UNIVERZITET U NOVOM SADU
FAKULTET TEHNIČKIH NAUKA

Sistemi za podršku odlučivanju u preciznoj poljoprivredi na nivou piksela zasnovani na obradi satelitskih snimaka korišćenjem naprednih algoritama mašinskog učenja

DOKTORSKA DISERTACIJA

Branislav Pejak

Novi Sad, 2026.

УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА **ОБРАЗАЦ – 5а**

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА[1]

Врста рада:	Докторска дисертација
Име и презиме аутора:	Бранислав Пејак
Ментор (титула, име, презиме, звање, институција):	др Никша Јаковљевић, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду; др Оскар Марко, виши научни сарадник, Истраживачко-развојни институт за информационе технологије биосистема БиоСенс, Универзитет у Новом Саду
Наслов рада:	Системи за подршку одлучивању у прецизној пољопривреди на нивоу пиксела засновани на обради сателитских снимака коришћењем напредних алгоритама машинског учења
Језик и писмо рада:	српски језик, латинично писмо
Физички опис рада:	Страница: 187 Поглавља: 12 Референци: 81 Табела: 30 Слика: 23 Графикона: 0 Прилога: 0
Научна област:	Електротехничко и рачунарско инжењерство
Ужа научна област (научна дисциплина):	Обрада сигнала
Кључне речи / предметна одредница:	прецизна пољопривреда; даљинска детекција; сателитски снимци; машинско учење; предвиђање приноса; генеративне адверсаријалне мреже; продуктивност усева; управљачке зоне; системи за подршку одлучивању
Апстракт на језику рада:	Ова докторска дисертација бави се развојем интегрисаног система за подршку одлучивању у прецизној пољопривреди на нивоу пиксела, заснованог на обради сателитских снимака, теренских података и примени напредних алгоритама машинског учења. Истраживање је усмерено на развој методолошког оквира који омогућава просторно експлицитну анализу усева кроз више међусобно повезаних задатака: класификацију усева, процену релативне продуктивности, детекцију потенцијалних штета, процену вероватноће наводњавања, делинеацију хомогених зона унутар парцела, као и предвиђање приноса и протеинског

[1] Аутор докторске дисертације потписао је и приложио следеће Обрасце:
5б – Изјава о ауторству;
5в – Изјава о истоветности штампане и електронске верзије докторског рада и дозвола за објављивање личних података;
5г – Изјава о коришћењу.
Ове Изјаве се чувају у институцији у штампаном и електронском облику и не кориче се са радом.

	састава усева. Основу система чине временске серије Sentinel-1 и Sentinel-2 сателитских података, подаци са монитора приноса, подаци о протеинском саставу, земљишни, топографски и метеоролошки подаци, као и референтне информације о типовима култура, штетним догађајима и наводњаваним парцелама. Посебна пажња посвећена је просторно-временском усклађивању хетерогених извора података и њиховој интеграцији у јединствен аналитички оквир прилагођен потребама прецизне пољопривреде. Један од кључних доприноса рада представља развој SAPI приступа (Satellite Agriculture Productivity Index), којим је омогућена процена релативне продуктивности усева на основу сезонске динамике вегетационог сигнала. Развијени су и модели за класификацију усева, предикцију приноса и протеинског садржаја, као и поступак за делинеацију хомогених управљачких зона и избор репрезентативних места узорковања. Посебан акценат стављен је на примену генеративних модела за допуну и реконструкцију сателитских временских серија у условима облачности. Резултати показују да комбиновање сателитских, теренских и помоћних агроеколошких података омогућава праћење фенолошке динамике усева, анализу просторне хетерогености и моделовање агрономски значајних показатеља на високој просторној резолуцији. Развијени систем представља методолошки конзистентну основу за дигитални мониторинг пољопривредне производње и подршку одлучивању у прецизној и одрживој пољопривреди.
Датум прихватања теме од стране надлежног већа:	
Датум одбране: (Попуњава накнадно институција)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: проф. др Татјана Грбић, редовни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: др Душан Јовановић, редовни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: др Александра Јовановић, редовни професор, Електронски факултет, Универзитет у Нишу Члан: др Бранко Бркљач, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду Члан, ментор: др Никша Јаковљевић, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду Члан, ментор: др Оскар Марко, виши научни сарадник, БиоСенс институт, Универзитет у Новом Саду
Напомена:	

UNIVERSITY OF NOVI SAD
FACULTY OF TECHNICAL SCIENCES

KEY WORD DOCUMENTATION[2]

Document type:	Doctoral dissertation
Author:	Branislav Pejak
Supervisor (title, first name, last name, position, institution)	Dr Nikša Jakovljević, Associate Professor, Faculty of Technical Sciences, University of Novi Sad; Dr Oskar Marko, Senior Research Associate, BioSense Institute – Research and Development Institute for Information Technologies in Biosystems, University of Novi Sad
Thesis title in English:	Pixel-level decision support systems in precision agriculture based on satellite image processing and advanced machine learning algorithms
Language and script:	Serbian, Latin script
Physical description:	Pages: 187 Chapters: 12 References: 81 Tables: 30 Illustrations: 23 Graphs: 0 Appendices: 0
Scientific field:	Electrical and computer engineering
Scientific subfield (scientific discipline):	Signal processing
Subject, Key words:	precision agriculture; remote sensing; satellite imagery; machine learning; crop yield prediction; generative adversarial networks; crop productivity; management zones; decision support systems
Abstract in English:	This doctoral dissertation focuses on the development of an integrated pixel-level decision support system for precision agriculture, based on satellite image processing, field data, and advanced machine learning algorithms. The research establishes a methodological framework for spatially explicit crop analysis through several interconnected tasks: crop classification, assessment of relative productivity, detection of potential crop damage, estimation of irrigation probability, delineation of homogeneous within-field management zones, and prediction of crop yield and protein content.

[2] The author of the doctoral dissertation has signed the following Statements:
5б – Statement on the authorship,
5в – Statement that the printed and e-version of the doctoral dissertation are identical and authorization to use personal data,
5г – Copyright statement.
The paper and e-versions of Statements are held at the institution and are not included into the printed thesis.

	The system is based on Sentinel-1 and Sentinel-2 time series, yield monitor data, protein content measurements, soil, topographic and meteorological data, as well as reference information on crop types, damage events, and irrigated parcels. Particular attention is given to the spatial and temporal harmonization of heterogeneous data sources and their integration into a unified analytical framework suitable for precision agriculture applications. One of the key contributions of the dissertation is the development of the SAPI approach (Satellite Agriculture Productivity Index), which enables the assessment of relative crop productivity based on the seasonal dynamics of vegetation signals. Crop classification models, regression and classification models for yield and protein content prediction, and a procedure for delineating homogeneous management zones and selecting representative sampling locations within fields were also developed. Particular emphasis is placed on generative models for completing and reconstructing satellite time series under cloudy conditions. The results indicate that combining satellite, field, and auxiliary agroecological data enables monitoring of crop phenological dynamics, analysis of spatial heterogeneity, and modeling of agronomically relevant indicators at high spatial resolution. The developed system provides a methodologically consistent basis for digital agricultural monitoring and decision support in precision and sustainable agriculture.
Date of endorsement by the scientific board:	
Date of defence: (Filled in by the institution)	
Thesis defence board: (title, first name, last name, position, institution)	Chair: Prof. Dr Tatjana Grbić, Full Professor, Faculty of Technical Sciences, University of Novi Sad Member: Dr Dušan Jovanović, Full Professor, Faculty of Technical Sciences, University of Novi Sad Member: Dr Aleksandra Jovanović, Full Professor, Faculty of Electronic Engineering, University of Niš Member: Dr Branko Brkljač, Associate Professor, Faculty of Technical Sciences, University of Novi Sad Member, Supervisor: Dr Nikša Jakovljević, Associate Professor, Faculty of Technical Sciences, University of Novi Sad Member, Supervisor: Dr Oskar Marko, Senior Research Associate, BioSense Institute, University of Novi Sad
Note:	

Sažetak

Ova doktorska disertacija bavi se razvojem integrisanog sistema za podršku odlučivanju u preciznoj poljoprivredi na nivou piksela, zasnovanog na obradi satelitskih snimaka, terenskih podataka i primeni naprednih algoritama mašinskog učenja. Istraživanje je usmereno na razvoj metodološkog okvira koji omogućava prostornu analizu useva kroz više međusobno povezanih zadataka: klasifikaciju useva, procenu relativne produktivnosti, detekciju potencijalnih šteta, procenu verovatnoće navodnjavanja, delineaciju homogenih zona unutar parcela, kao i predikciju prinosa i proteinskog sastava useva.

Osnovu sistema čine vremenske serije Sentinel-1 i Sentinel-2 satelitskih podataka, podaci sa monitora prinosa, podaci o proteinskom sastavu, zemljišni, topografski i meteorološki podaci, kao i referentne informacije o tipovima kultura, štetnim događajima i navodnjavanim parcelama. Posebna pažnja posvećena je prostorno-vremenskom usklađivanju heterogenih izvora podataka i njihovoj integraciji u jedinstven analitički okvir prilagođen potrebama precizne poljoprivrede.

Jedan od ključnih doprinosa rada predstavlja razvoj SAPI pristupa (Satellite Agriculture Productivity Index), kojim je omogućena procena relativne produktivnosti useva na osnovu sezonske dinamike vegetacionog signala. Pokazano je da SAPI uspešno identifikuje prostorne obrasce višeg i nižeg proizvodnog potencijala, omogućava analizu anomalija povezanih sa štetnim događajima, kao i preliminarnu procenu verovatnoće navodnjavanja. Dodatno, razvijeni su modeli za klasifikaciju useva, regresioni modeli za predikciju prinosa i proteinskog sastava, kao i postupak za delineaciju homogenih upravljačkih zona i izbor reprezentativnih mesta uzorkovanja unutar parcela.

U radu je poseban akcenat stavljen na primenu generativnih modela za dopunu i rekonstrukciju satelitskih vremenskih serija u uslovima oblačnosti, čime je unapređena konzistentnost ulaznih podataka za dalje modelovanje. Rezultati pokazuju da kombinovanje satelitskih, terenskih i pomoćnih agroekoloških podataka omogućava pouzdano praćenje fenološke dinamike useva, analizu prostorne heterogenosti i modelovanje agronomski značajnih pokazatelja na visokoj prostornoj rezoluciji.

Zaključeno je da razvijeni sistem predstavlja metodološku osnovu za razvoj operativnih sistema za digitalni monitoring poljoprivredne proizvodnje i podršku odlučivanju u preciznoj i održivoj poljoprivredi. Naučni doprinos disertacije ogleda se u razvoju novih pristupa za integraciju višeznačnih izvora podataka, procenu produktivnosti i modelovanje proizvodnih pokazatelja, dok se praktični doprinos ogleda u mogućnosti

njihove implementacije u operativni Geografski Informacioni Sistem (GIS) i softverske alate namenjene upravljanju poljoprivrednim površinama.

Abstract

This doctoral dissertation focuses on the development of an integrated pixel-level decision support system for precision agriculture, based on satellite image processing, field data, and advanced machine learning algorithms. The research is aimed at establishing a methodological framework for spatially explicit crop analysis through several interconnected tasks: crop classification, assessment of relative productivity, detection of potential crop damage, estimation of irrigation probability, delineation of homogeneous within-field management zones, and prediction of crop yield and protein content.

The system is based on Sentinel-1 and Sentinel-2 time series, yield monitor data, protein content measurements, soil, topographic and meteorological data, as well as reference information on crop types, damage events, and irrigated parcels. Particular attention is given to the spatial and temporal harmonization of heterogeneous data sources and their integration into a unified analytical framework suitable for precision agriculture applications.

One of the key contributions of the dissertation is the development of the Satellite Agriculture Productivity Index (SAPI) approach, which enables the assessment of relative crop productivity based on the seasonal dynamics of vegetation signals. The results show that SAPI successfully identifies spatial patterns of higher and lower production potential, supports the analysis of anomalies related to damaging events, and enables a preliminary estimation of irrigation probability. In addition, crop classification models, regression models for yield and protein content prediction, and a procedure for delineating homogeneous management zones and selecting representative sampling locations within fields were developed.

A particular emphasis of the study is placed on the application of generative models for the completion and reconstruction of satellite time series under cloudy conditions, thereby improving the consistency of input data for downstream modeling. The results indicate that combining satellite, field, and auxiliary agroecological data enables reliable monitoring of crop phenological dynamics, analysis of spatial heterogeneity, and modeling of agronomically relevant indicators at high spatial resolution.

It is concluded that the developed system provides a methodological basis for the development of operational systems for digital agricultural monitoring and decision support in precision and sustainable agriculture. The scientific contribution of the dissertation lies in the development of novel approaches for multi-source data integra-

tion, productivity assessment, and modeling of production-related indicators, while its practical contribution is reflected in the possibility of implementing these methods in operational GIS and software tools intended for agricultural land management.

Zahvalnica

Želim da izrazim iskrenu zahvalnost svima koji su na različite načine doprineli izradi ove doktorske disertacije i pružili mi podršku tokom mog dosadašnjeg akademskog i istraživačkog puta. Pre svega, zahvaljujem se mentorima na stručnom vođstvu, strpljenju, korisnim savetima i kontinuiranoj podršci tokom svih faza istraživanja. Njihovo znanje, iskustvo i poverenje bili su od velikog značaja za oblikovanje ove disertacije. Posebnu zahvalnost dugujem članovima komisije, kao i svim profesorima i kolegama sa fakulteta i instituta, koji su svojim sugestijama, diskusijama i profesionalnim odnosom doprineli kvalitetu ovog rada i mom naučnom razvoju. Zahvaljujem se i svojim kolegama i prijateljima, koji su ovaj put učinili lakšim i lepšim kroz saradnju, razmenu ideja, ali i kroz onu dragocenu svakodnevnu podršku koja često znači više nego što se može izraziti rečima. Najveću zahvalnost dugujem svojoj porodici, na bezuslovnoj ljubavi, razumevanju, strpljenju i veri u mene da ću uspeti na ovom putu. Njihova podrška bila je oslonac u trenucima umora, sumnje i izazova. Bez njih ovaj put ne bi bio moguć.

Na kraju, hvala svima koji su me podsećali da i najduži put može da se završi korak po korak.

Sadržaj

Sažetak

Abstract

Zahvalnica

1 Uvod **1**

1.1 Poljoprivreda u Srbiji . . . 1
1.2 Predmet i cilj istraživanja . . . 2
1.3 Istraživačka pitanja i hipoteze . . . 3
1.4 Naučni doprinos teze . . . 3
1.5 Organizacija disertacije . . . 5

2 Pregled literature i teorijske osnove **6**

2.1 Pregled oblasti i istraživački kontekst . . . 6
2.2 Daljinska detekcija . . . 6
2.3 Mašinsko učenje i predikcija prinosa . . . 7
2.4 Upravljanje zonama u parceli . . . 10
2.5 Generativni algoritmi . . . 10
2.6 Pregled korišćenih modela mašinskog učenja . . . 12
2.6.1 Stabla odlučivanja i slučajna šuma (Random Forest) . . . 13
2.6.2 Višestruka linearna regresija (Multiple Linear Regression) . . . 17
2.6.3 Metoda potpornih vektora (Support Vector Machine) . . . 20
2.6.4 eXtreme Gradient Boosting . . . 21
2.6.5 Duboke neuronske mreže (Deep Neural Network) . . . 23
2.6.6 Long Short-Term Memory . . . 26
2.6.7 Konvolucione neuronske mreže (Convolutional Neural Network) 30
2.6.8 K-means . . . 31

3 Podaci i područje istraživanja **34**

3.1 Područje istraživanja i organizacija podataka . . . 34
3.2 Satelitski snimci . . . 35
3.2.1 Radarski satelitski snimci . . . 35
3.2.2 Optički satelitski snimci . . . 38
3.3 Monitor prinosa . . . 41
3.4 Podaci zemljišnih karakteristika . . . 43

3.5 Digitalni elevacioni model . . . 46
3.6 Meteo podaci . . . 46
3.7 Podaci o vrstama poljoprivrednih kultura . . . 48
3.8 Podaci o datumima štete . . . 50
3.9 Podaci o navodnjavanim parcelama . . . 52
3.10 Podaci o proteinskom sastavu . . . 55
3.11 Zbirni pregled i uloga skupova podataka . . . 55

4 Opšti metodološki okvir i priprema podataka 58
4.1 Arhitektura i metodološki okvir predloženog sistema . . . 58
4.1.1 Ulazni podaci sistema . . . 61
4.1.2 Predobrada podataka . . . 62
4.1.3 Integracija skupova podataka . . . 63
4.1.4 Modeli sistema . . . 64
4.1.5 Izlazni podaci sistema . . . 66
4.1.6 Grafički korisnički interfejs . . . 67
4.2 Sistem za klasifikaciju useva . . . 67
4.3 Sistem za predikciju prinosa i proteinskog sastava useva . . . 68
4.4 Sistem za generisanje nedostajućih satelitskih podataka . . . 69
4.5 Sistem za zoniranje parcele . . . 70
4.6 Obrada vektorskih podataka . . . 71
4.6.1 Obrada podataka sa monitora prinosa . . . 71
4.7 Predobrada satelitskih slika . . . 75
4.7.1 Izračunavanje vegetacionih indeksa . . . 75
4.7.2 Predobrada satelitskih slika za CNN . . . 78
4.7.3 Generisanje obeležja pomoću GAN . . . 79

