	63.25	88.00	98.03	63.03
	GM		53.02	93.33	74.59	73.32	62.55	87.67	97.98	61.36
XGB	имплицитно		69.37	86.33	76.49	64.97	67.73	94.31	95.35	63.75
	без временске одреднице	MC	28.77	97.53	43.77	63.43	65.34	85.44	97.47	26.67
		SM	27.81	98.04	47.45	65.84	67.39	86.00	97.94	31.04
		GM	29.45	97.50	56.59	62.70	64.63	85.53	97.69	28.08
	doy	MC	38.67	94.59	63.14	74.62	61.16	87.27	97.66	68.28
		SM	38.53	95.07	65.26	75.97	60.98	87.43	97.99	77.13
		GM	40.00	94.84	68.35	74.91	59.66	87.03	97.96	69.49
		day + month	MC	38.07	94.64	63.00	74.73	60.95	86.82	97.72
	SM		38.11	95.24	65.11	76.06	60.86	87.45	97.98	77.47
	GM		39.46	94.95	68.18	75.17	59.59	87.14	98.04	69.49

Табела П2. Корисникова тачност различитих приступа респектовања временске
КОМПОНЕНТЕ

Временска одредница		Правило агрегације	1	2	3	4	5	6	7	8	
DT	имплицитно		56.14	74.10	47.04	65.82	53.67	71.98	85.71	0.00	
	без временске одреднице	MC	42.95	56.98	33.86	0.00	53.10	83.81	62.24	0.00	
		SM	69.75	55.45	34.86	0.00	55.45	82.79	63.79	0.00	
		GM	62.69	57.11	32.81	0.00	55.99	85.02	72.60	0.00	
	day	MC	38.87	56.81	33.93	0.00	53.10	76.95	65.46	0.00	
		SM	49.95	55.39	35.91	0.00	55.45	80.30	64.01	0.00	
		GM	65.63	57.23	32.57	0.00	55.98	82.28	73.78	0.00	
		day + month	MC	48.81	57.96	48.98	0.00	52.92	75.42	69.74	0.00
			SM	72.07	56.08	52.48	0.00	55.36	79.52	68.42	0.00
			GM	62.66	57.95	37.12	0.00	56.24	82.17	74.17	0.00
RF	имплицитно		73.60	88.53	62.65	76.99	72.68	91.34	95.05	96.50	
	без временске одреднице	MC	83.02	79.20	61.24	82.00	73.33	84.73	88.08	98.54	
		SM	86.70	80.23	62.91	85.23	75.76	84.81	88.51	100.00	
		GM	87.58	82.07	61.32	88.37	78.31	85.98	90.11	100.00	
	day	MC	83.87	84.21	62.33	84.32	74.80	86.98	91.95	96.55	
		SM	86.35	84.58	63.47	86.25	76.47	87.53	92.07	99.34	
		GM	87.15	85.71	62.44	87.32	77.61	88.00	92.63	99.23	
		day + month	MC	84.00	84.39	62.57	83.08	74.80	87.31	91.81	95.40
			SM	86.36	84.85	63.72	83.75	76.34	88.04	92.19	99.04
			GM	87.08	85.97	62.71	85.01	77.47	88.67	92.65	99.14
XGB	имплицитно		71.74	89.66	63.45	78.60	73.52	91.89	96.85	95.76	
	без временске одреднице	MC	84.58	74.17	64.34	78.51	71.12	82.74	88.65	95.64	
		SM	88.17	74.92	66.06	82.63	74.06	82.96	88.80	99.39	
		GM	88.93	77.02	63.41	86.37	76.84	84.61	90.52	99.46	
	day	MC	86.59	80.62	61.64	80.47	72.23	85.39	92.63	96.56	
		SM	88.65	80.84	62.40	80.99	74.52	86.13	92.90	98.54	
		GM	88.64	81.70	61.53	82.41	76.15	86.92	93.52	99.29	
		day + month	MC	86.71	80.36	61.51	80.01	72.22	86.04	92.92	95.72
			SM	88.67	80.68	62.55	80.57	74.55	86.60	93.21	98.50
			GM	88.36	81.49	61.68	82.27	76.25	87.39	93.79	99.24

Прилог 2 – Квантитативне мере перформанси различитих приступа избора репрезентативних атрибута

Табела ПЗ. Вредности укупне тачности и капа коефицијента слагања за различите методе селекције репрезентативних атрибута

Техника машинског учења	Временска одредница	Правило агрегације	сви атрибути		CFS		SFS		SBS		HCS		GA	
			ОА [%]	Капа	ОА [%]	Капа	ОА [%]	Капа	ОА [%]	Капа	ОА [%]	Капа	ОА [%]	Капа
DT	doy	имплицитно	75.98	0.65	75.48	0.64	76.08	0.65	75.98	0.65	75.12	0.63	75.62	0.64
		MC	72.32	0.60	72.29	0.60	72.29	0.60	72.32	0.60	69.73	0.57	72.27	0.60
		SM	71.59	0.59	71.55	0.59	71.59	0.59	71.59	0.59	69.69	0.57	71.53	0.59
	day + month	GM	71.47	0.59	71.43	0.59	71.50	0.59	71.47	0.59	69.70	0.57	71.41	0.59
		MC	72.32	0.60	72.29	0.60	72.37	0.61	72.32	0.60	72.52	0.61	72.16	0.60
		SM	71.59	0.59	71.55	0.59	71.64	0.59	71.59	0.59	71.78	0.60	71.45	0.59
GM	71.47	0.59	71.43	0.59	71.51	0.59	71.47	0.59	71.66	0.59	71.36	0.59		
RF	doy	имплицитно	83.33	0.76	81.20	0.73	76.51	0.65	83.33	0.76	82.81	0.75	83.29	0.76
		MC	80.76	0.74	74.50	0.64	75.02	0.65	80.87	0.74	80.39	0.73	80.39	0.73
		SM	81.12	0.74	74.65	0.65	75.06	0.65	81.24	0.74	80.67	0.73	80.70	0.73
	day + month	GM	81.20	0.74	74.70	0.65	75.06	0.65	81.30	0.74	80.71	0.73	80.76	0.74
		MC	80.84	0.74	74.50	0.64	74.74	0.64	80.55	0.73	80.92	0.74	80.56	0.73
		SM	81.24	0.74	74.65	0.65	74.90	0.65	80.91	0.74	81.26	0.74	80.87	0.74
GM	81.31	0.74	74.70	0.65	74.90	0.65	80.98	0.74	81.33	0.74	80.94	0.74		
XGB	doy	имплицитно	83.22	0.76	81.21	0.73	80.09	0.71	83.21	0.76	82.73	0.75	81.96	0.74
		MC	81.05	0.74	75.37	0.65	77.88	0.69	81.09	0.74	80.31	0.73	80.08	0.73
		SM	81.24	0.74	75.37	0.65	77.85	0.69	81.30	0.74	80.60	0.73	80.34	0.73
	day + month	GM	81.38	0.74	75.41	0.66	77.92	0.69	81.43	0.74	80.74	0.73	80.49	0.73
		MC	80.90	0.74	75.34	0.65	76.56	0.67	80.72	0.73	80.74	0.73	80.25	0.73
		SM	81.08	0.74	75.36	0.65	76.60	0.67	80.92	0.74	80.97	0.74	80.51	0.73
GM	81.20	0.74	75.39	0.65	76.56	0.67	81.03	0.74	81.08	0.74	80.63	0.73		