5 Rekonstrukcija nedostajućih satelitskih opservacija 86
5.1 GAN generisane satelitske slike . . . 86
5.1.1 Formiranje skupa podataka za obuku GAN . . . 86
5.1.2 Upotreba GAN modela u svrhu generisanja optičkih satelitskih snimaka . . . 88
5.1.3 Rezultati rekonstrukcije Sentinel-2 snimaka . . . 88

6 Klasifikacija useva 92
6.1 Klasifikacija useva . . . 92
6.1.1 Ulazni podaci modela . . . 94
6.1.2 Obuka klasifikatora . . . 95

6.1.3 Generisanje finalnih mapa klasifikacije useva 96
6.1.4 Metodološke prednosti i ograničenja 98

7 Predikcija prinosa i proteinskog sadržaja 99
7.1 Predikcija prinosa useva 99
7.1.1 Priprema satelitskih podataka 100
7.1.2 Referentni podaci o prinosu i prostorna priprema uzoraka 100
7.1.3 Konstrukcija vremenskih serija spektralnih atributa 101
7.1.4 Eksperimentalni dizajn: realni, GAN generisani i kombinovani atributi 102
7.1.5 Validacija metodom izostavljanja jedne parcele - Leave-One-Parcel-Out 102
7.1.6 Metodološke prednosti pristupa 103
7.1.7 Finalne mape predikcije prinosa useva 103
7.1.8 Rezultati modela za predikciju prinosa 104
7.1.9 Rezultati predikcije prinosa za 2018. godinu 109
7.1.10 Rezultati za 2019. godinu 112
7.1.11 Poređenje modela i eksperimentalnih režima 116
7.1.12 Interpretacija rezultata u odnosu na vrednosti prinosa 117
7.1.13 Evaluacija generalizacije modela na podacima iz različitih izvora 121
7.2 Predikcija proteinskog sastava useva 123
7.2.1 Opis podataka, transformacija vremenskih serija i konstrukcija ulaznih obeležja 123
7.2.2 Klasifikacija proteinskog sastava 125
7.2.3 Rezultati klasifikacije proteinskog sastava 125

8 SAPI i njegove primene 127
8.1 Mape produktivnosti useva – SAPI (Satellite Agriculture Productivity Index) 127
8.1.1 Ulazni podaci i izračunavanje spektralnih pokazatelja 128
8.1.2 Rekonstrukcija vremenske serije i integral sezonske aktivnosti . . 129
8.1.3 Klasifikacija relativne produktivnosti 129
8.1.4 Višegodišnja agregacija i prostorna interpretacija 130
8.1.5 Poređenje SAPI klasa i prinosa useva u 2018. godini 130
8.1.6 Poređenje SAPI klasa i prinosa useva u 2019. godini 134
8.1.7 Značaj i primena SAPI indeksa 138
8.2 Analiza anomalija produktivnosti povezanih sa štetnim događajima . . 138

8.2.1 Formiranje baze olujnih događaja, geolokacijska standardizacija i povezivanje olujnih događaja sa katastarskim opštinama 140
8.2.2 Izdvajanje statistika produktivnosti po katastarskim opštinama iz SAPI mapa 142
8.2.3 Rezultati analize anomalija u katastarskim opštinama pogođenim olujama 143
8.3 Analiza povezanosti SAPI sa poznatim navodnjavanim površinama . . . 143
8.3.1 Analiza usklađenosti između navodnjavanja i produktivnosti useva zasnovane na SAPI indeksu 145

9 Delineacija upravljačkih zona i izbor mesta uzorkovanja 150
9.1 Delineacija zona i određivanje mesta uzorkovanja unutar poljoprivrednih parcela 150
9.1.1 Ulazni podaci i formiranje spektralnih obeležja za segmentaciju 151
9.1.2 Klasterizacija piksela unutar parcele, prostorna homogenizacija i uklanjanje sitnih fragmenata 152
9.1.3 Određivanje reprezentativnih mesta uzorkovanja unutar zona . . 153
9.1.4 Izlazni proizvodi modela za određivanje homogenih zona 154
9.1.5 Validacija modela delineacije zona i strategije uzorkovanja . . . 155

10 Integracija modula i grafički korisnički interfejs 161
10.1 Povezivanje izlaza pojedinačnih modula 161
10.2 Grafički korisnički interfejs 162
10.3 Primer operativnog toka 165

11 Diskusija 167
11.0.1 Rekonstrukcija satelitskih podataka i predikcija prinosa 167
11.0.2 Klasifikacija useva, SAPI i prostorna analiza poljoprivrednih površina 170
11.0.3 Ograničenja i praktični značaj razvijenog sistema 172
11.0.4 Pravci budućih istraživanja 173

12 Zaključak 175
12.1 Odgovori na istraživačka pitanja i završni zaključci 175
12.2 Status postavljenih hipoteza 178
12.3 Završna napomena 179

Bibliografija 180

Spisak slika

3.1 Prikaz područja AP Vojvodine na osnovu Sentinel-1 radarskog snimka u VV polarizaciji. Svetlije vrednosti odgovaraju većem intenzitetu povratnog radarskog signala, dok tamnije oblasti predstavljaju površine sa manjim povratnim signalom. 36

3.2 Prostorni raspored Sentinel-2 granula koje prekrivaju teritoriju Autonomne pokrajine Vojvodine. Bela mreža predstavlja granice Sentinel-2 granula sa pripadajućim oznakama, dok je crvenom linijom prikazana državna granica Srbije. 40

3.3 Prostorna raspodela organskog ugljenika u zemljištu (SOC stock) u površinskom sloju od 0 do 30 cm, prema podacima SoilGrids. Zelenije nijanse označavaju područja sa većom koncetracijom organskog ugljenika, dok žute nijanse predstavljaju područja sa nižim vrednostima. 44

3.4 Digitalni elevacioni model (DEM) područja AP Vojvodine prostorne rezolucije 10 m. Zelene nijanse predstavljaju područja niže nadmorske visine, dok žute i braon nijanse označavaju više terene. 47

3.5 Korisnički interfejs mobilne aplikacije FieldApp za terensko prikupljanje podataka o poljoprivrednim kulturama. Aplikacija omogućava unos podataka o usevu, predusevu, sorti ili hibridu, datumu setve, fenofazi, opštem stanju biljke i prisustvu bolesti, korova i štetočina. Prikupljeni podaci mogu se koristiti kao referentne informacije za validaciju satelitskih analiza i modela mašinskog učenja. 49

3.6 Prostorni prikaz merenog sadržaja proteina unutar poljoprivredne parcele. Tamnije zelene nijanse predstavljaju više vrednosti procentualnog sadržaja proteina, dok svetlije zelene i bele nijanse označavaju niže vrednosti. Tačke predstavljaju pojedinačna merenja prikupljena duž putanje kretanja mehanizacije. 54

4.1 Dijagram toka obrade podataka za analizu poljoprivrednih useva korišćenjem satelitskih podataka i metoda mašinskog učenja. 60

4.2 Prikaz podataka sa monitora prinosa gde svaka tačka predstavlja merenu vrednost prinosa. 72

4.3 Grafički prikaz preklapanja poligona i piksela za PPI metodu. 73

4.4 Poligoni koji odgovaraju površini sa koje je soja požnjevena za svako pojedinačno merenje prinosa. 73
4.5 Mreža Sentinel-2 piksela (sivi kvadranti) postavljena preko poligona prinosa (magenta). 74
4.6 Grafički prikaz mreže Sentinel-2 piksela. Granični pikseli isključeni iz analize označeni su crvenom bojom, dok su pikseli za dalju obradu označeni sivom bojom. 75

5.1 Vizuelno poređenje originalnog Sentinel-2 snimka (slika levo) i rekonstruisanog snimka (slika desno 06.08.2025.) iz test skupa podataka. Oblast označena crvenom bojom je region koji je bio zahvaćen oblacima 91

6.1 Grafički prikaz klasifikacije 6 najzastupljenijih useva u Vojvodini. . . . 97

7.1 Grafički prikaz predikcije prinosa na nivou piksela za kukuruz, pšenicu, soju i suncokret na teritoriji AP Vojvodine 105
7.2 Grafički prikaz predikcije prinosa na nivou piksela za kukuruz, pšenicu, soju i suncokret na teritoriji AP Vojvodine, sezona 2019. 106

8.1 Grafički prikaz mape produktivnosti useva za 2019. godinu (SAPI). . . 138
8.2 Grafički prikaz parcele sa SAPI indeksima produktivnosti po pikselu. . 139
8.3 Dijagram NDVI vegetacionog indeksa kroz vreme. U trenutku t(110) može se videti nagli pad vrednosti indeksa koji je prouzrokovan pojavom oblaka. 140
8.4 Poređenje distribucija SAPI vrednosti između navodnjavanih i nenavodnjavanih piksela za tri posmatrane kulture u 2021. i 2022. godini. 149

9.1 Primer zoniranja parcele za potrebe uzorkovanja zemljišta. Parcela je podeljena na tri zone sličnih karakteristika, dok crveni markeri predstavljaju preporučene tačke za uzorkovanje unutar izdvojenih zona. 155
9.2 Poređenje tri strategije uzorkovanja unutar prethodno definisanih homogenih zona. Različite boje predstavljaju izdvojene prostorne zone, dok crne tačke označavaju mesta terenskog uzorkovanja. (a) GRID – uzorkovanje prema pravilnoj prostornoj mreži, koje predstavlja konvencionalni pristup; (b) OPT – uzorkovanje na lokacijama određenim razvijenim algoritmom, pri čemu je broj i položaj uzoraka optimizovan u odnosu na geometriju i prostornu strukturu zona; (v) ONE – redukovani pristup sa jednom reprezentativnom tačkom uzorkovanja po homogenoj zoni. . . . 157

10.1 Prikaz grafičkog korisničkog interfejsa za učitavanje i vizuelizaciju tematskih rastera. Interfejs omogućava podešavanje transparentnosti i palete boja (*colormap*), interaktivno očitavanje vrednosti piksela i prikaz osnovnih statističkih pokazatelja aktivnog sloja. 165

Spisak tabela

3.1 Osnovne karakteristike polarizacionih kanala Sentinel-1 IW podataka . 36
3.2 Osnovne karakteristike režima snimanja Sentinel-1 satelita 37
3.3 Karakteristike Sentinel-1 IW snimanja područja AP Vojvodine 38
3.4 Spektralni kanali multispektralnog senzora na Sentinel-2 satelitima . . 39
3.5 Zbirni pregled skupova podataka korišćenih u disertaciji 57

5.1 Rezultati evaluacije GAN rekonstrukcije Sentinel-2 snimaka po datumima za sezonu 2018 89
5.2 Rezultati evaluacije GAN rekonstrukcije Sentinel-2 snimaka po datumima 89

6.1 Performanse klasifikacije po klasama useva za 2019. godinu (TPR, FNR, FPR, TNR, PREC, FDR, ACC, INACC, F-score) 98

7.1 Reprezentativni fiksni hiperparametri i korišćene Python biblioteke . . 106
7.2 Rezultati svih modela za predikciju prinosa kukuruza u 2018. godini. . 109
7.3 Rezultati svih modela za predikciju prinosa soje u 2018. godini. 110
7.4 Rezultati svih modela za predikciju prinosa pšenice u 2018. godini. . . 111
7.5 Rezultati svih modela za predikciju prinosa suncokreta u 2018. godini. 112
7.6 Rezultati svih modela za predikciju prinosa kukuruza u 2019. godini. . 113
7.7 Rezultati svih modela za predikciju prinosa soje u 2019. godini. 114
7.8 Rezultati svih modela za predikciju prinosa pšenice u 2019. godini. . . 115
7.9 Rezultati svih modela za predikciju prinosa suncokreta u 2019. godini. 116
7.10 Prosečne vrednosti prinosa i najbolje ostvarene greške predikcije izražene u $\mathrm{t\,ha^{-1}}$. 118
7.11 Rezultati testiranja modela pri scenariju treniranja na podacima iz 2018. godine i testiranja na podacima iz 2019. godine. Oznaka *sat* odnosi se na satelitske ulazne podatke, *soil* na zemljišne karakteristike, dok *meteo* označava meteorološke promenljive. Vrednosti grešaka izražene su u t/pikselu. 122
7.12 Rezultati binarne klasifikacije proteinskog sastava ($\leq 40\%$ naspram $> 40\%$) u režimu *leave-one-source-out* evaluacije. 126

8.1 Broj piksela i prinos kukuruza po SAPI klasama u 2018. godini 131

8.2 Broj piksela i prinos soje po SAPI klasama u 2018. godini 132
8.3 Broj piksela i prinos suncokreta po SAPI klasama u 2018. godini 132
8.4 Broj piksela i prinos pšenice po SAPI klasama u 2018. godini 133
8.5 Broj piksela i prinos kukuruza po SAPI klasama u 2019. godini 134
8.6 Broj piksela i prinos soje po SAPI klasama u 2019. godini 135
8.7 Broj piksela i prinos suncokreta po SAPI klasama u 2019. godini 136
8.8 Broj piksela i prinos pšenice po SAPI klasama u 2019. godini 137
8.9 Poređenje SAPI vrednosti između navodnjavanih i nenavodnjavanih piksela za 2021. i 2022. godinu. Kolona "Broj uzoraka" označava broj nenavodnjavanih piksela, dok N_1 označava broj navodnjavanih piksela. Vrednost μ_{nenav} predstavlja prosečnu SAPI vrednost nenavodnjavanih piksela, μ_{nav} prosečnu SAPI vrednost navodnjavanih piksela, dok $\Delta\mu$ označava razliku između prosečnih vrednosti navodnjavanih i nenavodnjavanih piksela, odnosno $\Delta\mu = \mu_{nav} - \mu_{nenav}$. Cliff's δ predstavlja meru veličine efekta između dve grupe. 147

9.1 Sažetak rezultata validacije delineiranih zona i poređenja GRID i OPT strategija uzorkovanja 158

Poglavlje 1

Uvod

Sadržaj

1.1 Poljoprivreda u Srbiji 1
1.2 Predmet i cilj istraživanja 2
1.3 Istraživačka pitanja i hipoteze 3
1.4 Naučni doprinos teze 3
1.5 Organizacija disertacije 5

1.1 Poljoprivreda u Srbiji

Poljoprivreda je jedna od osnovnih i glavnih grana privrede za mnoge zemlje u svetu, jer obezbeđuje hranu, sirovine za industriju i značajan deo zaposlenosti u ruralnim područjima. Održiva poljoprivredna proizvodnja postaje sve važnija u kontekstu globalnih izazova, kao što su klimatske promene, rast populacije i smanjenje dostupnosti obradivog zemljišta. Jedan od ključnih faktora za definisanje održive poljoprivredne proizvodnje je procena količine prinosa pred žetvu i na kraju sezone Jabed and Murad 2024. Ova procena ne samo da omogućava efikasnije planiranje proizvodnje, već doprinosi i stabilizaciji tržišta kroz formiranje strategija u vezi sa cenama, uvozom, izvozom, procenom štete nakon elementarnih nepogoda, skladištenjem i određivanjem krajnjih potreba za poljoprivrednim proizvodima.

Precizne informacije o prinosu omogućavaju donositeljima odluka, kako na nivou pojedinačnih gazdinstava, tako i na nivou državne politike, da pravovremeno reaguju na tržišne fluktuacije, smanje gubitke i optimizuju lance snabdevanja. Digitalizacija i primena savremenih tehnologija, kao što su daljinska detekcija, satelitsko praćenje i digitalni blizanci poljoprivrednih sistema, omogućavaju preciznije i brže procene, što dodatno povećava efikasnost celokupnog sektora.

Vodeći se ovim činjenicama, može se zaključiti da poznavanje količine i dostupnosti određenog proizvoda može značajno uticati na celokupni ekonomski sistem, bilo sa gledišta proizvođača, otkupljivača, prerađivača ili državne strategije. Udeo poljoprivrede

u bruto domaćem proizvodu (BDP) Republike Srbije iznosi oko 6-7%, što je i dalje čini jednom od vodećih grana privrede u zemlji Mitrović et al. 2017. Ovaj sektor ne samo da obezbeđuje stabilne prihode za veliki deo stanovništva, već ima i značajnu ulogu u očuvanju ruralnih područja, unapređenju prehrambene sigurnosti i zaštiti prirodnih resursa. Razvoj održivih metoda proizvodnje i modernizacija poljoprivrednih praksi ključni su faktori za povećanje konkurentnosti Srbije na međunarodnom tržištu, dok istovremeno doprinose zaštiti ekosistema i racionalnom korišćenju vodnih resursa.

1.2 Predmet i cilj istraživanja

Predmet ove doktorske disertacije jeste razvoj i evaluacija integrisanog metodološkog okvira za podršku odlučivanju u preciznoj poljoprivredi na nivou piksela. Okvir povezuje viševremenske Sentinel-1 i Sentinel-2 podatke, terenska merenja prinosa i proteinskog sastava, prostorne podatke o parcelama, kao i zemljišne, topografske i meteorološke promenljive. Opšti cilj istraživanja jeste da se ispita u kojoj meri kombinovanje ovih izvora podataka i modela mašinskog učenja omogućava pouzdanu prostornu analizu useva, uz posebno razmatranje generalizacije između parcela, godina i izvora podataka.

Polazna inženjerska potreba ovog istraživanja jeste da se iz heterogenih i često nepotpunih podataka dobije skup pouzdanih prostornih informacija koje mogu pomoći u donošenju odluka na nivou parcele, regiona ili šireg sistema monitoringa. U tom kontekstu, potrebno je odgovoriti na nekoliko praktičnih pitanja: koja se kultura nalazi na posmatranoj površini, kakvo je relativno stanje i proizvodni potencijal useva, gde se unutar parcele javljaju stabilno produktivnije ili slabije zone, da li se mogu uočiti prostorne anomalije povezane sa štetnim događajima, u kojoj meri se poznate navodnjavane površine razlikuju od nenavodnjavanih prema satelitski izvedenim pokazateljima, kao i koliko pouzdano je moguće proceniti prinos ili proteinski sadržaj na osnovu dostupnih daljinskih i pomoćnih podataka.

Pojedinačni ciljevi obuhvataju: rekonstrukciju nedostajućih optičkih opservacija u uslovima oblačnosti; klasifikaciju poljoprivrednih kultura; predikciju prinosa i klasifikaciju proteinskog sadržaja; razvoj i ispitivanje SAPI (Satellite Agricultural Productivity Index) pokazatelja relativne produktivnosti; analizu anomalija povezanih sa štetnim događajima i poznatim navodnjavanim površinama; kao i delineaciju homogenih zona i izbor reprezentativnih mesta uzorkovanja.

1.3 Istraživačka pitanja i hipoteze

Istraživanje je organizovano oko sledećih pitanja: (1) da li prostorno konzistentna validacija izlaza modela na nivou parcela daje realističnu procenu generalizacije modela; (2) da li rekonstruisane optičke opservacije mogu da dopune vremenske serije bez značajnog gubitka informacija relevantnih za dalje modelovanje; (3) u kojoj meri sezonska dinamika vegetacionog signala omogućava procenu relativne produktivnosti; i (4) da li prostorno regularizovana klasterizacija može da izdvoji praktično upotrebljive unutarparcelne zone.

Na osnovu navedenih pitanja postavljene su radne hipoteze: **H1** – generativno rekonstruisane opservacije mogu da obezbede uporedivu korisnost sa realnim opservacijama; **H2** – integral sezonske aktivnosti vegetacije sadrži informaciju povezanu sa prostornim obrascima relativne produktivnosti; **H3** – poznate navodnjavane površine u proseku pokazuju više SAPI vrednosti; i **H4** – kombinovanje spektralne klasterizacije i prostorne homogenizacije daje kompaktnije zone pogodne za planiranje uzorkovanja.

1.4 Naučni doprinos teze

Naučni doprinos ove doktorske disertacije ogleda se u razvoju **integrisanog metodološkog okvira za analizu poljoprivrednih površina na nivou piksela, zasnovanog na kombinaciji satelitskih snimaka, terenskih podataka i naprednih algoritama mašinskog učenja**. U okviru disertacije razvijen je sistem koji objedinjuje više međusobno povezanih zadataka od značaja za preciznu poljoprivredu, uključujući klasifikaciju useva, predikciju prinosa i proteinskog sastava, procenu relativne produktivnosti, analizu anomalija povezanih sa štetnim događajima, procenu povezanosti sa navodnjavanjem, kao i delineaciju homogenih zona unutar parcela.

Prvi naučni doprinos odnosi se na razvoj integrisanog arhitekturnog i metodološkog okvira za povezivanje heterogenih izvora podataka, uključujući Sentinel-1 i Sentinel-2 satelitske snimke, podatke sa monitora prinosa, zemljišne, topografske i meteorološke podatke, kao i referentne podatke o tipovima useva, štetnim događajima i navodnjavanim parcelama. Predloženi okvir omogućava prostorno-vremensko usklađivanje podataka različite prirode i rezolucije i njihovu transformaciju u jedinstven analitički prostor pogodan za modelovanje na nivou piksela.

Drugi naučni doprinos ogleda se u razvoju pristupa za procenu relativne produktivnosti useva primenom indeksa SAPI (*Satellite Agriculture Productivity Index*). Ovim pristupom omogućena je višegodišnja analiza prostorne varijabilnosti proizvodnog potencijala na osnovu sezonske dinamike satelitski izvedenih pokazatelja vegetacije. Na taj

način uveden je metod za prostornu identifikaciju stabilno produktivnijih i manje produktivnih zona unutar poljoprivrednog prostora, čime je otvorena mogućnost za analizu dugoročnih obrazaca proizvodnje i donošenje odluka u okviru precizne poljoprivrede.

Treći naučni doprinos odnosi se na razvoj metodologije za primenu SAPI pristupa u dodatnim analitičkim zadacima, odnosno za procenu anomalija produktivnosti u područjima pogođenim štetnim događajima i za indirektnu procenu verovatnoće navodnjavanja parcela. Time je pokazano da satelitski izvedeni indeksi relativne produktivnosti mogu imati širu analitičku vrednost od same procene proizvodnog potencijala, odnosno da mogu poslužiti kao osnova za interpretaciju različitih procesa i pojava relevantnih za upravljanje poljoprivrednom proizvodnjom.

Četvrti naučni doprinos predstavlja razvoj modela za klasifikaciju useva, kao i regresionih i klasifikacionih modela za predikciju prinosa i proteinskog sastava useva na osnovu vremenskih serija satelitskih snimaka i dodatnih agroekoloških promenljivih. Poseban značaj ovog doprinosa ogleda se u primeni prostorno konzistentnih strategija validacije, čime je omogućena realističnija procena sposobnosti generalizacije modela u uslovima prostorne heterogenosti i parcelno uslovljene autokorelacije.

Peti naučni doprinos odnosi se na razvoj metodologije za delineaciju homogenih zona unutar poljoprivrednih parcela i određivanje reprezentativnih mesta uzorkovanja. Kombinacijom spektralnih obeležja, klasterizacije i prostorne homogenizacije omogućeno je izdvajanje upravljačkih zona koje predstavljaju osnovu za preciznije terensko uzorkovanje, diferencirano upravljanje parcelama i bolje razumevanje unutarparcelne heterogenosti.

Šesti naučni doprinos ogleda se u razvoju operativno primenljive metodologije za dopunu i rekonstrukciju satelitskih vremenskih serija u uslovima oblačnosti, zasnovane na generativnim modelima, sa posebnim fokusom na generativne adversarijalne mreže. Predloženi pristup omogućava nadomeštanje nedostajućih multispektralnih informacija i izvedenih vegetacionih indeksa, čime se obezbeđuje vremenski kontinualan i konzistentan skup ulaznih obeležja kroz više vegetacionih sezona. Na taj način unapređuje se stabilnost i robusnost modela koji zavise od gustih i vremenski usklađenih satelitskih vremenskih serija.

Sedmi od doprinosa disertacije predstavlja formiranje i harmonizacija višeslojne geoprostorne istraživačke baze podataka za područje AP Vojvodine. Baza objedinjuje višegodišnje Sentinel-1 i Sentinel-2 satelitske podatke, merenja prinosa, podatke o proteinskom sadržaju, informacije o navodnjavanju i olujnim događajima, kao i izvedene prostorne proizvode dobijene klasifikacijom useva, rekonstrukcijom oblačnih optičkih snimaka, izračunavanjem SAPI i delineacijom upravljačkih zona. Podaci su prostorno i vremenski usklađeni i organizovani tako da omogućavaju zajedničku primenu u modelima mašinskog učenja i njihovoj nezavisnoj evaluaciji.

Ukupno posmatrano, naučni doprinos disertacije ogleda se u povezivanju metoda daljinske detekcije, mašinskog učenja i geoprostorne analize u jedinstven sistem za podršku odlučivanju u preciznoj poljoprivredi. Razvijeni pristupi imaju ne samo teorijski i metodološki značaj, već i izražen potencijal za operativnu primenu u digitalnom monitoringu useva, upravljanju parcelama i razvoju savremenih softverskih alata za poljoprivredu.

1.5 Organizacija disertacije

U uvodnom delu nalazi se objašnjene i motivacija za izradu doktorske disertacije, dok su u drugom poglavlju prikazani pregled literature i teorijske osnove korišćenih metoda. Treće poglavlje opisuje područje istraživanja i skupove podataka, dok četvrto daje zajednički metodološki okvir, predobradu i integraciju podataka. Poglavlja od petog do devetog predstavljaju pojedinačne eksperimentalne celine: rekonstrukciju satelitskih opservacija, klasifikaciju useva, modelovanje prinosa i proteinskog sadržaja, SAPI i njegove primene, kao i delineaciju upravljačkih zona. Deseto poglavlje opisuje integraciju izlaza i grafički interfejs. U jedanaestom poglavlju rezultati su objedinjeno diskutovani, a u dvanaestom su dati zaključci, ograničenja i pravci daljeg rada.

Poglavlje 2

Pregled literature i teorijske osnove

Sadržaj

2.1 Pregled oblasti i istraživački kontekst 6
2.2 Daljinska detekcija 6
2.3 Mašinsko učenje i predikcija prinosa 7
2.4 Upravljanje zonama u parceli 10
2.5 Generativni algoritmi 10
2.6 Pregled korišćenih modela mašinskog učenja 12
2.6.1 Stabla odlučivanja i slučajna šuma (Random Forest) 13
2.6.2 Višestruka linearna regresija (Multiple Linear Regression) 17
2.6.3 Metoda potpornih vektora (Support Vector Machine) 20
2.6.4 eXtreme Gradient Boosting 21
2.6.5 Duboke neuronske mreže (Deep Neural Network) 23
2.6.6 Long Short-Term Memory 26
2.6.7 Konvolucione neuronske mreže (Convolutional Neural Network) 30
2.6.8 K-means 31

2.1 Pregled oblasti i istraživački kontekst

Ovo poglavlje sistematizuje teorijske osnove i prethodna istraživanja relevantna za razvijeni sistem. Najpre su razmotrene primene daljinske detekcije u poljoprivredi, zatim modelovanje prinosa, upravljačke zone i generativni pristupi, a potom matematičke i algoritamske osnove modela primenjenih u eksperimentima.

2.2 Daljinska detekcija

Upotreba platformi za daljinsku detekciju, poput satelita, značajno je poboljšala našu sposobnost praćenja i upravljanja u oblasti poljoprivrede. U okviru doktorske disertacije su opisane različite primene satelitskih snimaka za potrebe precizne poljoprivrede.

Snimci koji će činiti osnovu ovog istraživanja preuzeti su sa Sentinel-1 i Sentinel-2 satelita, koji su lansirani u okviru programa Copernicus pod pokroviteljstvom Evropske svemirske agencije (ESA). Copernicus je moderan svemirski program čiji sateliti služe za opservaciju Zemljine površine. Sateliti ovog programa generišu snimke visoke rezolucije (10 m $\times$ 10 m) na svakih 5 dana u godini. Podaci predstavljaju vremensku seriju multispektralnih i SAR (Synthetic Aperture Radar) snimaka nad regionom od interesa. Da bi se dobio što realniji prikaz stanja useva, neophodno je koristiti vremenske serije satelitskih snimaka kako bi se obuhvatili svi specifični vegetacioni periodi kultura od interesa Agency 2021.

Moderni satelitski senzori, kao što su multispektralne kamere, mogu da se koriste za otkrivanje promena na vegetaciji. Na očitavanje senzora satelita u infracrvenom delu elektromagnetnog spektra najviše utiče struktura biljnog pokrivača, dok na očitavanje u mikrotalasnom opsegu najveći uticaj ima količina, odnosno sadržaj vode u biljkama Anatoly A. Gitelson and Mark N. Merzlyak 2022. Najveća prednost satelitskih snimaka u odnosu na snimke sa ostalih platformi, kao što su avioni ili dronovi, je ta što sateliti mogu da osmatraju mnogo veće površine.

Satelitske tehnologije pružaju važne podatke kako za poljoprivredne proizvođače, tako i za državne institucije. Informacije dobijene iz satelitskih snimaka mogu se koristiti za razvoj poljoprivredne proizvodnje, zonalni menadžment parcela, logistiku, predviđanje cena, planiranje skladištenja, strategije uvoza i izvoza i slično Kooistra et al. 2023; Maimaitijiang et al. 2020.

Satelitske platforme postaju sve dostupnije kako za istraživačke, tako i za komercijalne svrhe. Njihova upotreba omogućava praćenje vegetacije na velikim površinama, dok se satelitske slike koriste za globalni monitoring stanja useva Gowda et al. 2026.

2.3 Mašinsko učenje i predikcija prinosa

Upotrebom satelitskih snimaka u kombinaciji sa metodama mašinskog učenja može se unaprediti klasifikacija useva Khan et al. 2026, što predstavlja osnovni sloj za sledeći korak ka predikciji prinosa. Ovi algoritmi donose zaključke obradom velike količine podataka u vidu snimaka visoke rezolucije, na osnovu kojih se donose precizne odluke.

U cilju predikcije prinosa u radu je korišćen jedan od već dokazanih algoritama nadgledanog mašinskog učenja poput: Slučajna šuma (engl. Random Forest - RF), Višestruka linearna regresija (engl. Multiple Linear Regression - MLR) sa parametrima procenjenim algoritmom stohastičkog gradijentnog spusta (engl. Stochastic Gradient Descent - SGD), Metod vektora nosača (engl. Support Vector Machine - SVM) i

XGBoost (eXtreme Gradient Boosting) Breiman 2001; T. Chen and Guestrin 2016. Da bi se išlo u korak s vremenom, u istu svrhu su isprobane i različite arhitekture dubokih neuronskih mreža (engl. Deep Neural Network - DNN) LeCun et al. 2015, između ostalog konvolucionih neuronskih mreža (engl. Convolutional Neural Network - CNN) Krizhevsky et al. 2017. Za formiranje nedostajućih podataka u bazi primenjene su posebne strukture neuronskih mreža (engl. Generative Adversarial Network - GAN) koje će omogućiti da se uklone smetnje i nedostaci u skupu podataka Goodfellow et al. 2020.

Savremene tehnike za obradu snimaka i analizu podataka daju izuzetno dobre rezultate, s velikim potencijalom za praktičnu primenu. Ključni algoritmi uključuju RF i DNN, koji su uspešno primenjeni u poljoprivredi Chowdhury 2024; Ngugi et al. 2024. Prvobitno su ove metode korišćene za klasifikaciju zemljišnog pokrivača, omogućavajući izradu preciznih mapa vodnih površina, šumskih područja, poljoprivrednih parcela i urbanih zona Bellón et al. 2017.

Metode RF i DNN takođe su uspešno korišćene za klasifikaciju poljoprivrednih kultura Zhong et al. 2019; Orynbaikyzy et al. 2019. Modeli zasnovani na DNN, u kombinaciji sa SAR snimcima, postižu prosečnu tačnost od 86%, dok Random Forest algoritam, koristeći multispektralne satelitske snimke, dostiže tačnost od 90% Patil et al. 2025. Slično tome, DNN se koristi za prepoznavanje stanja biljnih kultura i detekciju pojave bolesti na parcelama Lu et al. 2023.

Jedna od najvažnijih primena daljinske detekcije i mašinskog učenja je predikcija prinosa. Uspešnost predikcije zavisi od vrste kulture i njenog prosečnog prinosa. Performanse algoritama se obično izražavaju kao RMSE (Root Mean Square Error). Dosadašnja istraživanja pokazuju da kod predikcije prinosa kukuruza pomoću DNN prosečan RMSE iznosi 12% S. Khaki and L. Wang 2019.

Informacija o prinosima poljoprivrednih kultura nije značajna samo za poljoprivrednike već i za optimizaciju izbora semena i sorti koje najbolje odgovaraju određenom podneblju Marko et al. 2016. Ove metode mogu koristiti i semenske kuće za planiranje distribucije i potrebne količine semena na regionalnom nivou. Savremeni algoritmi mašinskog učenja analiziraju istorijske podatke i daju preporuke za sorte sa najvećim prinosima u određenim agroekološkim uslovima, čime poljoprivrednici mogu optimizovati skladišne kapacitete, logistiku i prodajne strategije.

U domenu precizne poljoprivrede, uređaji za monitoring prinosa predstavljaju moćan alat za sagledavanje zona s većim ili manjim prinosima unutar jedne parcele. Predikcija prinosa je jedan od najvažnijih zadataka mašinskog učenja u poljoprivredi Sharma et al. 2020. To omogućava analizu podataka o određenim segmentima proizvodnje, identifikaciju ključnih faktora uspeha i pripremu poboljšanih agrotehničkih mera za

sledeće sezone.