Табела П4. Вредности произвођачеве тачности за различите методе селекције репрезентативних атрибута коришћењем технике стабла одлучивања (DT)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
сви атриб.			44.09	87.31	75.80	33.40	62.68	77.30	89.06	0.00
CFS	имплицитно		30.60	89.66	77.54	1.15	64.18	81.44	89.59	0.00
SFS			45.69	87.95	76.31	5.33	63.24	75.55	89.03	0.00
SBS			44.09	87.31	75.80	33.40	62.68	77.30	89.06	0.00
HCS			28.87	89.43	76.07	35.78	62.72	76.90	89.25	0.00
GA			33.67	89.08	75.54	35.78	62.58	79.70	89.57	0.00
сви атриб.				18.05	92.50	73.68	0.00	53.05	72.22	93.91
CFS	MC		18.34	92.37	73.65	0.00	53.05	72.22	93.91	0.00
SFS			18.34	92.19	76.38	0.00	55.06	59.32	95.20	0.00
SBS			18.05	92.50	73.68	0.00	53.05	72.22	93.91	0.00
HCS			12.29	87.86	73.10	0.00	53.05	76.93	93.91	0.00
GA			18.34	92.30	73.68	0.00	53.05	72.22	93.91	0.00
сви атриб.				11.40	92.42	73.66	0.00	53.05	73.68	93.91
CFS	doy	SM	11.37	92.37	73.63	0.00	53.05	73.66	93.91	0.00
SFS			11.37	92.22	76.70	0.00	54.25	61.15	95.88	0.00
SBS			11.40	92.42	73.66	0.00	53.05	73.68	93.91	0.00
HCS			12.93	87.61	73.10	0.00	53.05	76.93	93.91	0.00
GA			11.37	92.30	73.66	0.00	53.05	73.68	93.91	0.00
сви атриб.					10.59	92.70	73.66	0.00	53.05	70.92
CFS	GM		10.57	92.65	73.63	0.00	53.05	70.90	93.91	0.00
SFS			10.57	92.50	77.47	0.00	53.54	58.29	95.91	0.00
SBS			10.59	92.70	73.66	0.00	53.05	70.92	93.91	0.00
HCS			12.93	87.61	73.17	0.00	53.05	76.92	93.91	0.00
GA			10.57	92.58	73.66	0.00	53.05	70.92	93.91	0.00
сви атриб.				18.05	92.50	73.68	0.00	53.05	72.22	93.91
CFS	MC		18.34	92.37	73.65	0.00	53.05	72.22	93.91	0.00
SFS			18.34	92.58	73.57	0.00	53.05	72.22	93.91	0.00
SBS			18.05	92.50	73.68	0.00	53.05	72.22	93.91	0.00
HCS			18.34	92.97	73.40	0.00	53.05	72.22	93.91	0.00
GA			18.34	91.81	76.56	0.00	55.06	59.32	95.20	0.00
сви атриб.				11.40	92.42	73.66	0.00	53.05	73.68	93.91
CFS	day + month	SM	11.37	92.37	73.63	0.00	53.05	73.66	93.91	0.00
SFS			11.37	92.58	73.55	0.00	53.05	73.70	93.91	0.00
SBS			11.40	92.42	73.66	0.00	53.05	73.68	93.91	0.00
HCS			11.37	92.96	73.39	0.00	53.05	73.71	93.91	0.00
GA			11.37	91.85	76.87	0.00	54.25	61.11	95.88	0.00
сви атриб.					10.59	92.70	73.66	0.00	53.05	70.92
CFS	GM		10.57	92.65	73.63	0.00	53.05	70.90	93.91	0.00
SFS			10.57	92.86	73.55	0.00	53.05	70.91	93.91	0.00
SBS			10.59	92.70	73.66	0.00	53.05	70.92	93.91	0.00
HCS			10.57	93.25	73.39	0.00	53.05	70.93	93.91	0.00
GA			10.57	92.12	77.64	0.00	53.54	58.20	95.91	0.00

Табела П5. Вредности корисникове тачности за различите методе селекције репрезентативних атрибута коришћењем технике стабла одлучивања (DT)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
сви атриб.			63.76	83.16	63.70	62.61	69.71	70.47	90.55	0.00
CFS			66.27	80.91	64.10	39.16	65.96	74.43	92.60	0.00
SFS			66.86	83.58	64.07	34.09	65.03	71.14	89.81	0.00
SBS		имплицитно	63.76	83.16	63.70	62.61	69.71	70.47	90.55	0.00
HCS			62.08	80.45	63.72	60.89	69.37	73.86	91.87	0.00
GA			65.59	81.35	63.68	60.96	69.55	71.35	91.88	0.00
сви атриб.			70.17	78.38	56.80	0.00	65.36	72.30	90.84	0.00
CFS			68.53	78.39	56.88	0.00	65.36	72.30	90.84	0.00
SFS			68.53	78.49	57.64	0.00	63.21	78.63	88.23	0.00
SBS		MC	70.17	78.38	56.80	0.00	65.36	72.30	90.84	0.00
HCS			41.90	77.47	58.15	0.00	65.36	54.18	90.84	0.00
GA			68.53	78.39	56.80	0.00	65.36	72.30	90.84	0.00
сви атриб.			60.22	77.34	57.00	0.00	65.36	71.09	90.84	0.00
CFS			58.68	77.31	57.07	0.00	65.36	71.10	90.84	0.00
SFS			58.68	77.41	57.47	0.00	63.80	79.09	87.39	0.00
SBS		SM	60.22	77.34	57.00	0.00	65.36	71.09	90.84	0.00
HCS	day		41.81	77.53	58.55	0.00	65.36	53.18	90.84	0.00
GA			58.68	77.31	57.00	0.00	65.36	71.09	90.84	0.00
сви атриб.			61.41	77.02	57.00	0.00	65.36	70.73	90.84	0.00
CFS			59.68	76.99	57.08	0.00	65.36	70.73	90.84	0.00
SFS			59.68	77.10	57.07	0.00	64.80	81.02	87.07	0.00
SBS		GM	61.41	77.02	57.00	0.00	65.36	70.73	90.84	0.00
HCS			41.81	77.53	58.17	0.00	65.36	54.26	90.84	0.00
GA			59.68	76.99	57.00	0.00	65.36	70.73	90.84	0.00
сви атриб.			70.17	78.38	56.80	0.00	65.36	72.30	90.84	0.00
CFS			68.53	78.39	56.88	0.00	65.36	72.30	90.84	0.00
SFS			68.53	78.40	57.11	0.00	65.36	72.30	90.84	0.00
SBS		MC	70.17	78.38	56.80	0.00	65.36	72.30	90.84	0.00
HCS			68.53	78.43	57.51	0.00	65.36	72.30	90.84	0.00
GA			68.53	78.47	57.26	0.00	63.21	78.63	88.23	0.00
сви атриб.			60.22	77.34	57.00	0.00	65.36	71.09	90.84	0.00
CFS			58.68	77.31	57.07	0.00	65.36	71.10	90.84	0.00
SFS			58.68	77.33	57.30	0.00	65.36	71.09	90.84	0.00
SBS		SM	60.22	77.34	57.00	0.00	65.36	71.09	90.84	0.00
HCS	day + month		58.68	77.36	57.70	0.00	65.36	71.11	90.84	0.00
GA			58.68	77.38	57.09	0.00	63.80	79.09	87.39	0.00
сви атриб.			61.41	77.02	57.00	0.00	65.36	70.73	90.84	0.00
CFS			59.68	76.99	57.08	0.00	65.36	70.73	90.84	0.00
SFS			59.68	77.01	57.30	0.00	65.36	70.75	90.84	0.00
SBS		GM	61.41	77.02	57.00	0.00	65.36	70.73	90.84	0.00
HCS			59.68	77.04	57.70	0.00	65.36	70.78	90.84	0.00
GA			59.68	77.07	56.72	0.00	64.80	80.90	87.07	0.00