Postoji više tipova algoritama mašinskog učenja koji su primenjeni za predikciju prinosa useva. Jedan od pristupa je upotreba DNN modela treniranih na podacima o karakteristikama zemljišta i meteorološkim podacima Jabed and Murad 2024. Drugi pristup podrazumeva korišćenje CNN, kombinovanih sa javno dostupnim satelitskim snimcima Peng et al. 2024. U oblasti poljoprivrede, primena CNN modela i Vision Transformera je u ekspanziji, s obzirom na to da su podaci sa različitih senzora (satelitski, RGB, RGB+D, termalni, hiperspektralni, dron snimci) sve dostupniji Bourriz et al. 2025.

Multi-Modal Spatial-Temporal Vision Transformer (MMST-ViT) je model dubokog učenja namenjen za predviđanje prinosa useva uzimajući u obzir klimatske promene, kombinujući višemodalne prostorno-vremenske podatke kako bi se analizirali kratkoročni vremenski uticaji i dugoročne klimatske promene. Model se sastoji od Multi-Modal Transformera koji integriše satelitske snimke i meteorološke podatke, Prostornog Transformera za hvatanje visoko-rezolucijskih zavisnosti između okruga i Vremenskog Transformera za modelovanje dugoročnih vremenskih efekata. Kako bi poboljšali efikasnost učenja, autori su razvili pristup kontrastivnog učenja za višemodalno prethodno treniranje modela. Evaluacijom na više od 200 okruga u SAD, koristeći Sentinel-2 satelitske snimke, meteorološke podatke iz WRF-HRRR modela i statistiku useva iz USDA baze podataka, MMST-ViT je pokazao superiorne rezultate u odnosu na postojeće modele prema ključnim metrikama predikcije, poput RMSE i R^2 Lin et al. 2023.

U radu Najjar et al. 2025 autori se fokusiraju na predikciju prinosa soje, pšenice i uljane repice u Argentini, Urugvaju i Nemačkoj. Koristeći veliku bazu satelitskih snimaka, dodatnih modaliteta podataka i mapa prinosa, razvili su prediktivne modele zasnovane na dugoročnim-kratkoročnim memorijskim mrežama (engl. Long Short Term Memory - LSTM). Istražili su uticaj različitih vremenskih uzoraka satelitskih podataka i dodavanje relevantnih modaliteta na performanse modela. Za objašnjivost modela primenili su metode atribucije karakteristika kako bi kvantifikovali doprinos ulaznih karakteristika, identifikovali kritične faze rasta, analizirali varijabilnost prinosa na nivou polja i objasnili manje tačne predikcije. Rezultati su pokazali poboljšanje performansi dodavanjem više modaliteta ili korišćenjem svih dostupnih instanci satelitskih podataka. Objašnjivost je otkrila različite obrasce važnosti karakteristika za svaki usev i region, pri čemu su najuticajnije faze rasta zavisile od vremenskog uzorkovanja ulaznih podataka, što se poklapa sa postojećom literaturom u agronomiji i biologiji razvoja useva.

2.4 Upravljanje zonama u parceli

Upravljanje zonama na poljoprivrednim parcelama zasniva se na preciznoj poljoprivredi, koja koristi tehnologije daljinskog osmatranja, geoprostorne analize i mašinskog učenja za optimizaciju prinosa i resursa. Satelitske slike omogućavaju kontinuirano praćenje stanja useva, detekciju stresa izazvanog nedostatkom vode, štetočinama ili bolestima, kao i mapiranje prostorne heterogenosti zemljišta i vegetacije. Mašinsko učenje omogućava analizu velikih skupova podataka iz različitih izvora, uključujući multispektralne i hiperspektralne satelitske slike, kako bi se generisale precizne preporuke za upravljanje zonama unutar parcela Bourriz et al. 2025.

CNN su postale ključna tehnika u analizi satelitskih podataka, omogućavajući visoku tačnost u klasifikaciji kultura i detekciji biljnih bolesti Thakur et al. 2023. Pored toga, metode kao što su RF i SVM koriste se za mapiranje varijabilnosti u prinosima i procenu stanja zemljišta na osnovu spektralnih karakteristika Khanal et al. 2018. Integracija vremenskih serija satelitskih snimaka sa Recurrent Neural Networks (RNN) omogućava preciznije predikcije prinosa i dinamike rasta useva Saeed Khaki et al. 2020.

Dodatno, kombinacija Sentinel-2 i Landsat-8 podataka pruža visoku prostornu i vremensku rezoluciju za monitoring poljoprivrednih sistema, dok algoritmi dubokog učenja mogu automatski klasifikovati poljoprivredne zone i detektovati anomalije u rastu biljaka T. Chen and Guestrin 2016. Napredne tehnike, poput GAN, koriste se za poboljšanje kvaliteta satelitskih slika i generisanje sintetičkih podataka koji pomažu u treniranju modela P. Wang et al. 2025.

Dalji razvoj upravljanja zonama na poljoprivrednim parcelama uključuje fuziju satelitskih podataka sa IoT senzorima i meteorološkim podacima kako bi se obezbedila još preciznija analiza stanja zemljišta i optimalnih strategija navodnjavanja Fuentes-Peñailillo et al. 2024. Ovi pristupi omogućavaju donošenje odluka zasnovanih na podacima, što vodi ka smanjenju troškova, efikasnijem korišćenju resursa i održivoj poljoprivredi.

2.5 Generativni algoritmi

Oblaci predstavljaju značajan izazov u analizi optičkih satelitskih slika jer prekrivaju delove scene, što ograničava dostupnost podataka i smanjuje tačnost daljinske detekcije. U poslednjih nekoliko godina, GAN su postale dominantan pristup za uklanjanje oblaka i rekonstrukciju satelitskih slika. Modeli poput Cloud-GAN i Pix2Pix GAN koriste nadgledano i nenadgledano učenje kako bi generisali visokokvalitetne optičke slike sa

parcijalnom pokrivenošću oblacima Nikolaev et al. 2025.

Dodatno, CycleGAN omogućava rekonstrukciju slika na osnovu istorijskih podataka i višespektralnih informacija, čak i u slučajevima kada referentne slike nisu dostupne. Ova metoda koristi dvosmernu mapu između oštećenih i obnovljenih slika, čime poboljšava konzistentnost boja i teksture Ma et al. 2023. GAN sa pažnjom (Attention-GAN) dodatno poboljšava uklanjanje oblaka isticanjem relevantnih delova slike, što omogućava bolju rekonstrukciju struktura u područjima sa oblacima Wu et al. 2023.

Jedna od ključnih primena GAN-a u ovom domenu je u kombinaciji sa vremenskim serijama podataka, gde se više satelitskih snimaka koristi za interpolaciju i predikciju izgubljenih informacija. Korišćenjem Spatio-Temporal GAN-a, moguće je generisati konzistentne slike koje održavaju prostorne i vremenske karakteristike scene Sebastianelli et al. 2021.

Radarske satelitske slike (SAR) imaju prednost u odnosu na optičke slike jer mogu snimati površinu Zemlje nezavisno od vremenskih uslova i osvetljenja. Međutim, interpretacija SAR podataka je složenija zbog *speckle* šuma i razlika u vizuelnim karakteristikama u odnosu na optičke slike. GAN modeli su se pokazali kao efikasni alati za transformaciju SAR slika u optičke slike, omogućavajući bolju integraciju podataka iz različitih izvora Anuradha et al. 2024.

Jedan od prvih uspešnih modela u ovom domenu bio je Pix2Pix GAN, koji koristi uparene SAR i optičke slike za treniranje mreže koja može generisati optičke slike visoke rezolucije Wei et al. 2023. Međutim, zbog ograničene dostupnosti uparenih podataka, CycleGAN je postao dominantna tehnika jer omogućava transformaciju SAR u optičke slike bez potrebe za direktno uparenim uzorcima J. Wang et al. 2024.

Nove tehnike kao što su Conditional GANs (cGAN) i StyleGAN omogućavaju adaptivnu transformaciju SAR slika u optičke slike, uz očuvanje strukturnih karakteristika terena Huang et al. 2024. Ove metode su posebno korisne u oblastima sa gustim oblacima, gde optičke slike često nisu dostupne. Pored toga, Fusion-GAN omogućava kombinaciju SAR i optičkih podataka u zajednički prikaz koji poboljšava vizuelnu interpretaciju i preciznost klasifikacije zemljišta Liu et al. 2022.

Ovi pristupi predstavljaju ključne korake ka poboljšanju dostupnosti satelitskih podataka u oblastima sa visokim procentom oblačnosti i omogućavaju efikasnije korišćenje SAR podataka u preciznoj poljoprivredi, kartografiji i praćenju prirodnih katastrofa.

2.6 Pregled korišćenih modela mašinskog učenja

Radi doslednosti matematičke notacije, u ovom poglavlju koristi se n za ukupan broj uzoraka, p za ukupan broj ulaznih obeležja, i za indeks uzorka i j za indeks obeležja. Ulazni vektor i-tog uzorka označen je sa $\mathbf{x}_i$, njegova stvarna ciljna vrednost sa y_i, a predikcija modela sa $\hat{y}_i$. Broj klasa označen je sa c, broj stabala u ansamblu sa B, dok je broj klastera u K-means algoritmu označen sa K.

Modeli mašinskog učenja predstavljaju srž savremene analize podataka, omogućavajući računarima da prepoznaju obrasce i donose predikcije bez eksplicitnog programiranja pravila. Ovi modeli nalaze primenu u različitim oblastima, medicine, finansija i poljoprivrede.

Postoje različiti pristupi u mašinskom učenju, od kojih je jedan od najčešćih nadgledano učenje. U ovom pristupu, model se trenira na skupu podataka koji sadrži ulazne vrednosti i njihove odgovarajuće izlaze. Tokom procesa treniranja, model uči da prepoznaje veze između ulaznih podataka i željenih rezultata, što mu omogućava da donosi tačne predikcije kada se susretne sa novim podacima. Primeri algoritama koji se koriste u ovom kontekstu uključuju linearnu i logističku regresiju, stabla odlučivanja, kao i neuronske mreže.

Sa druge strane, nenadgledano učenje funkcioniše bez unapred definisanih izlaznih vrednosti. Model sam istražuje podatke i otkriva skrivene obrasce, često grupišući podatke u različite klastere ili smanjujući dimenzionalnost kako bi se izdvojile ključne karakteristike. Ovaj pristup je naročito koristan u analizi velikih skupova podataka (engl. Big Data), gde su obrasci često nepoznati ili teško definisani.

Još jedan zanimljiv pristup je učenje potkrepljivanjem (engl. reinforcement learning), koje se temelji na principu interakcije agenta sa okruženjem. Kroz sistem nagrada i kazni, model uči koje akcije donose najbolje rezultate u određenim situacijama. Ovaj pristup je ključan u razvoju autonomnih sistema, kao što su roboti i inteligentni igrači u video igrama.

Proces razvoja modela mašinskog učenja obuhvata nekoliko ključnih koraka. Prvi je prikupljanje i obrada podataka, koji uključuju čišćenje, normalizaciju i pripremu za analizu. Zatim sledi faza treniranja modela, tokom koje algoritam analizira podatke i prilagođava svoje parametre kako bi minimizovao odstupanje stvarnog izlaza od željenog izlaza modela. Nakon toga, model se evaluira pomoću različitih evaluacionih metričkih metoda, poput tačnosti, preciznosti ili F1-score, kako bi se procenilo ponašanje modela na novim podacima.

Zahvaljujući sve većoj dostupnosti podataka i napretku računarske snage, modeli mašinskog učenja postaju neizostavan deo savremenih tehnologija. Njihova primena u

medicini omogućava precizniju dijagnostiku, u finansijama pomaže u predikciji tržišnih trendova, dok u poljoprivredi omogućava optimizaciju procesa poljoprivredne proizvodnje i praćenje stanja useva. Kako se tehnologija dalje razvija, mašinsko učenje će sve više doprinositi automatizaciji i donošenju odluka u različitim domenima, otvarajući nove mogućnosti za inovacije i unapređenje kvaliteta života.

2.6.1 Stabla odlučivanja i slučajna šuma (Random Forest)

Stabla odlučivanja predstavljaju jednu od najčešće korišćenih metoda mašinskog učenja za rešavanje problema klasifikacije i regresije. Osnovna ideja ove metode jeste da se proces donošenja odluke predstavi u obliku hijerarhijske strukture, odnosno stabla. U takvoj strukturi svaki unutrašnji čvor predstavlja jedno pitanje ili uslov koji se odnosi na vrednost nekog obeležja, dok grane predstavljaju moguće odgovore na to pitanje. Krajnji čvorovi stabla nazivaju se listovi i oni sadrže konačnu odluku modela, odnosno predikciju klase u slučaju klasifikacije ili numeričku vrednost u slučaju regresije.

Tokom izgradnje stabla cilj je da se skup podataka u svakom čvoru podeli na dva ili više podskupova tako da novonastali podskupovi budu što homogeniji. Homogenost, odnosno čistoća skupa, označava stepen u kome uzorci u jednom čvoru pripadaju istoj klasi. Što je veći broj uzoraka iste klase u čvoru, to se taj čvor smatra čistijim. Za merenje čistoće najčešće se koriste Gini indeks i entropija.

Gini indeks meri verovatnoću pogrešne klasifikacije slučajno izabranog uzorka iz datog čvora ukoliko bi mu se klasa dodelila prema raspodeli klasa u tom čvoru. Manja vrednost Gini indeksa ukazuje na veću čistoću čvora. Gini indeks se može zapisati kao:

$$G = 1 - \sum_{k=1}^{c} \pi_k^2 \tag{2.1}$$

gde je π_k udeo uzoraka koji pripadaju klasi k, dok je c ukupan broj klasa.

Entropija predstavlja drugu meru nečistoće skupa i zasniva se na količini nesigurnosti u raspodeli klasa. Ukoliko svi uzorci u čvoru pripadaju istoj klasi, entropija je jednaka nuli. Ukoliko su klase ravnomerno raspoređene, entropija ima veću vrednost. Entropija se definiše kao:

$$H = -\sum_{k=1}^{c} \pi_k \log_2 \pi_k \tag{2.2}$$

Obuka jednog stabla odlučivanja odvija se rekurzivno, počevši od korenog čvora koji sadrži ceo skup za obuku. U svakom čvoru algoritam razmatra različita pitanja, odnosno moguće podele skupa podataka na osnovu vrednosti pojedinačnih obeležja. Za svaku moguću podelu izračunava se koliko ona poboljšava čistoću novonastalih podskupova.

Kao optimalna podela bira se ona koja najviše smanjuje nečistoću, odnosno ona koja najbolje razdvaja uzorke različitih klasa. Nakon toga se isti postupak ponavlja u svakom novom čvoru, sve dok se ne ispuni neki od kriterijuma zaustavljanja. Ti kriterijumi mogu biti maksimalna dubina stabla, minimalan broj uzoraka u čvoru ili postizanje dovoljno velike čistoće čvora.

Kod regresionog stabla postupak obuke zasniva se na istom principu rekurzivne podele prostora obeležja, ali se kvalitet podele ne određuje na osnovu čistoće klasa, već prema homogenosti numeričkih ciljnih vrednosti. Neka skup uzoraka koji se nalazi u čvoru m bude označen sa Q_m, pri čemu svaki uzorak sadrži vektor obeležja $\mathbf{x}_i$ i odgovarajuću ciljnu vrednost y_i. U slučaju predikcije prinosa, y_i predstavlja izmerenu vrednost prinosa za dati uzorak.

Za svako raspoloživo obeležje j i potencijalnu graničnu vrednost t razmatra se binarna podela skupa uzoraka na levi i desni podskup:

$$Q_m^L(j,t) = \Big\{(\mathbf{x}_i, y_i) \in Q_m \mid x_{ij} \leq t\Big\}, \tag{2.3}$$

$$Q_m^R(j,t) = \Big\{(\mathbf{x}_i, y_i) \in Q_m \mid x_{ij} > t\Big\}. \tag{2.4}$$

Homogenost ciljnih vrednosti u svakom novonastalom čvoru najčešće se procenjuje pomoću srednje kvadratne greške ili varijanse. Za skup uzoraka Q kriterijum nečistoće može se zapisati kao:

$$I(Q) = \frac{1}{|Q|} \sum_{(\mathbf{x}_i, y_i) \in Q} \left(y_i - \bar{y}_Q\right)^2, \tag{2.5}$$

gde je $\bar{y}_Q$ srednja vrednost ciljnih vrednosti uzoraka u posmatranom čvoru:

$$\bar{y}_Q = \frac{1}{|Q|} \sum_{(\mathbf{x}_i, y_i) \in Q} y_i. \tag{2.6}$$

Za svaku moguću podelu izračunava se ponderisana nečistoća levog i desnog podskupa:

$$J(j,t) = \frac{|Q_m^L|}{|Q_m|} I\left(Q_m^L\right) + \frac{|Q_m^R|}{|Q_m|} I\left(Q_m^R\right). \tag{2.7}$$

Kao optimalna bira se kombinacija obeležja i granične vrednosti koja minimizuje ponderisanu nečistoću novonastalih čvorova:

$$(j^*, t^*) = \arg\min_{j,t} J(j,t). \tag{2.8}$$

Ekvivalentno, optimalna podela može se posmatrati kao ona koja ostvaruje najveće smanjenje nečistoće u odnosu na roditeljski čvor:

$$\Delta I = I(Q_m) - J(j^*, t^*). \tag{2.9}$$

Nakon izbora optimalne podele, isti postupak se rekurzivno primenjuje na levom i desnom podskupu. Grananje se nastavlja sve dok se ne ispuni unapred definisan kriterijum zaustavljanja, kao što su maksimalna dubina stabla, minimalan broj uzoraka potreban za podelu čvora, minimalan broj uzoraka u listu ili nedovoljno smanjenje greške nakon nove podele. Kada se formira završni list, njegova predikciona vrednost određuje se kao srednja vrednost ciljnih vrednosti svih trening uzoraka koji su dospeli u taj list:

$$\hat{y}_m = \frac{1}{|Q_m|} \sum_{(\mathbf{x}_i, y_i) \in Q_m} y_i. \tag{2.10}$$

Na taj način obuka regresionog stabla formira skup hijerarhijskih pravila koja dele prostor obeležja na regione sa što sličnijim vrednostima ciljne promenljive. Svakom završnom regionu pridružuje se jedna numerička vrednost, koja se zatim koristi kao predikcija za sve nove uzorke koji tokom zaključivanja dospeju u odgovarajući list.