Табела П6. Вредности произвођачеве тачности за различите методе селекције репрезентативних атрибута коришћењем технике случајне шуме (RF)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
сви атриб.			70.34	88.26	84.01	58.16	67.04	94.64	97.81	77.29
CFS			65.92	86.78	81.36	54.12	65.23	93.05	96.34	87.87
SFS		имплицитно	39.16	89.59	79.45	6.33	54.26	87.21	94.44	68.55
SBS			70.39	88.24	84.09	58.06	67.03	94.44	97.79	79.55
HCS			70.40	87.79	83.23	54.29	66.67	94.07	96.96	86.46
GA			71.02	88.18	83.46	55.16	67.81	94.53	97.26	81.66
сви атриб.			58.75	88.52	82.54	63.85	63.25	88.40	96.81	94.03
CFS			28.96	89.84	73.96	44.06	57.08	79.64	93.11	95.74
SFS		MC	23.51	91.35	77.11	15.69	58.61	83.57	95.75	84.19
SBS			59.58	88.50	82.57	64.50	63.19	88.56	96.82	94.79
HCS			57.14	88.87	80.37	57.16	65.51	86.67	96.50	95.36
GA			56.61	88.57	81.64	63.14	64.08	87.05	96.64	94.76
сви атриб.			59.64	89.01	82.86	65.13	62.79	88.52	96.80	94.34
CFS			28.84	89.92	74.41	44.61	57.15	79.98	93.30	95.86
SFS		SM	23.46	91.55	77.13	14.14	58.41	83.80	95.77	83.62
SBS	day		60.41	89.00	82.97	65.51	62.78	88.61	96.82	94.87
HCS			57.20	89.33	80.75	59.11	65.09	86.80	96.51	95.21
GA			57.08	88.95	82.13	64.34	63.86	87.10	96.63	94.79
сви атриб.			59.98	89.08	83.00	64.95	62.70	88.44	96.84	94.30
CFS			29.14	89.93	74.42	44.47	57.26	79.99	93.26	95.86
SFS		GM	23.57	91.61	77.12	13.38	58.46	83.35	95.73	84.00
SBS			60.66	89.08	83.11	65.31	62.66	88.53	96.85	94.95
HCS			57.38	89.40	80.89	58.78	64.92	86.74	96.52	95.21
GA			57.39	89.00	82.27	64.06	63.76	87.05	96.65	94.79
сви атриб.			59.07	88.38	83.24	64.03	63.10	88.30	96.87	94.15
CFS			28.96	89.84	73.96	44.06	57.08	79.64	93.11	95.74
SFS		MC	22.21	90.84	77.28	25.71	58.64	80.88	95.71	85.33
SBS			56.23	88.59	82.84	62.56	63.49	88.20	96.91	93.16
HCS			58.00	88.93	81.51	60.20	65.95	87.84	96.90	95.21
GA			56.66	88.55	82.60	63.47	64.04	87.10	96.77	94.68
сви атриб.			59.57	89.01	83.53	65.82	62.75	88.39	96.89	94.26
CFS			28.84	89.92	74.41	44.61	57.15	79.98	93.30	95.86
SFS		SM	23.25	91.02	77.39	23.07	58.76	80.64	95.74	83.89
SBS	day + month		56.75	89.17	83.19	64.58	62.93	88.34	96.91	93.47
HCS			58.35	89.43	82.03	61.52	65.59	87.88	96.83	95.10
GA			57.08	88.96	83.00	65.20	63.75	87.20	96.74	94.79
сви атриб.			59.86	89.09	83.65	65.64	62.66	88.34	96.94	94.26
CFS			29.14	89.93	74.42	44.47	57.26	79.99	93.26	95.86
SFS		GM	23.61	91.04	77.47	21.84	58.74	80.07	95.66	84.50
SBS			57.12	89.23	83.31	64.38	62.84	88.26	96.95	93.50
HCS			58.65	89.49	82.21	61.13	65.45	87.86	96.87	95.17
GA			57.33	89.01	83.13	65.07	63.69	87.24	96.78	94.87

Табела П7. Вредности корисникове тачности за различите методе селекције репрезентативних атрибута коришћењем технике случајне шуме (RF)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8	
сви атриб.			76.76	91.25	66.63	79.56	75.59	94.23	96.34	100.00	
CFS			71.79	89.84	65.07	70.07	72.28	91.83	96.20	100.00	
SFS	имплицитно		69.13	83.00	60.72	51.67	66.87	88.09	91.36	98.18	
SBS			76.59	91.27	66.68	79.15	75.64	94.27	96.39	100.00	
HCS			74.97	91.17	66.38	75.62	74.40	94.03	96.11	100.00	
GA			76.32	91.35	66.91	78.19	74.98	93.86	96.21	99.83	
сви атриб.				80.44	88.34	61.50	85.49	77.64	89.03	95.91	100.00
CFS			62.12	81.80	59.89	57.58	66.04	79.65	93.67	99.64	
SFS	MC		69.90	80.10	60.67	65.85	69.19	80.93	93.27	99.77	
SBS			80.36	88.56	61.57	85.36	77.58	89.22	95.96	100.00	
HCS			80.74	87.84	61.43	84.09	76.39	87.44	95.19	99.88	
GA			80.33	87.82	61.39	85.21	76.87	88.16	95.37	99.96	
сви атриб.				82.49	88.57	61.58	86.90	78.03	88.97	95.92	100.00
CFS	day		62.71	81.77	60.11	58.51	66.38	79.96	93.73	99.64	
SFS		SM		70.71	80.12	60.79	64.34	69.03	80.79	93.14	99.77
SBS				82.41	88.77	61.67	87.04	78.06	89.14	95.97	100.00
HCS				82.75	87.88	61.50	86.18	76.76	87.41	95.36	99.88
GA				81.93	87.94	61.56	86.66	77.42	88.11	95.52	99.92
сви атриб.				82.86	88.65	61.57	87.46	78.10	89.01	95.91	100.00
CFS	GM		63.00	81.82	60.05	58.88	66.25	80.76	93.71	99.64	
SFS			70.83	80.10	60.81	64.33	68.89	81.05	93.18	99.77	
SBS			82.82	88.84	61.65	87.63	78.14	89.15	95.95	100.00	
HCS			83.13	87.93	61.42	86.96	76.81	87.56	95.35	99.88	
GA			82.28	88.01	61.52	87.28	77.48	88.24	95.52	99.92	
сви атриб.			80.76	88.45	61.46	85.71	78.13	89.13	95.96	100.00	
CFS	MC		62.12	81.80	59.89	57.58	66.04	79.65	93.67	99.64	
SFS			67.09	79.70	60.19	79.11	71.28	78.66	93.11	99.78	
SBS			80.84	87.79	61.40	85.72	77.91	89.09	95.73	100.00	
HCS			81.15	88.10	62.05	86.29	77.76	88.16	95.84	99.96	
GA			80.61	87.87	61.63	86.47	77.75	87.91	95.49	100.00	
сви атриб.			82.57	88.61	61.78	87.12	78.57	89.03	95.95	100.00	
CFS	day + month		62.71	81.77	60.11	58.51	66.38	79.96	93.73	99.64	
SFS		SM		69.35	79.83	60.26	80.62	71.08	79.18	92.88	99.77
SBS				83.04	87.95	61.58	87.26	78.30	89.05	95.66	100.00
HCS				83.28	88.20	62.17	87.78	78.47	88.14	95.80	99.96
GA				82.20	88.00	61.82	87.26	78.27	87.90	95.56	100.00
сви атриб.				82.96	88.69	61.76	87.55	78.65	89.06	95.94	100.00
CFS	GM		63.00	81.82	60.05	58.88	66.25	80.76	93.71	99.64	
SFS			69.67	79.83	60.20	81.56	70.97	79.42	92.95	99.78	
SBS			83.40	88.04	61.56	87.51	78.38	89.10	95.64	100.00	
HCS			83.69	88.27	62.12	88.26	78.64	88.18	95.81	99.96	
GA			82.50	88.06	61.80	87.72	78.36	88.03	95.60	100.00	