Kod problema klasifikacije, zaključivanje pomoću obučenog stabla vrši se tako što se novi uzorak propušta kroz stablo od korenog čvora ka listu. U svakom unutrašnjem čvoru proverava se odgovarajući uslov nad obeležjima uzorka i na osnovu rezultata tog uslova bira se odgovarajuća grana. Kada uzorak stigne do lista, dodeljuje mu se klasa koja je dominantna u tom listu. Kod regresionih problema postupak prolaska kroz stablo je isti, ali se u listu ne nalazi klasa, već numerička vrednost. Ta vrednost se najčešće računa kao srednja vrednost ciljnih vrednosti uzoraka koji su tokom obuke dospeli u taj list.

Iako su stabla odlučivanja jednostavna za razumevanje i interpretaciju, pojedinačna stabla često mogu biti osetljiva na promene u trening podacima. Mala promena u skupu za obuku može dovesti do formiranja drugačije strukture stabla, što može uticati na stabilnost i tačnost predikcije. Zbog toga se često koristi ansambl pristup, u kome se kombinuje više stabala kako bi se dobio robusniji model. Jedna od najpoznatijih metoda ovog tipa jeste slučajna šuma.

Slučajna šuma, odnosno RF, predstavlja ansambl metodu koja se sastoji od velikog broja stabala odlučivanja. Umesto da se predikcija zasniva na jednom stablu, slučajna šuma kombinuje predikcije više pojedinačnih stabala. Osnovna ideja je da se više relativno jednostavnih modela može kombinovati tako da zajedno daju stabilniju i tačniju predikciju. Svako stablo u šumi trenira se na drugačijem podskupu podataka, a dodatna slučajnost uvodi se i prilikom izbora obeležja koja se razmatraju u pojedinačnim čvorovima.

Skup za obuku pojedinačnog stabla formira se postupkom koji se naziva *bagging*, odnosno *bootstrap aggregating*. U ovom postupku se iz originalnog trening skupa nasumično biraju uzorci sa vraćanjem. To znači da isti uzorak može biti izabran više puta, dok neki uzorci mogu ostati neizabrani za obuku konkretnog stabla. Na taj način svako stablo dobija nešto drugačiji trening skup, što doprinosi raznovrsnosti stabala u šumi. Uzorci koji nisu izabrani za obuku određenog stabla nazivaju se *out-of-bag* uzorci i mogu se koristiti za procenu greške modela.

Pored slučajnog izbora uzoraka, slučajna šuma uvodi slučajnost i prilikom obuke svakog pojedinačnog stabla. Naime, u svakom čvoru stabla ne razmatraju se sva dostupna obeležja, već samo slučajno izabran podskup obeležja. Iz tog podskupa bira se obeležje i prag podele koji daju najbolju podelu prema odabranoj meri čistoće, kao što su Gini indeks ili entropija kod klasifikacije, odnosno smanjenje varijanse kod regresije. Ovakav pristup smanjuje međusobnu korelaciju između stabala i doprinosi boljoj generalizaciji modela.

Zaključivanje kod slučajne šume zavisi od tipa problema. Kod klasifikacije, svaki pojedinačni model, odnosno svako stablo, daje svoju predikciju klase za dati ulazni uzorak. Konačna klasa određuje se većinskim glasanjem svih stabala u ansamblu:

$$\hat{y} = \text{mode}\{h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_B(\mathbf{x})\} \tag{2.11}$$

gde $h_b(\mathbf{x})$ predstavlja predikciju b-tog stabla, dok je B ukupan broj stabala u šumi.

Kod regresionih problema, svako stablo daje numeričku predikciju, dok se konačna vrednost najčešće dobija kao prosečna vrednost predikcija svih stabala:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} h_b(\mathbf{x}) \tag{2.12}$$

Na ovaj način se smanjuje uticaj pojedinačnih grešaka stabala, jer konačna predikcija zavisi od celog ansambla, a ne od jednog modela.

Glavne prednosti slučajne šume su visoka tačnost, robusnost i sposobnost rada sa velikim brojem obeležja. Zbog kombinovanja većeg broja stabala, model je manje osetljiv na šum u podacima i manje sklon preprilagođavanju u odnosu na pojedinačno stablo odlučivanja. Slučajna šuma može se koristiti i za klasifikaciju i za regresiju, a dodatna prednost je mogućnost procene važnosti pojedinačnih obeležja, što može biti korisno za interpretaciju rezultata.

Ipak, ova metoda ima i određena ograničenja. U poređenju sa jednim stablom odlučivanja, slučajna šuma je složenija za interpretaciju, jer se konačna odluka donosi na osnovu velikog broja stabala. Takođe, obuka i predikcija mogu zahtevati više računarskih resursa, posebno kada se koristi veliki broj stabala ili veoma veliki skup podataka. Uprkos

tome, slučajna šuma se smatra jednom od najpouzdanijih i najčešće korišćenih metoda klasičnog mašinskog učenja, naročito u slučajevima kada je potrebna dobra tačnost i stabilnost predikcije Louppe 2014.

2.6.2 Višestruka linearna regresija (Multiple Linear Regression)

Višestruka linearna regresija (engl. Multiple Linear Regression – MLR) predstavlja statistički metod kojim se modelira odnos između jedne zavisne promenljive (engl. dependent variable, target) i dve ili više nezavisnih promenljivih (engl. independent variables, predictors). Ova metoda predstavlja proširenje jednostavne linearne regresije i omogućava analizu uticaja većeg broja prediktora na ciljnu promenljivu.

Za i-ti uzorak, MLR model može se zapisati kao:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i, \tag{2.13}$$

odnosno u razvijenom obliku:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \tag{2.14}$$

gde su:

- y_i – stvarna vrednost zavisne, odnosno ciljne promenljive za i-ti uzorak;
- x_{ij} – vrednost j-tog prediktora za i-ti uzorak;
- β_0 – slobodni član modela (engl. *intercept*);
- β_j – regresioni koeficijent pridružen j-tom prediktoru;
- ϵ_i – slučajna greška modela za i-ti uzorak;
- p – ukupan broj prediktora.

Za pravilno korišćenje višestruke linearne regresije razmatraju se sledeće pretpostavke:

- **Linearnost** – očekivana vrednost zavisne promenljive predstavlja linearnu funkciju prediktora;
- **Nezavisnost reziduala** – greške modela ne bi trebalo da budu međusobno zavisne;

- **Homoskedastičnost** – varijansa grešaka trebalo bi da bude približno konstantna za različite vrednosti prediktora;
- **Normalnost reziduala** – reziduali bi trebalo da imaju približno normalnu raspodelu, posebno kada se sprovodi statističko zaključivanje;
- **Odsustvo izražene multikolinearnosti** – prediktori ne bi trebalo da budu međusobno veoma snažno korelisani.

Najčešće korišćena metoda za procenu regresionih koeficijenata jeste metoda najmanjih kvadrata, kojom se minimizuje suma kvadrata razlika između stvarnih i predviđenih vrednosti:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2, \tag{2.15}$$

gde je n ukupan broj uzoraka, dok $\hat{y}_i$ predstavlja vrednost koju je model predvideo za i-ti uzorak.

Pored klasične procene koeficijenata metodom najmanjih kvadrata, u radu je korišćena i varijanta linearnog regresionog modela kod koje se parametri procenjuju iterativno primenom stohastičkog gradijentnog spusta (engl. Stochastic Gradient Descent – SGD). Za razliku od klasičnog MLR modela, kod kojeg se koeficijenti određuju direktnim rešavanjem problema najmanjih kvadrata, SGD postepeno ažurira koeficijente u smeru smanjenja funkcije gubitka. U implementaciji korišćenoj u ovom radu primenjena je Huber funkcija gubitka uz L_2 regularizaciju, a ova varijanta modela u rezultatima je označena kao MLR_SGD.

Regularizovana proširenja linearne regresije uključuju:

- **Ridge regresiju**, koja funkciji greške dodaje L_2 penalizaciju $\lambda \sum_{j=1}^{p} \beta_j^2$, čime se ograničavaju velike vrednosti koeficijenata i ublažavaju posledice multikolinearnosti između prediktora;
- **Lasso regresiju**, koja dodaje L_1 penalizaciju $\lambda \sum_{j=1}^{p} |\beta_j|$ i može dovesti do toga da pojedini regresioni koeficijenti postanu jednaki nuli, čime se vrši implicitna selekcija prediktora.

Jedna od metrika koja se može koristiti za ocenjivanje slaganja predikcija MLR modela sa stvarnim vrednostima jeste koeficijent determinacije:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}, \tag{2.16}$$

gde y_i predstavlja stvarnu vrednost ciljne promenljive za i-ti uzorak, $\hat{y}_i$ predviđenu vrednost modela, dok $\bar{y}$ označava srednju vrednost svih stvarnih vrednosti:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{2.17}$$

Brojilac u izrazu za R^2 predstavlja sumu kvadrata grešaka predikcije, dok imenilac predstavlja ukupnu varijabilnost stvarnih vrednosti u odnosu na njihovu srednju vrednost. Veće vrednosti R^2, odnosno vrednosti bliže jedinici, ukazuju na bolje slaganje predikcija sa stvarnim vrednostima. Međutim, dodavanje novih prediktora može povećati R^2 čak i kada ti prediktori ne doprinose značajno modelu.

Zbog toga se može koristiti prilagođeni koeficijent determinacije, koji uzima u obzir i broj prediktora:

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}, \tag{2.18}$$

gde je n broj uzoraka, a p broj prediktora. Za razliku od običnog koeficijenta determinacije, prilagođeni koeficijent može se smanjiti kada se modelu dodaju prediktori koji ne doprinose dovoljno poboljšanju predikcije. Pored toga, važne metrike uključuju:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.20}$$

Ako nezavisne promenljive imaju visoku međusobnu korelaciju, to može dovesti do nestabilnih procena koeficijenata. Može se detektovati pomoću *Variance Inflation Factor (VIF)*:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{2.21}$$

Ako je $VIF_j > 10$, potrebno je razmotriti eliminaciju ili transformaciju odgovarajućeg prediktora. Ovde R_j^2 predstavlja koeficijent determinacije dobijen regresijom j-tog prediktora na preostale prediktore. Ako postoji zavisnost između reziduala, može se koristiti *Durbin-Watson test*. U slučaju heteroskedastičnosti, može se koristiti *Breusch-Pagan test*, a rešenje uključuje logaritamsku i/ili Box-Cox transformaciju podataka.

MLR se široko koristi u različitim oblastima kao što su predviđanje cena i/ili potrošačkih trendova, analiza faktora koji utiču na dobit kompanije, analiza efekata različitih marketing kampanja, predviđanje temperature na osnovu klimatskih faktora, istraživanje uticaja više faktora na zdravlje pacijenata...

2.6.3 Metoda potpornih vektora (Support Vector Machine)

Metoda potpornih vektora (engl. Support Vector Machine – SVM) je nadgledani algoritam mašinskog učenja koji se koristi za probleme klasifikacije i regresije, pri čemu je najpoznatiji po svojoj primeni u binarnoj klasifikaciji. Osnovna ideja SVM-a je pronalaženje optimalne hiperravni (engl. hyperplane) koja najbolje razdvaja klase u višedimenzionalnom prostoru.

Matematički, hiperravan se može zapisati kao:

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0 \tag{2.22}$$

gde su:

- $\mathbf{w}$ – vektor težina,
- $\mathbf{x}$ – ulazni vektor obeležja,
- b – slobodan član (bias).

Cilj SVM-a je da pronađe hiperravan koja maksimizuje marginu, tj. razdaljinu između najbližih tačaka iz svake klase (engl. support vectors) i same hiperravni. Margina se definiše kao:

$$\text{margin} = \frac{2}{\|\mathbf{w}\|} \tag{2.23}$$

i treba da bude maksimalna da bi model bio što stabilniji i generalizabilan.

Da bismo rešili ovaj optimizacioni problem, koristimo sledeću formulaciju:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{uz ograničenje} \quad & y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1, \qquad i = 1, \ldots, n. \end{aligned} \tag{2.24}$$

gde je $y_i \in \{-1, 1\}$ oznaka klase.

Ukoliko podaci nisu linearno separabilni, koristi se meka margina (engl. soft margin) dodavanjem relaksacionih varijabli ξ_i, što daje novi optimizacioni problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{2.25}$$

uz ograničenje:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \tag{2.26}$$

gde je C hiperparametar koji balansira marginu i penalizaciju pogrešnih klasifikacija.

U slučajevima kada podaci nisu linearno separabilni ni sa mekanom marginom, koristi se trik sa kernelom koji omogućava preslikavanje podataka u viši dimenzionalni prostor pomoću funkcije:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \tag{2.27}$$

Najčešće korišćeni kernel funkcije su:

- **Linearni kernel**: $K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$
- **Polinomialni kernel**: $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + c_0)^q$
- **Radial Basis Function (RBF)**: $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$
- **Sigmoidni kernel**: $K(\mathbf{x}, \mathbf{x}') = \tanh(\alpha \mathbf{x}^T \mathbf{x}' + c_0)$

gde q označava stepen polinomskog kernela, a c_0 konstantni član kernela.

SVM se može proširiti i na probleme višeklasne klasifikacije (engl. multi-class classification) korišćenjem strategija poput jedan na jedan (engl. one-vs-one OvO) i jedan protiv svih (engl. one-vs-all OvA).

Za ocenu performansi SVM modela koriste se metrike poput tačnosti (engl. accuracy), preciznosti (engl.- precision), specifičnost (engl. recall) i F1-mere. Model se trenira optimizacijom pomoću algoritama poput *Sequential Minimal Optimization (SMO)*.

SVM je veoma moćan algoritam koji daje dobre rezultate u visoko-dimenzionalnim podacima i kada su klase dobro separabilne. Međutim, može biti računski zahtevan za velike skupove podataka i osetljiv na izbor hiperparametara.

2.6.4 eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGBoost) je optimizovana verzija *Gradient Boosting* algoritma, razvijena da bude brža i efikasnija u skalabilnim aplikacijama mašinskog učenja. XGBoost je posebno dizajniran da bude računski optimizovan, koristeći paralelnu obradu podataka, optimizacije memorije i poboljšane heuristike za redukciju složenosti modela.

Osnovna ideja XGBoost-a je gradnja ansambla oslabljenih učenika (slabih modela) kroz iterativno poboljšavanje predikcija koristeći gradijentni spust. Model se zasniva na drveću odlučivanja, gde svako novo stablo pokušava da koriguje greške prethodnog skupa stabala.

Matematički, cilj XGBoost-a je minimizacija sledeće funkcije gubitka:

$$L(\theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{b=1}^{B} \Omega(f_b) \tag{2.28}$$

gde je:

- $\ell(y_i, \hat{y}_i)$ - funkcija greške između stvarnih vrednosti y_i i predviđenih vrednosti $\hat{y}_i$,
- f_b - b-to stablo u ansamblu, pri čemu je B ukupan broj stabala,
- $\Omega(f_b)$ - regularizaciona funkcija kojom se penalizuju složenija stabla kako bi se sprečilo prekomerno prilagođavanje (engl. overfitting).

Regularizacija se implementira pomoću sledeće formule:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{\ell=1}^{T} w_\ell^2 \tag{2.29}$$

gde su:

- T - broj listova u stablu,
- w_ℓ - težina ℓ-tog lista stabla,
- γ i λ - hiperparametri koji regulišu složenost modela.

Glavna karakteristika XGBoost-a je njegov mehanizam **drveća odlučivanja sa rastom u širinu** (engl. depth-wise growth), gde se čvorovi stabla dodaju prema strategiji smanjenja gubitka. Ovo se razlikuje od tradicionalnog *boosting*-a, gde su stabla plića, ali pažljivo prilagođena.

XGBoost koristi **drugu rednu aproksimaciju** za optimizaciju funkcije greške, zasnovanu na Taylor-ovoj ekspanziji:

$$L^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t) \tag{2.30}$$

gde su g_i i h_i prvi i drugi izvod funkcije greške:

$$g_i = \frac{\partial \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 \ell(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \tag{2.31}$$

XGBoost nudi različite metode optimizacije, uključujući **paralelnu implementaciju** za brže treniranje, **histogramsku binarnu optimizaciju** radi ubrzanja računanja i **pruning strategiju** koja odbacuje grane koje ne doprinose smanjenju greške.