Табела П8. Вредности произвођачеве тачности за различите методе селекције репрезентативних атрибута коришћењем технике градијентног појачавања (XGB)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
сви атриб.			74.08	87.36	83.14	57.55	68.67	93.46	95.14	92.24
CFS			67.71	86.82	80.55	56.17	65.20	92.46	93.84	93.09
SFS			63.64	87.63	80.60	50.92	59.30	89.05	90.28	86.18
SBS			75.13	87.20	82.84	57.60	68.86	93.20	94.79	92.67
HCS			72.66	87.26	82.60	58.20	67.50	92.47	94.87	92.95
GA			69.03	86.78	82.83	56.52	66.99	92.61	94.46	92.95
сви атриб.			57.09	89.64	81.52	67.77	65.06	87.24	95.09	94.72
CFS			28.71	90.24	79.10	42.93	55.47	80.73	93.02	94.72
SFS			36.08	91.01	80.49	52.32	59.09	85.67	95.44	94.72
SBS			57.16	89.59	81.46	67.56	65.52	87.21	95.34	94.72
HCS			53.74	89.39	80.41	62.16	65.77	86.19	95.88	94.72
GA			52.96	89.57	80.33	66.38	63.66	86.25	95.26	94.72
сви атриб.			57.59	89.92	82.03	68.15	64.43	87.19	95.06	94.72
CFS			28.59	90.23	78.98	44.03	55.44	80.90	93.27	94.72
SFS			36.01	90.86	80.98	53.42	58.54	85.67	95.51	94.72
SBS			57.69	89.86	82.04	68.17	64.86	87.23	95.30	94.72
HCS			54.22	89.79	81.07	63.06	65.12	86.21	95.90	94.72
GA			53.73	89.82	80.86	66.91	63.37	86.12	95.22	94.72
сви атриб.			58.21	89.99	82.30	67.95	64.37	87.08	95.04	94.72
CFS			28.86	90.25	79.00	43.75	55.50	80.72	93.29	94.72
SFS			36.92	90.80	81.15	52.82	58.42	85.51	95.52	94.72
SBS			58.39	89.92	82.30	68.01	64.78	87.12	95.27	94.72
HCS			55.08	89.84	81.38	62.56	64.95	86.11	95.90	94.72
GA			54.64	89.87	81.09	66.77	63.34	86.07	95.19	94.72
сви атриб.			56.92	89.55	81.44	66.91	64.63	87.15	95.12	94.72
CFS			28.59	90.22	79.05	42.95	55.45	80.77	92.98	94.72
SFS			30.75	90.30	80.50	38.20	58.75	84.34	96.11	94.72
SBS			56.46	89.54	81.22	66.91	64.03	87.08	95.13	94.72
HCS			55.67	89.60	81.16	66.15	65.04	86.82	95.10	94.72
GA			54.18	89.54	80.65	63.23	63.64	86.58	96.00	94.72
сви атриб.			57.28	89.81	81.81	67.65	64.16	87.22	95.09	94.72
CFS			28.50	90.22	78.93	44.04	55.39	81.04	93.28	94.72
SFS			31.51	90.01	80.61	39.12	58.86	84.65	96.21	94.72
SBS			56.92	89.81	81.70	67.68	63.43	87.12	95.16	94.72
HCS			56.14	89.91	81.61	67.15	64.42	86.96	95.17	94.72
GA			54.82	89.87	81.10	63.78	63.18	86.57	96.05	94.72
сви атриб.			57.83	89.88	82.00	67.72	64.14	87.07	95.08	94.72
CFS			28.75	90.24	78.95	43.68	55.45	80.84	93.29	94.72
SFS			31.98	89.85	80.80	37.17	58.88	84.45	96.18	94.72
SBS			57.41	89.87	81.86	67.77	63.42	87.01	95.12	94.72
HCS			56.73	89.97	81.79	67.08	64.38	86.84	95.17	94.72
GA			55.50	89.92	81.28	63.44	63.16	86.47	96.05	94.72

Табела П9. Вредности корисникове тачности за различите методе селекције репрезентативних атрибута коришћењем технике градијентног појачавања (XGB)

Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
сви атриб.			74.30	92.20	66.97	79.31	73.97	93.66	96.74	99.85
CFS			71.79	90.20	64.97	70.62	71.02	92.23	96.51	96.21
SFS			72.20	89.08	63.78	64.07	69.80	87.90	93.78	55.19
SBS		имплицитно	74.62	92.35	66.78	78.00	74.00	92.97	96.64	99.24
HCS			73.31	91.75	66.76	76.85	73.26	92.84	96.66	98.51
GA			71.75	90.68	66.17	74.72	73.89	92.92	96.33	96.77
сви атриб.			81.88	87.99	62.82	83.05	77.02	88.21	95.95	99.60
CFS			65.30	81.86	60.53	64.66	70.25	80.03	93.68	99.56
SFS			75.81	83.56	61.71	75.31	73.23	86.13	96.11	99.60
SBS		MC	81.77	88.01	62.88	83.31	77.09	88.34	96.04	99.44
HCS			81.69	86.98	61.92	80.00	77.61	86.74	95.59	99.40
GA			80.72	87.07	61.75	83.44	75.42	87.03	95.82	99.60
сви атриб.			83.22	88.13	62.76	83.64	77.40	88.23	96.03	99.60
CFS			65.25	81.84	60.68	64.70	69.99	80.11	93.53	99.56
SFS			76.55	83.53	61.42	76.86	73.48	86.11	96.05	99.60
SBS		doj	83.04	88.16	62.85	84.01	77.57	88.33	96.10	99.52
HCS		SM	83.23	87.13	62.06	81.96	78.14	86.64	95.60	99.40
GA			82.10	87.25	61.81	84.27	75.96	86.98	95.88	99.56
сви атриб.			83.65	88.28	62.77	84.21	77.60	88.47	96.02	99.60
CFS			65.51	81.89	60.59	65.15	69.82	80.93	93.53	99.56
SFS			76.91	83.69	61.28	77.67	73.40	86.67	96.02	99.60
SBS		GM	83.47	88.32	62.85	84.67	77.75	88.56	96.09	99.44
HCS			83.69	87.32	62.00	83.10	78.33	86.86	95.59	99.44
GA			82.59	87.46	61.80	84.85	76.17	87.22	95.87	99.56
сви атриб.			81.98	87.95	62.49	82.54	76.62	88.27	96.03	99.36
CFS			64.99	81.84	60.52	64.77	70.18	80.04	93.70	99.56
SFS			71.31	82.13	60.88	73.03	73.04	83.93	95.34	99.01
SBS		MC	81.68	87.83	62.22	82.69	76.26	88.18	96.10	99.56
HCS			82.17	87.66	62.45	82.03	76.69	87.59	95.70	99.48
GA			81.52	87.19	61.47	83.15	76.77	87.07	95.97	99.44
сви атриб.			83.15	88.05	62.48	83.60	76.92	88.16	96.12	99.36
CFS			65.11	81.82	60.67	64.84	69.92	80.09	93.58	99.56
SFS			72.09	82.24	60.74	75.03	73.02	83.59	95.21	99.13
SBS		day + month	82.99	87.97	62.19	83.50	76.65	88.10	96.16	99.56
HCS		SM	83.66	87.80	62.44	82.95	77.01	87.65	95.78	99.52
GA			82.98	87.36	61.56	84.43	77.07	87.05	95.95	99.40
сви атриб.			83.55	88.19	62.49	84.31	77.03	88.38	96.11	99.36
CFS			65.35	81.86	60.57	65.12	69.76	80.93	93.60	99.56
SFS			72.39	82.29	60.39	76.05	72.78	84.66	95.31	99.20
SBS		GM	83.39	88.08	62.19	84.18	76.75	88.29	96.18	99.56
HCS			84.05	87.94	62.43	83.65	77.11	87.93	95.79	99.52
GA			83.40	87.52	61.53	85.42	77.18	87.24	95.97	99.40

Прилог 3 – Квантитативне мере перформанси постигнуте коришћењем скупа података балансираног применом SMOTE алгоритма

Табела П10. Вредности укупне тачности и капа коефицијента слагања постигнуте коришћењем балансираног скупа података

Техника машинског учења	Скуп атрибута	Временска одредница	Правило агрегације	ОА [%]	Капа
DT	SFS		имплицитно	74.51	0.63
			MC	69.23	0.57
	сви атрибути	doy	SM	69.16	0.57
			GM	69.28	0.58
			MC	69.24	0.57
			SM	69.04	0.57
HCS	day + month	GM	69.26	0.57	
RF	сви атрибути		имплицитно	83.36	0.76
			MC	81.25	0.74
	SBS	doy	SM	81.67	0.75
			GM	81.76	0.75
			MC	81.35	0.74
			SM	81.75	0.75
HCS	day + month	GM	81.82	0.75	
XGB	сви атрибути		имплицитно	83.20	0.76
			MC	81.17	0.74
	SBS	doy	SM	81.42	0.74
			GM	81.57	0.75
			MC	80.91	0.74
			SM	81.13	0.74
сви атриб.	day + month	GM	81.27	0.74	

Табела П11. Вредности произвођачеве тачности постигнуте коришћењем балансираног скупа података

Техника машинског учења	Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
DT	SFS		имплицитно	43.16	87.07	64.32	66.28	61.04	84.76	85.48	93.09
			MC	22.04	85.86	61.32	62.01	50.20	80.81	91.94	95.90
	сви атрибути	doy	SM	21.95	85.54	61.77	63.26	49.38	82.32	92.01	95.90
			GM	21.95	85.10	63.78	62.88	49.47	81.67	92.00	95.90
	HCS	day + month	MC	22.04	86.06	60.82	53.77	51.62	80.97	91.96	95.90
			SM	21.95	85.69	61.23	60.17	49.04	82.55	92.02	95.90
			GM	21.95	85.42	63.88	56.48	49.27	81.32	91.97	95.90
RF	сви атрибути		имплицитно	70.26	88.16	82.20	61.80	69.93	94.97	98.00	91.11
			MC	62.89	87.79	81.62	72.33	64.89	90.00	96.87	97.11
	SBS	doy	SM	63.59	88.38	82.21	72.68	64.64	90.08	96.86	97.00
			GM	63.85	88.47	82.42	72.48	64.45	90.03	96.91	97.00
	HCS	day + month	MC	61.40	88.17	80.61	68.62	67.80	89.65	97.10	97.15
			SM	61.75	88.78	81.22	69.66	67.56	89.64	96.93	97.07
			GM	62.07	88.83	81.51	69.26	67.32	89.61	96.99	97.07
XGB	сви атрибути		имплицитно	74.26	87.33	81.47	62.48	70.11	93.94	95.12	93.09
			MC	58.53	89.34	79.16	75.04	66.78	89.31	95.34	95.36
	SBS	doy	SM	59.07	89.65	79.72	75.73	66.20	89.39	95.38	95.36
			GM	59.62	89.72	80.09	75.75	66.13	89.31	95.35	95.36
	сви атрибути	day + month	MC	57.39	89.35	79.99	73.02	65.15	88.86	95.18	95.33
			SM	57.87	89.68	80.33	73.90	64.67	88.98	95.22	95.33
			GM	58.42	89.75	80.58	73.85	64.67	88.88	95.21	95.33

Табела П12. Вредности корисникове тачности постигнуте коришћењем балансираног скупа података

Техника машинског учења	Скуп атрибута	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
DT	SFS		имплицитно	64.39	83.64	64.56	26.68	69.81	63.84	94.73	55.84
			MC	46.28	78.69	59.81	23.67	70.30	60.36	94.27	87.00
	сви атрибути	doy	SM	46.17	78.70	60.22	23.65	70.92	59.17	94.11	87.00
			GM	46.17	78.66	59.17	24.15	71.08	62.93	94.16	87.00
	HCS	day + month	SM	46.17	78.72	59.06	24.42	70.65	57.84	94.06	87.00
			GM	46.17	78.70	58.21	24.28	70.18	62.61	94.18	87.00
RF	сви атрибути		имплицитно	76.74	91.32	67.35	74.53	74.59	93.33	96.40	99.85
			MC	79.30	89.29	62.42	76.32	78.59	88.14	96.19	99.88
	SBS	doy	SM	81.37	89.49	62.59	78.17	79.28	88.10	96.20	100.00
			GM	81.82	89.56	62.55	79.07	79.38	88.16	96.21	100.00
	HCS	day + month	SM	79.86	88.85	63.15	74.69	78.92	87.45	95.99	99.61
			GM	82.11	88.95	63.33	76.76	79.64	87.33	96.00	99.77
XGB	сви атрибути		имплицитно	74.04	92.22	67.58	74.34	73.37	93.12	96.65	86.39
			MC	80.79	88.26	64.07	68.70	77.48	86.39	96.09	99.48
	SBS	doy	SM	82.33	88.42	63.97	69.82	78.00	86.49	96.17	99.48
			GM	82.73	88.55	63.97	71.28	78.21	86.67	96.18	99.48
	сви атрибути	day + month	SM	80.99	88.00	63.36	70.42	77.42	86.57	96.11	99.52
			GM	82.50	88.15	63.29	71.68	77.67	86.56	96.17	99.52
			GM	82.89	88.28	63.29	72.81	77.83	86.82	96.16	99.52