U pogledu hiperparametara, najvažniji su:

- **learning rate** (η) - određuje veličinu koraka prilikom učenja modela,
- **max depth** - maksimalna dubina stabala,
- **subsample** - uzorak podataka koji se koristi za treniranje svakog stabla,
- **colsample_bytree** - procenat kolona korišćenih pri gradnji svakog stabla,
- **lambda** i **alpha** - regularizacioni parametri L2 i L1 penalizacije.

XGBoost podržava različite funkcije cilja, uključujući *log-loss* za binarnu klasifikaciju, *mean squared error* za regresione zadatke i *multi-class classification loss* za probleme sa više klasa.

Zahvaljujući svojoj optimizaciji, XGBoost je postao dominantan algoritam u mnogim takmičenjima, kao što su Kaggle izazovi. Koristi se u oblastima kao što su predviđanje finansijskih trendova, detekcija prevara, analize medicinskih podataka i preporučivački sistemi.

XGBoost nudi implementaciju u različitim programskim jezicima, uključujući Python (`xgboost` biblioteka), R, Julia i C++. Njegova fleksibilnost i visoka efikasnost čine ga jednim od najmoćnijih alata za analizu velikih skupova podataka.

2.6.5 Duboke neuronske mreže (Deep Neural Network)

Duboke neuronske mreže (engl. Deep Neural Networks – DNN) su složeni neuronski modeli koji se sastoje od više slojeva neurona organizovanih u hijerarhiju. Oni predstavljaju osnovu dubokog učenja (engl. deep learning) i koriste se za obradu nelinearnih podataka u oblastima kao što su računarska vizija, obrada prirodnog jezika i analiza vremenskih serija.

Neuronske mreže se sastoje od tri osnovne vrste slojeva:

- **Ulazni sloj (engl. input layer)** – Prvi sloj koji prima podatke.
- **Skriveni slojevi (engl. hidden layers)** – Jedan ili više slojeva koji primenjuju nelinearne transformacije podataka.
- **Izlazni sloj (engl. output layer)** – Poslednji sloj koji generiše predikcije.

Matematički, svaki neuron u mreži izračunava linearnu kombinaciju ulaznih vrednosti, kojoj se zatim primenjuje nelinearna aktivaciona funkcija:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)} \tag{2.32}$$

$$a^{(l)} = \sigma(z^{(l)}) \tag{2.33}$$

gde su:

- $z^{(l)}$ – linearna transformacija na sloju l,
- $W^{(l)}$ – matrica težina sloja l,
- $b^{(l)}$ – vektor pristrasnosti (engl. bias) sloja l,
- $a^{(l)}$ – izlaz aktivacione funkcije sloja l,
- σ – aktivaciona funkcija.

Parametri neuronske mreže mogu se objedinjeno označiti sa θ, pri čemu θ obuhvata sve matrice težina $W^{(l)}$ i sve vektore pristrasnosti $b^{(l)}$ za slojeve mreže koji sadrže parametre koji se uče. Tokom obuke vrednosti ovih parametara prilagođavaju se tako da se minimizuje odabrana funkcija gubitka.

U DNN modelima koriste se različite aktivacione funkcije kako bi se omogućilo modelovanje nelinearnih odnosa:

- **Sigmoid**: $\sigma(x) = \frac{1}{1+e^{-x}}$, čiji se izlaz nalazi u intervalu $(0, 1)$ i koja se često koristi u izlaznom sloju kod binarne klasifikacije.
- **ReLU (Rectified Linear Unit)**: $f(x) = \max(0, x)$, koja uvodi nelinearnost i široko se koristi u skrivenim slojevima dubokih mreža.
- **Tanh**: $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, čiji se izlaz nalazi u intervalu $(-1, 1)$.
- **Softmax**: koristi se najčešće u izlaznom sloju kod višeklasne klasifikacije i transformiše izlazne vrednosti mreže u raspodelu verovatnoća po klasama.

Trening DNN modela vrši se minimizacijom funkcije gubitka $J(\theta)$ korišćenjem optimizacionih algoritama kao što je *Stochastic Gradient Descent* (SGD):

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla J(\theta) \tag{2.34}$$

gde je η stopa učenja (*learning rate*). Poboljšani optimizatori kao što su Adam, RMSprop i Adagrad prilagođavaju stope učenja na osnovu prethodnih iteracija.

Ključna tehnika u obuci dubokih mreža je propagacija u nazad, algoritam koji računa gradijent funkcije gubitka propagirajući grešku od izlaznog ka ulaznom sloju. Ažuriranje težina u sloju l definisano je kao:

$$\Delta W^{(l)} = -\eta \frac{\partial J}{\partial W^{(l)}} \tag{2.35}$$

DNN modeli mogu biti podložni različitim problemima tokom obuke, među kojima su prekomerno prilagođavanje (engl. overfitting), kao i nestajanje i eksplozija gradijenata. Za ublažavanje ovih problema razvijen je veliki broj tehnika, pri čemu se razlikuju metode namenjene poboljšanju generalizacije modela i metode namenjene stabilizaciji procesa optimizacije (Goodfellow et al. 2014).

Za smanjenje prekomernog prilagođavanja i poboljšanje sposobnosti generalizacije najčešće se koriste:

- L_1 **i** L_2 **regularizacija** – funkciji gubitka dodaje se penalizacioni član kojim se ograničava veličina parametara modela. Na primer, kod L_2 regularizacije važi:

$$\widetilde{J}(\theta) = J(\theta) + \lambda \|\theta\|_2^2, \tag{2.36}$$

  gde je λ hiperparametar koji određuje intenzitet regularizacije.

- **Dropout** – tokom obuke se nasumično deaktivira deo neurona, odnosno njihovih izlaza, čime se smanjuje zavisnost mreže od pojedinačnih aktivacija i podstiče robusnija reprezentacija.

- **Rano zaustavljanje (engl. early stopping)** – obuka se prekida kada se performanse na validacionom skupu više ne poboljšavaju, čime se ograničava dalje prilagođavanje modela skupu podataka za obuku.

- **Augmentacija podataka (engl. data augmentation)** – skup podataka za obuku proširuje se dodatnim, odgovarajuće transformisanim uzorcima, čime se povećava raznovrsnost podataka dostupnih modelu.

- **Dodavanje šuma** – kontrolisano dodavanje šuma ulaznim podacima, aktivacijama ili parametrima može povećati robusnost modela i smanjiti njegovu osetljivost na male promene ulaza.

- **Ansambl metode** – kombinovanjem predikcija više modela može se smanjiti varijansa pojedinačnog modela i poboljšati generalizacija.

- **Deljenje parametara (engl. parameter sharing)** – isti parametri koriste se u više delova modela, čime se ograničava broj nezavisnih parametara i kapacitet modela.

Sa druge strane, problemi nestajanja i eksplozije gradijenata odnose se prvenstveno na stabilnost optimizacije dubokih mreža. Za njihovo ublažavanje koriste se odgovarajuća inicijalizacija težina, izbor pogodnih aktivacionih funkcija, normalizacija aktivacija i, kod eksplozije gradijenata, ograničavanje njihove magnitude (engl. *gradient clipping*).

- **Odgovarajuća inicijalizacija težina** – početne vrednosti težina biraju se tako da se varijansa aktivacija i gradijenata što bolje očuva kroz slojeve mreže.
- **Batch normalizacija (engl. Batch Normalization)** – aktivacije unutar mreže normalizuju se korišćenjem statistika iz mini-paketa, što može doprineti stabilnijoj i efikasnijoj obuci.
- **Izbor aktivacionih funkcija** – korišćenje funkcija kao što je ReLU može ublažiti problem nestajanja gradijenta u poređenju sa zasićujućim aktivacionim funkcijama u određenim delovima mreže.
- **Ograničavanje gradijenta (engl. gradient clipping)** – kada gradijent postane veoma velik, njegova norma se ograničava kako bi se sprečila nestabilna ažuriranja parametara.

DNN se koristi u širokom spektru primena, uključujući prepoznavanje slika, analizu govora, predikciju vremenskih serija i biomedicinsku dijagnostiku. Zahvaljujući modernim GPU i TPU akceleratorima, treniranje ovih modela postalo je efikasnije, omogućavajući razvoj sofisticiranih dubokih arhitektura poput konvolucionih neuronskih mreža (CNN) i rekurentnih neuronskih mreža (RNN).

2.6.6 Long Short-Term Memory

Za modelovanje sekvencijalnih i vremenski zavisnih podataka u ovom istraživanju razmatrana je arhitektura *Long Short-Term Memory* (LSTM), koja predstavlja specijalizovanu RNN. Osnovna prednost LSTM pristupa ogleda se u sposobnosti da efikasno modeluje dugoročne zavisnosti u vremenskim serijama, odnosno da zadrži korisnu informaciju kroz veći broj vremenskih koraka, što kod standardnih RNN arhitektura često nije moguće zbog problema nestajanja ili eksplozije gradijenta.

U kontekstu analize vremenskih serija daljinske detekcije, LSTM je posebno pogodan jer ulazni podaci prirodno imaju sekvencijalnu strukturu. Na primer, za svaki piksel, parcelu ili uzorak moguće je formirati niz opservacija kroz vreme, pri čemu svaka opservacija sadrži vrednosti spektralnih kanala, vegetacionih indeksa ili drugih izvedenih atributa za određeni datum snimanja. Na taj način LSTM ne posmatra svaku opservaciju

izolovano, već uči kako se stanje vegetacije menja kroz sezonu, odnosno kako prethodne fenološke faze utiču na kasnije stanje useva i konačni ishod, kao što je klasna pripadnost ili prinos.

Za razliku od klasičnih neuronskih mreža koje obrađuju ulaz kao statički vektor, LSTM obrađuje ulaznu sekvencu korak po korak. U svakom vremenskom koraku mreža prima ulazni vektor x_t, koji može predstavljati skup spektralnih i izvedenih atributa u trenutku t, kao i informacije iz prethodnog vremenskog koraka. Te informacije sadržane su u skrivenom stanju h_{t-1}, koje predstavlja neposredno dostupnu, kratkoročniju reprezentaciju, i stanju ćelije c_{t-1}, koje omogućava prenos informacija kroz duži niz vremenskih koraka i može se posmatrati kao dugoročnija memorija mreže. Ključna osobina LSTM arhitekture jeste postojanje memorijske ćelije i mehanizma kapija (engl. gates) kojima se kontroliše protok, zadržavanje i ažuriranje informacija kroz mrežu.

Kapija zaboravljanja f_t određuje u kojoj meri se pojedinačne komponente prethodnog stanja ćelije c_{t-1} prenose u novi vremenski korak. Njene vrednosti nalaze se u intervalu od 0 do 1 i element-po-element se množe sa c_{t-1}. Kada je odgovarajuća vrednost f_t bliska jedinici, informacija iz prethodnog stanja ćelije gotovo u potpunosti se zadržava, dok vrednost bliska nuli dovodi do njenog gotovo potpunog potiskivanja. Vrednosti između 0 i 1 omogućavaju delimično zadržavanje prethodne informacije.

LSTM ćelija sastoji se od tri glavne kapije: ulazne kapije (engl. input gate), kapije zaboravljanja (engl. forget gate) i izlazne kapije (engl. output gate). Njihova uloga je da adaptivno odlučuju koje informacije iz trenutnog ulaza treba zapamtiti, koje ranije informacije treba zadržati ili odbaciti i koji deo internog memorijskog stanja treba proslediti dalje kao izlaz. Na taj način mreža može da očuva relevantne informacije kroz duži vremenski period i da potisne one koje nisu značajne za zadati problem.

Matematički, rad LSTM ćelije u vremenskom koraku t može se opisati sledećim jednačinama:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2.37}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2.38}$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{2.39}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{2.40}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{2.41}$$

$$h_t = o_t \odot \tanh(c_t) \tag{2.42}$$

gde su: x_t ulazni vektor u vremenskom koraku t, h_{t-1} skriveno stanje iz prethodnog koraka, c_{t-1} memorijsko stanje iz prethodnog koraka, f_t kapija zaboravljanja, i_t ulazna kapija, $\tilde{c}_t$ kandidat za novo memorijsko stanje, c_t ažurirano memorijsko stanje, o_t izlazna kapija, h_t novo skriveno stanje, dok su W i b težine i bias parametri koji se uče tokom procesa treniranja. Funkcija σ označava sigmoidnu aktivaciju, a $\odot$ element-po-element množenje.

Kapija zaboravljanja f_t određuje koliki deo prethodnog memorijskog stanja treba sačuvati. Ukoliko je vrednost ove kapije bliska jedinici, prethodna memorija se uglavnom zadržava; ukoliko je bliska nuli, ranija informacija se potiskuje. Ulazna kapija i_t kontroliše koliko nove informacije iz trenutnog ulaza ulazi u memorijsko stanje, dok kandidat $\tilde{c}_t$ predstavlja novu sadržinu koja može biti upisana u memoriju. Kombinacijom prethodne memorije i nove informacije formira se ažurirano memorijsko stanje c_t. Konačno, izlazna kapija o_t određuje koji deo memorijskog stanja postaje skriveno stanje h_t, koje se koristi kao izlaz trenutnog koraka i ulaz za sledeći vremenski korak.

Prednost ovakve arhitekture ogleda se u tome što LSTM može naučiti da prepozna značajne obrasce u vremenskim serijama, kao što su početak intenzivnog razvoja vegetacije, periodi stagnacije, maksimum vegetacionog signala ili nagli pad usled stresa, bolesti ili sazrevanja. U primenama u poljoprivredi i daljinskoj detekciji ovo je naročito važno, jer različite kulture imaju karakteristične fenološke putanje kroz sezonu, a prinos često zavisi od celokupne dinamike razvoja, a ne od jedne izolovane opservacije.

U praktičnoj primeni LSTM modela ulazni podaci se najčešće organizuju u obliku trodimenzionalnog tenzora dimenzija $n \times \tau \times p$, gde je n broj uzoraka, τ broj vremenskih koraka, a p broj atributa po vremenskom koraku.

gde svaki uzorak predstavlja jednu parcelu, piksel ili lokaciju uzorkovanja, svaki vremenski korak jedan datum snimanja, a svaki atribut po koraku može biti spektralni kanal, vegetacioni indeks ili druga pomoćna promenljiva. Na primer, ukoliko se za svaki datum koristi 12 Sentinel-2 kanala i 10 vegetacionih indeksa, tada jedan vremenski korak sadrži 22 atributa, dok celokupna ulazna sekvenca opisuje sezonski razvoj useva.

U zavisnosti od zadatka, LSTM može biti korišćen za klasifikaciju ili regresiju. U slučaju višeklasne klasifikacije, izlazni sloj može sadržati po jedan neuron za svaku klasu. Ako je cilj samo određivanje klase, moguće je koristiti linearne izlaze i odabrati klasu sa najvećom izlaznom vrednošću. Kada je potrebno da se izlazi interpretiraju kao verovatnoće pripadnosti pojedinačnim klasama, na izlaz se primenjuje *softmax* funkcija.

Osnovna razlika između klasifikacionog i regresionog zadatka ogleda se i u izboru ciljne

funkcije. Kod klasifikacije se najčešće koristi unakrsna entropija (engl. cross-entropy), dok se kod regresije, kao što je predikcija prinosa, često koristi srednja kvadratna greška (engl. mean squared error – MSE). U regresionom slučaju završni sloj najčešće sadrži jedan linearni neuron koji daje kontinuiranu izlaznu vrednost.

U oba slučaja, skrivena reprezentacija koju LSTM formira tokom prolaska kroz celu sekvencu predstavlja sažetu informaciju o vremenskoj dinamici posmatranog uzorka, dok se način formiranja konačnog izlaza i funkcija gubitka prilagođavaju konkretnom zadatku.

Pored osnovne LSTM arhitekture, često se koriste i njene proširene varijante. Jedna od njih je *stacked LSTM*, kod koje se više LSTM slojeva postavlja jedan iznad drugog radi modelovanja složenijih vremenskih obrazaca. Druga česta varijanta je *bidirectional LSTM*, gde se sekvenca obrađuje istovremeno unapred i unazad, čime model koristi informaciju i iz prošlih i iz budućih vremenskih koraka. Međutim, u problemima gde je cilj simulacija kauzalnog predviđanja tokom sezone, jednosmerni LSTM je često prikladniji, jer ne koristi buduće informacije koje u trenutku predikcije nisu dostupne.

Iako LSTM ima značajne prednosti u modelovanju vremenskih serija, njegova primena nosi i određena ograničenja. Pre svega, obuka ovakvih mreža je računski zahtevnija u odnosu na klasične metode mašinskog učenja, kao što su RF ili XGBoost. Takođe, uspešnost modela zavisi od kvaliteta i gustine vremenske serije: ukoliko postoji veliki broj nedostajućih datuma ili neregularna raspodela opservacija, performanse modela mogu biti smanjene.