Прилог 4 – Квантитативне мере перформанси постигнуте наглашавањем просторне компоненте класификационог модела

Табела П13. Вредности укупне тачности и капа коефицијента слагања постигнуте наглашавањем просторне компоненте класификационог модела

Величина суседства	Временска одредница	Правило агрегације	DT		RF		XGB	
			ОА [%]	Капа	ОА [%]	Капа	ОА [%]	Капа
1x1	имплицитно		74.51	0.63	83.36	0.76	83.20	0.76
		MC	69.23	0.57	81.25	0.74	81.17	0.74
	doy	SM	69.16	0.57	81.76	0.75	81.57	0.75
		GM	69.28	0.58	81.67	0.75	81.42	0.74
		MC	69.24	0.57	81.35	0.74	80.91	0.74
	day + month	SM	69.04	0.57	81.82	0.75	81.27	0.74
		GM	69.26	0.57	81.75	0.75	81.13	0.74
3x3	flat	MC	76.23	0.66	85.04	0.79	84.83	0.78
		SM	76.24	0.65	85.15	0.79	84.96	0.79
		GM	76.66	0.66	85.11	0.79	84.91	0.79
	doy	MC	70.20	0.59	82.96	0.77	82.30	0.76
		SM	70.13	0.59	83.22	0.77	82.54	0.76
		GM	70.52	0.59	83.20	0.77	82.43	0.76
	day + month	MC	70.21	0.59	83.33	0.77	81.97	0.75
		SM	70.05	0.58	83.41	0.77	82.17	0.75
		GM	70.43	0.59	83.43	0.77	82.06	0.75
5x5	flat	MC	76.95	0.67	85.33	0.79	85.17	0.79
		SM	76.91	0.66	85.42	0.79	85.29	0.79
		GM	77.36	0.67	85.43	0.79	85.22	0.79
	doy	MC	70.65	0.59	83.13	0.77	82.53	0.76
		SM	70.65	0.59	83.20	0.77	82.68	0.76
		GM	71.08	0.60	83.32	0.77	82.58	0.76
	day + month	MC	70.55	0.59	83.57	0.77	82.17	0.75
		SM	70.47	0.59	83.44	0.77	82.29	0.76
		GM	70.87	0.60	83.62	0.78	82.20	0.75

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П14. Вредности произвођачеве тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике стабла одлучивања (DT)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
1x1	имплицитно		43.16	87.07	64.32	66.28	61.04	84.76	85.48	93.09
		MC	22.04	85.86	61.32	62.01	50.20	80.81	91.94	95.90
		SM	21.95	85.54	61.77	63.26	49.38	82.32	92.01	95.90
	day + month	GM	21.95	85.10	63.78	62.88	49.47	81.67	92.00	95.90
		MC	22.04	86.06	60.82	53.77	51.62	80.97	91.96	95.90
		SM	21.95	85.69	61.23	60.17	49.04	82.55	92.02	95.90
		GM	21.95	85.42	63.88	56.48	49.27	81.32	91.97	95.90
		MC	44.66	88.54	66.65	66.22	63.06	86.58	88.26	93.23
3x3	имплицитно	SM	40.21	89.62	66.71	67.08	62.71	86.84	89.31	93.23
		GM	37.61	90.03	70.09	67.54	62.30	86.56	90.65	87.45
		MC	21.86	86.05	63.91	67.34	52.11	82.29	92.30	95.93
	day + month	SM	20.29	86.23	64.37	68.86	50.92	83.17	92.75	95.93
		GM	20.16	85.79	67.42	67.04	52.04	82.44	93.05	95.17
		MC	21.86	86.20	63.26	59.46	53.88	82.34	92.34	95.93
		SM	20.29	86.41	63.74	64.22	51.30	83.15	92.76	95.93
		GM	20.16	86.12	67.02	60.43	51.84	82.21	93.09	95.17
5x5	имплицитно	MC	45.49	89.14	67.74	66.76	63.84	87.55	88.57	93.23
		SM	39.54	90.45	67.87	67.79	63.55	87.77	89.75	93.23
		GM	36.75	90.60	72.24	68.13	63.31	87.34	91.30	83.50
	day + month	MC	21.81	86.16	65.27	71.39	52.68	82.26	92.53	95.90
		SM	19.67	86.54	65.80	73.16	51.69	83.13	93.08	96.01
		GM	19.50	85.99	69.25	70.66	53.19	82.46	93.15	94.57
		MC	21.81	86.28	64.29	61.91	54.39	82.36	92.56	95.90
		SM	19.67	86.72	64.97	67.10	51.80	83.16	93.08	96.01
GM	19.50	86.35	68.56	63.28	52.49	82.23	93.23	94.57		

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П15. Вредности корисникове тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике стабла одлучивања (DT)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
1x1	имплицитно	MC	64.39	83.64	64.56	26.68	69.81	63.84	94.73	55.84
		SM	46.28	78.69	59.81	23.67	70.30	60.36	94.27	87.00
	day	SM	46.17	78.70	60.22	23.65	70.92	59.17	94.11	87.00
		GM	46.17	78.66	59.17	24.15	71.08	62.93	94.16	87.00
	day + month	MC	46.28	78.72	58.76	23.99	68.61	60.05	94.24	87.00
		SM	46.17	78.72	59.06	24.42	70.65	57.84	94.06	87.00
		GM	46.17	78.70	58.21	24.28	70.18	62.61	94.18	87.00
3x3	имплицитно	MC	71.60	84.37	64.64	30.45	69.64	68.43	95.04	57.33
		SM	74.81	83.64	64.72	30.66	70.25	67.76	94.78	57.03
		GM	76.55	83.20	64.66	33.00	70.54	73.85	94.50	83.78
	day	MC	46.68	78.92	60.20	28.02	70.41	64.98	94.06	86.83
		SM	46.79	78.78	60.51	27.19	71.74	63.57	93.62	86.21
		GM	46.78	78.82	59.60	29.18	71.42	68.56	93.64	98.31
	day + month	MC	46.68	78.94	59.33	28.53	68.95	65.04	94.06	86.83
		SM	46.77	78.81	59.59	27.66	71.13	62.64	93.57	86.21
		GM	46.78	78.86	58.87	28.73	70.92	67.45	93.61	98.31
5x5	имплицитно	MC	74.46	84.59	64.65	32.69	70.24	70.65	95.44	59.12
		SM	79.07	83.61	64.84	32.77	71.05	69.63	95.22	58.65
		GM	79.69	83.10	64.87	36.61	71.24	77.95	94.75	94.12
	day	MC	46.81	78.94	60.31	31.33	70.72	66.76	94.06	88.34
		SM	46.85	78.74	60.66	30.27	72.24	65.81	93.60	87.99
		GM	46.66	78.80	59.74	33.45	71.93	70.66	93.72	99.20
	day + month	MC	46.80	78.95	59.55	30.55	69.25	66.83	94.10	88.34
		SM	46.83	78.77	59.82	29.43	71.77	64.92	93.63	87.99
		GM	46.67	78.85	59.15	31.11	71.66	69.65	93.69	99.20

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П16. Вредности произвођачеве тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике случајне шуме (RF)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
1x1	имплицитно	MC	70.26	88.16	82.20	61.80	69.93	94.97	98.00	91.11
		SM	62.89	87.79	81.62	72.33	64.89	90.00	96.87	97.11
	doy	SM	63.59	88.38	82.21	72.68	64.64	90.08	96.86	97.00
		GM	63.85	88.47	82.42	72.48	64.45	90.03	96.91	97.00
	dm	MC	61.40	88.17	80.61	68.62	67.80	89.65	97.10	97.15
		SM	61.75	88.78	81.22	69.66	67.56	89.64	96.93	97.07
		GM	62.07	88.83	81.51	69.26	67.32	89.61	96.99	97.07
3x3	имплицитно	MC	73.94	89.46	83.76	64.58	71.19	96.53	99.09	90.97
		SM	74.01	89.54	83.78	64.91	71.23	96.85	99.13	91.82
		GM	74.01	89.54	84.13	64.60	71.13	96.78	98.96	91.82
	doy	MC	66.47	89.17	83.85	76.14	65.59	91.26	97.80	97.34
		SM	66.67	89.53	84.13	76.41	65.47	91.20	98.11	97.38
		GM	66.49	89.64	84.61	75.58	65.17	90.90	97.74	97.38
	dm	MC	65.01	89.67	83.33	72.71	69.22	91.24	98.13	97.23
		SM	65.02	89.90	83.51	73.11	68.90	91.11	98.25	97.34
		GM	65.04	89.94	84.05	71.34	68.47	90.69	97.88	97.42
5x5	имплицитно	MC	74.43	89.58	84.57	65.12	71.20	96.89	99.39	91.26
		SM	74.30	89.72	84.59	66.23	71.38	97.12	99.37	93.23
		GM	74.34	89.72	85.22	65.60	71.03	96.65	98.15	93.23
	doy	MC	66.62	89.28	84.74	76.56	65.04	91.34	97.84	97.42
		SM	66.73	89.61	84.93	77.00	64.82	91.24	98.22	97.49
		GM	66.30	89.72	85.56	76.12	64.41	89.34	97.23	97.57
	day + month	MC	65.37	89.85	84.07	73.10	68.95	91.40	98.21	97.23
		SM	65.17	90.01	84.29	73.51	68.61	91.29	98.38	97.45
		GM	65.27	90.04	85.08	70.62	67.94	88.87	97.25	97.57

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П17. Вредности корисникове тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике случајне шуме (RF)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
1x1	имплицитно	MC	76.74	91.32	67.35	74.53	74.59	93.33	96.40	99.85
		SM	79.30	89.29	62.42	76.32	78.59	88.14	96.19	99.88
	doy	SM	81.37	89.49	62.59	78.17	79.28	88.10	96.20	100.00
		GM	81.82	89.56	62.55	79.07	79.38	88.16	96.21	100.00
	day + month	MC	79.86	88.85	63.15	74.69	78.92	87.45	95.99	99.61
		SM	82.11	88.95	63.33	76.76	79.64	87.33	96.00	99.77
		GM	82.48	89.04	63.25	77.88	79.78	87.36	96.00	99.80
		MC	83.70	92.72	67.41	78.50	75.11	95.33	97.05	100.00
3x3	имплицитно	SM	83.77	92.74	67.44	78.77	75.42	95.51	97.13	100.00
		GM	83.71	92.73	67.43	79.72	75.73	95.63	97.43	100.00
		MC	87.14	90.65	63.04	80.43	80.11	89.47	96.47	100.00
	doy	SM	88.38	90.71	63.20	80.71	80.54	89.59	96.56	100.00
		GM	88.16	90.67	63.16	81.74	80.87	89.91	97.00	100.00
		MC	88.23	90.31	64.27	78.73	80.94	89.05	96.32	100.00
	day + month	SM	89.07	90.31	64.25	79.43	81.14	89.04	96.41	100.00
		GM	88.25	90.32	64.17	80.62	81.45	89.20	96.90	100.00
5x5	имплицитно	MC	84.31	92.89	67.49	81.25	75.79	96.24	97.31	100.00
		SM	84.54	92.86	67.62	81.72	76.14	96.35	97.30	100.00
		GM	83.95	92.86	67.56	82.94	76.71	96.53	97.71	100.00
	doy	MC	87.70	90.73	63.03	81.69	80.87	89.88	96.63	100.00
		SM	88.74	90.74	63.19	81.70	81.20	89.94	96.77	100.00
		GM	86.95	90.63	63.15	82.70	81.81	90.32	97.44	100.00
	day + month	MC	89.03	90.42	64.34	78.99	81.79	89.58	96.44	100.00
		SM	89.46	90.36	64.40	79.67	81.87	89.72	96.58	100.00
		GM	86.17	90.40	64.39	81.68	82.53	89.42	97.36	100.00

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П18. Вредности произвођачеве тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике град. појачавања (XGB)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8	
1x1	имплицитно		74.26	87.33	81.47	62.48	70.11	93.94	95.12	93.09	
		MC	58.53	89.34	79.16	75.04	66.78	89.31	95.34	95.36	
	doy	SM	59.07	89.65	79.72	75.73	66.20	89.39	95.38	95.36	
		GM	59.62	89.72	80.09	75.75	66.13	89.31	95.35	95.36	
	day + month	MC	57.39	89.35	79.99	73.02	65.15	88.86	95.18	95.33	
		GM	58.42	89.75	80.58	73.85	64.67	88.88	95.21	95.33	
3x3	имплицитно	MC	77.36	88.68	82.80	64.77	71.81	95.52	96.39	93.23	
		SM	77.45	88.75	82.89	65.13	71.78	95.67	96.65	93.23	
		GM	77.48	88.76	83.21	65.00	71.74	95.56	96.66	92.52	
	doy	MC	59.39	90.33	80.96	77.52	67.39	90.65	96.52	95.55	
		SM	59.33	90.46	81.78	78.22	66.76	90.62	96.61	95.63	
		GM	59.52	90.50	82.53	77.34	66.59	90.44	96.34	95.67	
	day + month	MC	58.02	90.32	81.59	75.84	65.68	90.43	96.25	95.55	
		SM	57.82	90.44	82.25	76.39	65.16	90.38	96.39	95.63	
		GM	58.05	90.48	82.83	75.63	65.16	90.16	96.21	95.71	
	5x5	имплицитно	MC	77.91	88.81	83.52	65.42	72.37	95.89	96.56	93.23
			SM	78.08	88.84	83.54	66.12	72.22	95.97	96.69	93.23
			GM	78.25	88.82	84.14	65.67	72.01	95.73	96.73	91.54
doy		MC	59.49	90.43	81.99	77.76	67.23	90.87	96.47	95.67	
		SM	59.08	90.51	82.83	78.45	66.46	90.85	96.55	95.74	
		GM	59.12	90.53	83.80	77.32	66.21	90.52	96.32	95.48	
day + month		MC	58.08	90.41	82.54	76.31	65.41	90.62	96.25	95.67	
		SM	57.49	90.49	83.28	76.61	64.85	90.62	96.38	95.74	
		GM	57.53	90.49	84.11	75.77	64.77	90.23	96.19	95.59	