Dodatno, LSTM modeli zahtevaju pažljivo podešavanje hiperparametara, kao što su broj skrivenih jedinica, broj slojeva, stopa učenja i veličina mini-paketa (engl. batch size), koja predstavlja broj uzoraka obrađenih pre jednog ažuriranja parametara modela. Važan parametar je i dužina ulazne sekvence, odnosno broj vremenskih koraka koji se prosleđuju modelu. Teorijski, dužina sekvence može biti proizvoljno velika, ali je u praksi ograničena dostupnošću podataka, računskim resursima i sposobnošću modela da očuva relevantne informacije kroz veliki broj vremenskih koraka. Veća dužina sekvence omogućava modelu da koristi informacije iz udaljenije prošlosti, ali istovremeno povećava računsku složenost obuke.

Uprkos tome, LSTM predstavlja veoma snažan alat za analizu vremenskih serija u poljoprivredi, naročito kada je potrebno modelovati sezonsku dinamiku useva. Njegova sposobnost da integriše informacije kroz vreme čini ga pogodnim za zadatke kao što su klasifikacija useva, procena stanja vegetacije, detekcija anomalija i predikcija prinosa. U poređenju sa modelima koji koriste statički skup atributa, LSTM omogućava direktno učenje iz fenološke sekvence, čime se dobija interpretativno i metodološki opravdan okvir za rad sa višedatumskim satelitskim podacima.

2.6.7 Konvolucione neuronske mreže (Convolutional Neural Network)

Konvolucione neuronske mreže (engl. Convolutional Neural Networks – CNN) su specijalizovane duboke neuronske mreže dizajnirane za obradu podataka u obliku slika, videa i drugih strukturalnih podataka. CNN modeli su postali ključni u oblastima računarskog vida, prepoznavanja objekata, segmentacije slika i analize medicinskih podataka.

CNN se sastoji od nekoliko ključnih slojeva:

- **Konvolucioni sloj** (engl. Convolutional Layer) – Ekstrahuje relevantne karakteristike slike pomoću filtera (kernela).
- **Sloj aktivacije** (engl. Activation Layer) – Primena nelinearne funkcije (npr. ReLU) kako bi mreža mogla da nauči složene odnose u podacima.
- **Sloj sažimanja** (engl. Pooling Layer) – Smanjuje dimenzionalnost podataka i poboljšava robusnost modela.

Osnovna operacija CNN-a je konvolucija, koja se definiše kao:

$$S_k(r,s) = \sum_{c=1}^{C} \sum_{u} \sum_{v} I_c(r+u, s+v) K_{k,c}(u,v) + b_k, \tag{2.43}$$

gde su:

- I_c – c-ti kanal ulaznih podataka,
- $K_{k,c}$ – kernel k-tog filtera koji se primenjuje na c-ti ulazni kanal,
- C – broj ulaznih kanala,
- b_k – član pristrasnosti pridružen k-tom filteru,
- $S_k(r,s)$ – izlaz k-tog filtera na prostornoj poziciji (r,s).

Svaki filter unutar konvolucionog sloja otkriva specifične karakteristike slike, kao što su ivice, teksture i oblici. Mreža uči optimalne vrednosti filtera kroz proces treniranja.

Nakon konvolucionog sloja, često se koristi aktivaciona funkcija kao što je ReLU:

$$f(x) = \max(0, x) \tag{2.44}$$

koja poboljšava sposobnost mreže da uči nelinearne odnose.

Pooling sloj smanjuje dimenzionalnost slike i poboljšava invarijantnost na translacije. Najčešće korišćeni pooling operatori su:

- **Sažimanje na osnovu lokalnog maksimuma (*Max pooling*)**:

$$P(r,s) = \max_{(u,v)\in\mathcal{R}} S(r+u, s+v) \tag{2.45}$$

- **Sažimanje lokalnim usrednjavanjem (*Average pooling*)**:

$$P(r,s) = \frac{1}{|\mathcal{R}|} \sum_{(u,v)\in\mathcal{R}} S(r+u, s+v) \tag{2.46}$$

Gubitak modela se minimizuje pomoću funkcije unakrsne entropije u klasifikacionim zadacima:

$$L = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{2.47}$$

gde su y_i stvarne klase, a $\hat{y}_i$ predviđene verovatnoće.

Trening CNN-a se vrši pomoću propagacije u nazad i optimizacijskih algoritama kao što su Stochastic Gradient Descent (SGD) i Adam. Parametri se ažuriraju prema sledećoj formuli:

$$W^{(t+1)} = W^{(t)} - \eta \nabla L(W^{(t)}) \tag{2.48}$$

gde je η stopa učenja.

Da bi se sprečilo prekomerno prilagođavanje, koriste se tehnike poput:

- **Dropout** – Nasumično isključivanje neurona tokom treniranja.
- **Batch Normalization** – Normalizacija podataka u svakom sloju.
- **Data Augmentation** – Veštačko povećanje veličine skupa podataka rotacijom, translacijom i promenom osvetljenja slika.

CNN modeli su osnova za napredne arhitekture poput AlexNet, VGG, ResNet i EfficientNet, koje su revolucionalizovale oblast računarskog vida. CNN se koristi u medicinskoj dijagnostici, autonomnim vozilima, analizi satelitskih snimaka i mnogim drugim oblastima.

2.6.8 K-means

Jedna od najčešće korišćenih nenadgledanih metoda za grupisanje podataka u oblasti daljinske detekcije i precizne poljoprivrede jeste algoritam *K-means* klasterizacije MacQueen 1967; Jain 2010. U okviru ove disertacije, K-means se koristi za formiranje

približno homogenih zona unutar poljoprivredne parcele, na osnovu spektralnih, vegetacionih i/ili drugih izvedenih obeležja računatih za svaki piksel. Na taj način se omogućava prostorna podela parcele na regione koji pokazuju slične biološke i agroekološke karakteristike, što je od posebnog značaja za analizu prostorne varijabilnosti useva i podršku upravljanju na nivou zone.

Osnovna ideja K-means algoritma jeste da se skup od n uzoraka podeli na K klastera tako da se minimizuje unutarklasterska varijansa, odnosno suma kvadrata rastojanja između svakog uzorka i centroida klastera kojem je dodeljen. Neka je $\mathbf{x}_i \in \mathbb{R}^p$ vektor obeležja za i-ti uzorak, gde p predstavlja broj ulaznih obeležja, a $\boldsymbol{\mu}_k$ centroid k-tog klastera. Tada se kriterijumska funkcija može zapisati kao:

$$J = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 , \tag{2.49}$$

gde C_k označava skup uzoraka dodeljenih k-tom klasteru. Oznaka klastera predstavlja samo identifikator grupe koju je algoritam formirao na osnovu sličnosti uzoraka i ne treba je poistovećivati sa oznakom klase. Za razliku od klase, kojoj se unapred može pridružiti određeno semantičko značenje, klaster nema unapred definisano značenje, već se ono eventualno interpretira nakon izvršene klasterizacije.

Algoritam se izvodi iterativno kroz dva osnovna koraka: (i) dodeljivanje svakog uzorka klasteru čiji mu je centroid najbliži i (ii) ponovno izračunavanje centroida svakog klastera kao srednje vrednosti uzoraka koji su mu dodeljeni. Ovaj postupak se ponavlja do konvergencije, odnosno dok više ne dolazi do promene dodele uzoraka ili dok promena položaja centroida ne postane manja od unapred definisanog praga Lloyd 1982.

U kontekstu zoniranja parcela, K-means se primenjuje nad višedimenzionalnim prostorom obeležja formiranim od satelitskih kanala, vegetacionih indeksa, parametara golog zemljišta ili kombinovanih agroekoloških slojeva. Na izlazu algoritma dobija se tematska karta klastera, pri čemu svaki klaster može biti interpretiran kao zona sa sličnim stanjem vegetacije, produktivnosti ili potencijalom za upravljanje. Ovakve zone mogu biti dalje korišćene za definisanje strategija diferenciranog uzorkovanja zemljišta, varijabilne primene đubriva, navodnjavanja ili drugih mera precizne poljoprivrede Basso et al. 2016.

Prednost K-means algoritma ogleda se u njegovoj jednostavnosti, računskoj efikasnosti i lakoj interpretaciji rezultata. Metoda je pogodna za rad nad velikim skupovima piksela/vektora i omogućava brzo formiranje zona čak i na scenama visoke prostorne rezolucije. Međutim, algoritam ima i određena ograničenja. Pre svega, broj klastera K mora biti unapred definisan, što u praksi često zahteva dodatnu heuristiku ili ekspertsko znanje. Takođe, rezultat zavisi od inicijalnog izbora centroida, a standardna formulacija

algoritma ne uzima eksplicitno u obzir prostorne odnose između susednih piksela. Zbog toga se može desiti da dobijene zone budu prostorno fragmentisane ili da sadrže male izolovane regione koji nemaju jasnu agronomsku interpretaciju.

Uprkos navedenim ograničenjima, K-means predstavlja veoma pogodnu početnu metodu za zoniranje parcela, posebno kada je cilj identifikacija osnovnih obrazaca prostorne varijabilnosti u satelitskim podacima. U ovoj disertaciji K-means služi kao referentni i interpretabilni pristup za formiranje upravljačkih zona, koji se dalje može upoređivati sa naprednijim metodama zasnovanim na grafovima i prostornoj regularizaciji.

Poglavlje 3

Podaci i područje istraživanja

Sadržaj

3.1 Područje istraživanja i organizacija podataka 34
3.2 Satelitski snimci 35
3.2.1 Radarski satelitski snimci 35
3.2.2 Optički satelitski snimci 38
3.3 Monitor prinosa 41
3.4 Podaci zemljišnih karakteristika 43
3.5 Digitalni elevacioni model 46
3.6 Meteo podaci 46
3.7 Podaci o vrstama poljoprivrednih kultura 48
3.8 Podaci o datumima štete 50
3.9 Podaci o navodnjavanim parcelama 52
3.10 Podaci o proteinskom sastavu 55
3.11 Zbirni pregled i uloga skupova podataka 55

3.1 Područje istraživanja i organizacija podataka

Istraživanje je usmereno na poljoprivredne površine AP Vojvodine i odabrane parcele za koje su bili dostupni satelitski i referentni terenski podaci. Budući da nisu za sve parcele dostupni svi potrebni podaci u svim vremenskim trenucima, za svaki eksperiment odabran je podskup parcela i vremenskih trenutaka koji obezbeđuje potpunost svih neophodnih podataka. Ovo za posledicu ima da u rezultatima i disikusiji broj piksela, merenja i parcela zavisi od konkretnog problema kojim se bavi eksperiment. U interpretaciji rezultata zato se razlikuju broj piksela, broj merenja i broj nezavisnih parcela.

3.2 Satelitski snimci

U okviru projekta Copernicus, jednog od najznačajnijih programa za posmatranje Zemlje, Evropska svemirska agencija (engl. European Space Agency – ESA) lansirala je nekoliko Sentinel satelita. Ovi sateliti imaju ključnu ulogu u prikupljanju podataka o stanju atmosfere, kopna i okeana, omogućavajući precizno i kontinuirano praćenje i analizu naselja, useva, vodenih površina, šuma, klimatskih promena, ekoloških procesa, prirodnih katastrofa i drugih relevantnih fenomena. Sentinel misije podeljene su u nekoliko kategorija, a za potrebe ovog istraživanja najbitnije su Sentinel-1 i 2. Sentinel-1 pruža radarske slike visoke rezolucije koje su nezavisne od meteorolških uslova i doba dana, a Sentinel-2 optičke slike visoke rezolucije koje su neizostavne pri praćenju promena vegetacije. Iako ovi sateliti generišu slike nepestrano obuhvatajući različite delove Zemljine površine, koristi se samo deo ovih snimaka na kojima se nalaze regioni od interesa u vremenskom periodu koji prati fenološke faze rasta posmatrane poljoprivredne kulture. Na ovaj način se veličina satelitskih slika pratično svodi na veličinu regiona od interesa, a broj slika u posmatranoj vremenskoj seriji zavisi od fenoloških faza posmatranih kultura.

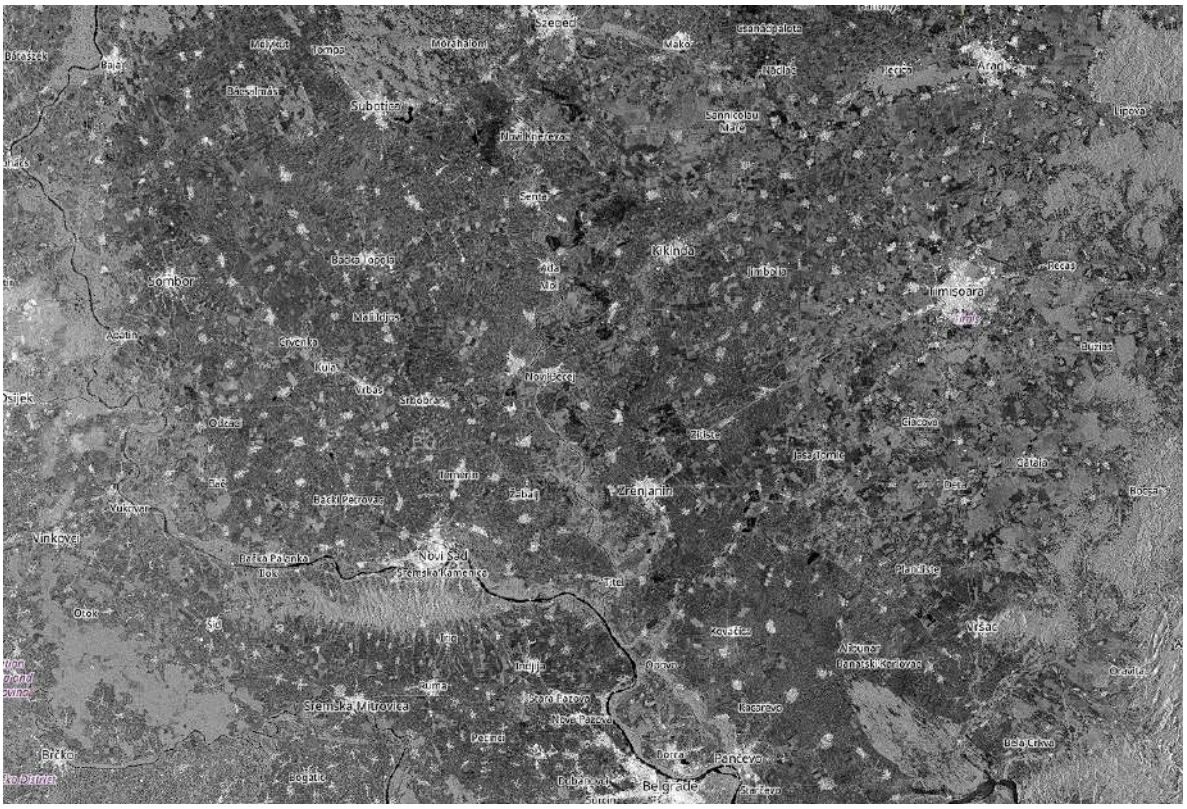
3.2.1 Radarski satelitski snimci

Sentinel-1 predstavlja seriju radarskih satelita za posmatranje Zemlje, razvijenu u okviru programa Copernicus. Seriju čine sateliti Sentinel-1A i Sentinel-1B, kao i njihovi naslednici Sentinel-1C i Sentinel-1D. Sateliti se nalaze u približno polarnoj, Sunčevo-sinhronoj orbiti na nominalnoj visini od oko 693 km.

Sateliti Sentinel-1 koriste radar sa sintetičkom aperturom (engl. Synthetic Aperture Radar – SAR), koji omogućava prikupljanje podataka nezavisno od Sunčeve svetlosti, oblačnosti, padavina i većine drugih atmosferskih uslova. Instrument radi u C mikrotalasnom opsegu, sa centralnom frekvencijom od 5,405 GHz, što odgovara talasnoj dužini od približno 5,6 cm.

Za snimanje kopnenih površina najčešće se koristi režim *Interferometric Wide Swath* (IW), u kojem su podaci uglavnom dostupni u dvostrukoj polarizaciji VV+VH. Oznake polarizacija predstavljaju orijentaciju emitovanog i primljenog elektromagnetnog talasa. Kanal VV označava vertikalno emitovanje i vertikalni prijem signala, dok kanal VH označava vertikalno emitovanje i horizontalni prijem signala. Oba polarizaciona kanala imaju jednaku prostornu rezoluciju, jer rezolucija zavisi od režima snimanja i vrste proizvoda, a ne od polarizacije.

Nominalna prostorna rezolucija IW režima iznosi približno 5×20 m, u pravcu



Slika 3.1: Prikaz područja AP Vojvodine na osnovu Sentinel-1 radarskog snimka u VV polarizaciji. Svetlije vrednosti odgovaraju većem intenzitetu povratnog radarskog signala, dok tamnije oblasti predstavljaju površine sa manjim povratnim signalom.