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

Табела П19. Вредности корисникове тачности постигнуте наглашавањем просторне компоненте класификационог модела коришћењем технике град. појачавања (XGB)

Величина суседства	Временска одредница	Правило агрегације	1	2	3	4	5	6	7	8
1x1	имплицитно		74.04	92.22	67.58	74.34	73.37	93.12	96.65	86.39
		MC	80.79	88.26	64.07	68.70	77.48	86.39	96.09	99.48
	doy	SM	82.33	88.42	63.97	69.82	78.00	86.49	96.17	99.48
		GM	82.73	88.55	63.97	71.28	78.21	86.67	96.18	99.48
	day + month	MC	80.99	88.00	63.36	70.42	77.42	86.57	96.11	99.52
		SM	82.50	88.15	63.29	71.68	77.67	86.56	96.17	99.52
		GM	82.89	88.28	63.29	72.81	77.83	86.82	96.16	99.52
3x3	имплицитно	MC	80.41	93.38	67.87	78.07	74.14	94.53	97.02	87.67
		SM	80.50	93.40	67.90	78.47	74.47	94.63	97.02	88.96
		GM	80.58	93.40	67.84	80.00	74.68	95.00	97.14	93.05
	doy	MC	87.54	88.84	64.29	73.20	78.50	87.81	96.58	100.00
		SM	88.26	88.84	64.27	73.67	79.36	87.92	96.63	100.00
		GM	88.49	88.89	64.26	75.03	80.01	87.72	96.92	100.00
day + month	MC	87.48	88.52	63.67	74.08	78.28	88.05	96.60	100.00	
	SM	88.19	88.51	63.63	74.43	79.00	88.11	96.61	100.00	
		GM	88.35	88.55	63.63	75.87	79.57	87.96	96.91	100.00
5x5	имплицитно	MC	80.94	93.56	68.18	80.65	74.95	94.93	97.15	87.55
		SM	80.96	93.59	68.17	80.80	75.23	94.91	97.13	89.20
		GM	81.02	93.62	68.06	82.98	75.71	95.27	97.20	96.15
	doy	MC	88.25	88.85	64.41	75.06	79.43	88.28	96.74	100.00
		SM	88.79	88.78	64.30	75.45	80.37	88.32	96.76	100.00
		GM	88.94	88.78	64.32	76.88	81.24	88.00	97.10	100.00
day + month	MC	88.18	88.53	63.74	75.65	79.18	88.55	96.75	100.00	
	SM	88.61	88.42	63.72	75.86	80.00	88.60	96.73	100.00	
		GM	88.67	88.43	63.74	77.49	80.84	88.21	97.07	100.00

* коришћени су скупови атрибута који су се показали као оптимални у претходним експериментима

БИОГРАФИЈА АУТОРА

Јован Ковачевић је рођен 4. 4. 1993. године у Зрењанину, Република Србија. Основну школу „Др Јован Цвијић“ у Зрењанину је завршио 2008. године, а затим је похађао Зрењанинску гимназију у Зрењанину коју је завршио 2012. године. У току основног и средњошколског образовања био је одличан ученик и носилац Вукове дипломе. Поред тога, Јован поседује и ниже музичко образовање – инструмент кларинет.

На основне академске студије, студијски програм Геодезија и геоинформатика, на Грађевинском факултету Универзитета у Београду Јован Ковачевић се уписао школске 2012/2013. године. Основне академске студије завршио је 29. 9. 2015. године са просечном оценом 9.90, чиме је стекао звање инжењера геодезије. Завршни рад под називом „Могућности примене *DAVID SLS-2* система за израду 3Д модела људског лица“ одбранио је са оценом 10. За успехе постигнуте током основних академских студија проглашен је за најбољег студента Грађевинског факултета у Београду на студијском програму Геодезија и геоинформатика који је дипломирао школске 2014/2015. године.

Мастер академске студије уписао је школске 2015/2016. на Грађевинском факултету Универзитета у Београду, модул Геоинформатика. Током мастер академских студија положио је све испите уз просечну оцену 10.00. Мастер рад под називом „Обрада података ласерског скенирања коришћењем софтверског пакета *LAStools*“ одбранио је 28. 9. 2017. године са оценом 10, чиме је стекао звање мастер инжењера геодезије.

Докторске студије на Грађевинском факултету – студијски програм Геодезија и геоинформатика, уписао је школске 2017/2018. године.

Јован је од 2017. је запослен на Грађевинском факултету Универзитета у Београду као асистент - студент докторских студија. Од 2018. године учествује у пројекту Министарства науке, просвете и технолошког развоја TR36020: „Унапређење геодетске основе Србије за потребе савременог државног премера“.

Поседује одлично знање енглеског и основно знање немачког језика.

ИЗЈАВА О АУТОРСТВУ

Име и презиме аутора Јован Ковачевић

Број индекса 905/17

Изјављујем

да је докторска дисертација под насловом

КАРТИРАЊЕ ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ
ОСМАТРАЊА ЗЕМЉЕ КОРИШЋЕЊЕМ ТЕХНИКА МАШИНСКОГ УЧЕЊА

MAPPING FOREST VEGETATION FROM SATELLITE EARTH OBSERVATION DATA
USING MACHINE LEARNING TECHNIQUES

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 13. 12. 2021.

ИЗЈАВА О ИСТОВЕТНОСТИ ШТАМПАНЕ И ЕЛЕКТРОНСКЕ ВЕРЗИЈЕ ДОКТОРСКОГ РАДА

Име и презиме аутора	Јован Ковачевић
Број индекса	905/17
Студијски програм	Геодезија и геоинформатика
Наслов рада	КАРТИРАЊЕ ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ ОСМАТРАЊА ЗЕМЉЕ КОРИШЋЕЊЕМ ТЕХНИКА МАШИНСКОГ УЧЕЊА MAPPING FOREST VEGETATION FROM SATELLITE EARTH OBSERVATION DATA USING MACHINE LEARNING TECHNIQUES
Ментор:	В. проф. др Жељко Цвијетиновић, дипл. инж. геод.

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у Дигиталном репозиторијуму Универзитета у Београду.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 13. 12. 2021.

ИЗЈАВА О КОРИШЋЕЊУ

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

КАРТИРАЊЕ ШУМСКЕ ВЕГЕТАЦИЈЕ НА ОСНОВУ ПОДАТАКА САТЕЛИТСКОГ
ОСМАТРАЊА ЗЕМЉЕ КОРИШЋЕЊЕМ ТЕХНИКА МАШИНСКОГ УЧЕЊА

MAPPING FOREST VEGETATION FROM SATELLITE EARTH OBSERVATION DATA
USING MACHINE LEARNING TECHNIQUES

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (*Creative Commons*) за коју сам се одлучио/ла.

- 1) Ауторство (CC BY)
- 2) Ауторство – некомерцијално (CC BY-NC)
- 3) Ауторство – некомерцијално – без прерада (CC BY-NC-ND)**
- 4) Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
- 5) Ауторство – без прерада (CC BY-ND)
- 6) Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци. Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 13. 12. 2021.

- 1) **Ауторство.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
- 2) **Ауторство – некомерцијално.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
- 3) **Ауторство – некомерцијално – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
- 4) **Ауторство – некомерцијално – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
- 5) **Ауторство – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
- 6) **Ауторство – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.