Tabela 3.1: Osnovne karakteristike polarizacionih kanala Sentinel-1 IW podataka

Kanal	**Vrsta polarizacije**	**Nominalna prostorna rezolucija IW režima (m)**	**Prostorna rezolucija IW GRD-HR proizvoda (m)**
VV	Vertikalno emitovanje i vertikalni prijem	5×20	približno 20×22
VH	Vertikalno emitovanje i horizontalni prijem	5×20	približno 20×22

rastojanja od senzora (engl. range) i pravcu kretanja satelita (engl. azimuth). Međutim, standardni IW GRD proizvod visoke rezolucije (*GRD High Resolution – GRD-HR*), koji se najčešće koristi za analizu intenziteta povratnog radarskog signala, ima efektivnu prostornu rezoluciju od približno 20×22 m i razmak između centara susednih piksela od 10×10 m. Zbog toga veličinu piksela od 10 m ne treba poistovećivati sa stvarnom prostornom rezolucijom snimka.

Sentinel-1 podržava više režima snimanja: *Interferometric Wide Swath* (IW), *Extra-Wide Swath* (EW), *Stripmap* (SM) i *Wave* (WV). Osnovne karakteristike ovih režima prikazane su u Tabeli 3.2.

Tabela 3.2: Osnovne karakteristike režima snimanja Sentinel-1 satelita

Režim snimanja	**Širina snimanja (km)**	**Nominalna prostorna rezolucija (m)**	**Najčešća primena**
IW	250	5×20	Kopnene površine, poljoprivreda i interferometrijska analiza
EW	400–410	20×40	Morske, priobalne i polarne oblasti
SM	80	približno 5×5	Detaljna snimanja manjih područja
WV	snimci dimenzija 20×20	približno 5×5	Praćenje talasa i otvorenih okeana

Režim IW predstavlja osnovni režim snimanja kopnenih površina. On koristi tehniku TOPSAR (engl. Terrain Observation with Progressive Scans SAR) i sastoji se od tri delimično preklapajuća podpojasa, označena kao IW1, IW2 i IW3. Ukupna širina snimljene trake iznosi približno 250 km, što omogućava pokrivanje velikih teritorija tokom jednog satelitskog prolaza.

Sentinel-1 proizvodi nisu organizovani prema fiksnoj mreži granula kao proizvodi Sentinel-2 misije. Umesto toga, proizvodi se formiraju kao scene ili vremenski isečci duž putanje satelita. Zbog dimenzija, oblika i geografskog položaja AP Vojvodine, za njeno potpuno pokrivanje u okviru jednog orbitalnog prolaza najčešće su potrebna dva susedna IW GRD proizvoda, odnosno dve scene ili granule. Jedan proizvod može obuhvatiti veliki deo teritorije AP Vojvodine, dok tačan broj proizvoda zavisi od relativne orbite, smera prolaza, datuma snimanja i načina podele originalnog segmenta na proizvode. Zbog toga se za konkretan skup podataka broj granula mora proveriti na osnovu prostornih

Tabela 3.3: Karakteristike Sentinel-1 IW snimanja područja AP Vojvodine

Karakteristika	**Vrednost**
Nominalna visina orbite	približno 693 km
Širina snimanja u IW režimu	približno 250 km
Broj polarizacionih kanala	2, VV i VH
Veličina piksela IW GRD-HR proizvoda	10×10 m
Efektivna prostorna rezolucija IW GRD-HR proizvoda	približno 20×22 m
Tipičan broj scena za pokrivanje AP Vojvodine	približno 2 scene po orbitalnom prolazu
Nominalni ciklus ponavljanja jednog satelita	12 dana

granica preuzetih proizvoda.

Različiti režimi snimanja omogućavaju primenu Sentinel-1 podataka u praćenju poljoprivrede, vlažnosti zemljišta, poplava, klizišta, morskog leda, deformacija tla i drugih prirodnih i antropogenih procesa.

Interferometrijska SAR analiza (*InSAR*) koristi fazne razlike između radarskih snimaka prikupljenih iz slične geometrije snimanja u različitim vremenskim trenucima. U povoljnim uslovima ova metoda omogućava detekciju veoma malih pomeranja Zemljine površine, uključujući pomeranja milimetarskog do centimetarskog reda u pravcu linije pogleda satelita. Zbog toga se InSAR primenjuje za praćenje sleganja tla, klizišta, seizmičkih deformacija, vulkanske aktivnosti i drugih geodinamičkih procesa.

Nominalni ciklus ponavljanja orbite jednog Sentinel-1 satelita iznosi 12 dana. Kada su dva satelita raspoređena u istoj orbitalnoj ravni, nominalno vreme ponovnog snimanja iste oblasti može se smanjiti na približno šest dana. Stvarna učestalost dostupnih snimaka može zavisiti od geografske lokacije, smera orbite i operativnog plana snimanja.

Podaci ove misije dostupni su besplatno za naučnu, operativnu i komercijalnu upotrebu putem *Copernicus Open Access Hub platforme*.

3.2.2 Optički satelitski snimci

Sentinel-2 misiju čine dva indentična satelita Sentinel-2A i Sentinel-2B koji služe za monitoring Zemljine površine pomoću optičkih multispektralnih senzora visoke rezolucije. Ovi sateliti imaju ugrađen optički multispektralni instrument (engl. MultiSpectral Instrument – MSI) koji detektuje reflektansu Zemljine površine u vidljivom - VIS (engl.

visible), blisko infracrvenom - NIR (engl. Near-Infra Red) i kratkotalasnom infracrvenom - SWIR (engl. Short Wave Infra Red) opsegu koji su izdeljeni na 13 kanala. Raspored kanala i talasne dužine date su u tabeli 3.4. Treba primetiti da prostorna rezolucija kanala varira (može biti 10 m, 20 m i 60 m), stoga je neophodno obezbediti da rezolucije svih kanala budu 10 m, što se može realizovati interpolacijom piksela - *upsampling* slika niže rezolucije.

Tabela 3.4: Spektralni kanali multispektralnog senzora na Sentinel-2 satelitima

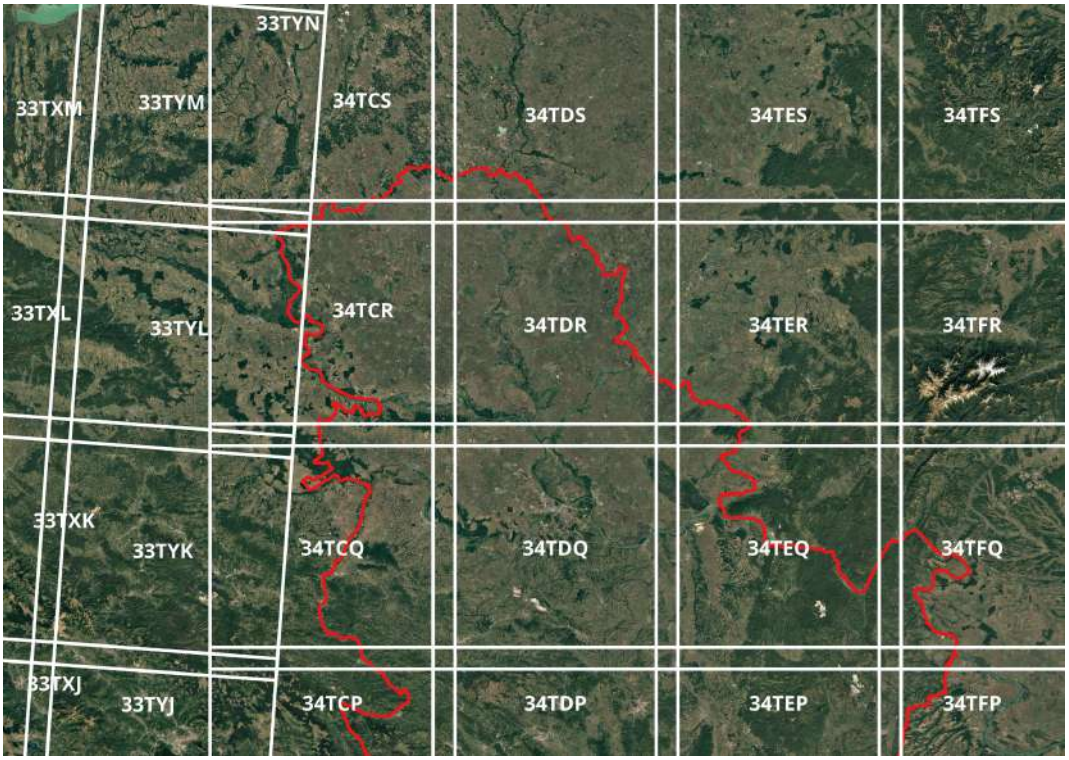
Kanal	**Naziv kanala**	**Akronim**	**Centralna talasna dužina (nm)**	**Prostorna rezolucija (m)**
B01	*Coastal aerosol*	CA	443	60
B02	*Blue*	BLUE	490	10
B03	*Green*	GREEN	560	10
B04	*Red*	RED	665	10
B05	*Vegetation red edge 1*	RedEdge1	705	20
B06	*Vegetation red edge 2*	RedEdge2	740	20
B07	*Vegetation red edge 3*	RedEdge3	783	20
B08	*Near infrared*	NIR	842	10
B8A	*Narrow near infrared*	NIRn	865	20
B09	*Water vapour*	WV	945	60
B10	*Short-wave infrared / Cirrus*	CIRRUS	1375	60
B11	*Short-wave infrared 1*	SWIR1	1610	20
B12	*Short-wave infrared 2*	SWIR2	2190	20

Konstelacijom ova dva satelita omogućeno je da se vreme snimanja istog područija smanji sa početnih 10 dana sa jednim satelitom na svakih 5 dana koristeći oba satelita. Kako njihove orbite nisu identične, iako su na istoj udaljenosti od površine zemlje (786 km), snimke jednog i drugog satelit neophodno je registrovati kako bi se dobilo preklapanje piksela u istoj geografskoj poziciji. Na ovaj način formira se baza podataka koja se može sačuvati u vidu tabelarnih podataka za vrednosti piksela za svaki pojedinačni kanal i dodatnim informacija u vidu geolokacije i identifikacionog broja svakog piksela.

MSI koristi *push-broom* senzore koji zahvataju širinu snimanja od 290 km u jednom prolazu, kreirajući slike koje zahvataju površinu od 10000 km^2. Veličina satelitskog snimka varira oko 1 GB u zavisnosti od širine zahvata i putanje satelita.

Podaci sa Sentinel-2 satelita mogu se pronaći u nekoliko modova: Level-1B: Podatke na ovom nivou čine vrednosti radijacije koji se očitavaju na senzorima satelita. Satelitski podaci izdeljeni su na manje slike, ondnosno granule, čime se omogućava lakša pretraga i vizuelizacija snimaka. Level-1C: Podaci prestavljaju vrednosti reflektanse. Ove slike predstavljalju merenja multispektralnih senzora satelita koja u sebi sadrže atmosferski uticaj i nalaze se u UTM/WGS84 projekciji. Level-2A: Podaci predstavljaju vrednosti reflektanse s tim što je uključena atmosferska i geometrijska korekcija (uključujući ortorektifikaciju i prostornu registraciju). U radu su isključivo korišćene slike ovog nivoa obrade.

Područije AP Vojvodine zahvata 8 granula (Slika 3.2): 34TCS, 34TCR i 34TCQ iz putanje R136 i T34TDS, T34TDR, T34TDQ, T34TER i T34TEQ iz putanje R036. Od najvećeg interesa su granule T34TCR, T34TCQ, T34TDR i T34TDQ jer zahvataju najveći deo površine AP Vojvodine dok ostale granule zahvataju samo male pogranične delove.



Slika 3.2: Prostorni raspored Sentinel-2 granula koje prekrivaju teritoriju Autonomne pokrajine Vojvodine. Bela mreža predstavlja granice Sentinel-2 granula sa pripadajućim oznakama, dok je crvenom linijom prikazana državna granica Srbije.

3.3 Monitor prinosa

Vektorski podaci predstavljaju standardizovan format za skladištenje i razmenu georeferenciranih prostornih informacija i široko se koriste u okviru GIS (Geografski Informacioni Sistem) analiza.

Prvi tip vektorskih podataka čine poligonalni vektori granica parcela, koji omogućavaju precizno definisanje prostornog obuhvata analiziranih polja. Ovi vektori se koriste kao osnovni referentni sloj za isecanje (engl. clipping) satelitskih snimaka i drugih rasterizovanih prostornih podataka, čime se analiza ograničava isključivo na površine od interesa. Na taj način obezbeđuje se prostorna konzistentnost između satelitskih snimaka, podataka o zemljištu i terenskih merenja, kao i smanjenje uticaja okolnih, nepovezanih površina.

Drugi tip vektorskih podataka obuhvata tačkaste vektorske zapise dobijene sa monitora prinosa, koji predstavljaju diskretna merenja količine prikupljenog useva duž putanje kretanja kombajna. Svaka tačka sadrži informaciju o geografskoj poziciji, trenutnom prinosu i vremenskoj oznaci, čime se omogućava rekonstrukcija putanje kombajna i analiza prostorne varijabilnosti prinosa unutar parcele. Integracijom ovih tačkastih podataka sa poligonalnim granicama parcela i satelitskim snimcima omogućava se detaljna analiza odnosa između stanja vegetacije i ostvarenog prinosa.

Najjednostavniji pristup za transformaciju tačkasto-vektorskih podataka sa monitora prinosa u kontinuirane prostorne mape jeste primena interpolacionih metoda. Međutim, ovakav pristup uvodi sistematske greške usled neujednačene prostorne gustine merenja, koja je direktno zavisna od brzine kretanja kombajna. U zonama sporijeg kretanja kombajna broj merenja je veći, dok su u zonama bržeg kretanja merenja ređa, iako ukupna količina prikupljenog useva može biti ista. Kao posledica toga, standardne interpolacione tehnike, poput jednostavnog pokretnog proseka (engl. Simple Moving Average – SMA), mogu dovesti do pristrasne procene prinosa i njegovog veštačkog uvećavanja ili umanjenja.

Kako bi se prevazišla navedena ograničenja, u ovom istraživanju primenjena je Metoda preseka poligona i piksela (engl. Polygon–Pixel Intersection – PPI), predložena u Pejak et al. 2022, koja omogućava direktno usklađivanje podataka sa monitora prinosa i satelitskih snimaka. Ova metoda zasniva se na prostornom preklapanju vektorskih zapisa prinosa sa rasterizovanim pikselima satelitskih snimaka, čime se obezbeđuje konzistentna raspodela prinosa nezavisno od brzine kretanja mašine i gustine merenja. Performanse PPI metode upoređene su sa rezultatima dobijenim primenom SMA interpolacije, a detaljna analiza i poređenje prikazani su u poglavlju Rezultati.

Monitori prinosa na kombajnima predstavljaju ključnu tehnologiju u preciznoj

poljoprivredi, omogućavajući prikupljanje podataka o prinosu useva. Ovi sistemi koriste senzore postavljene na kombajnima kako bi merili različite parametre koji utiču na produktivnost poljoprivrednih parcela. Prikupljeni podaci se kasnije mogu koristiti za optimizaciju poljoprivredne proizvodnje, analizu faktora koji utiču na prinos i razvoj modela mašinskog učenja za predikciju i poboljšanje prinosa.

Moderni monitori prinosa sadrže više senzora koji beleže sledeće podatke:

1. Količina prinosa – Merenje mase ili zapremine useva koji prolazi kroz kombajn, obično korišćenjem senzora mase i protoka.
2. Vlažnost zrna – Određuje sadržaj vode u prinosu, što je važno za procenu kvaliteta i potrebe za sušenjem.
3. Brzina kombajna – Utvrđuje koliko brzo se mašina kreće kroz njivu, što može uticati na preciznost merenja prinosa.
4. Širina zahvata hedera – Informacija o širini sečenja koja omogućava tačno određivanje površine s koje je prikupljen prinos.
5. GPS lokacija – Precizno geolociranje svake tačke merenja omogućava kreiranje mapa prinosa za celokupnu parcelu.
6. Ugao terena i nagib kombajna – Pomaže u korekciji merenja jer različiti nagibi terena mogu uticati na protok zrna kroz kombajn.

Podaci prikupljeni monitorima prinosa predstavljaju bogat izvor informacija za obuku modela mašinskog učenja. Ovi modeli mogu pomoći u donošenju odluka u vezi sa agronomskim praksama, optimizaciji upotrebe resursa i predikciji prinosa. Ključne primene uključuju:

1. Predikcija prinosa – Modeli mašinskog učenja mogu koristiti istorijske podatke o prinosu, vremenskim uslovima i karakteristikama zemljišta za predviđanje budućih prinosa.
2. Optimizacija gajenja useva – Na osnovu analize faktora koji utiču na prinos, mogu se optimizovati strategije đubrenja, navodnjavanja i zaštite useva.
3. Detekcija anomalija – Analiza podataka može otkriti probleme poput neujednačenog prinosa, bolesti biljaka ili problema sa zemljištem.
4. Automatizovana kontrola mašina – Modeli mogu pomoći u automatskom podešavanju parametara kombajna u zavisnosti od uslova na terenu kako bi se povećala efikasnost žetve.

5. Izrada mapa prinosa i zona upravljanja – Na osnovu prikupljenih podataka mogu se kreirati precizne mape koje pomažu u diferenciranoj obradi polja.

3.4 Podaci zemljišnih karakteristika

Podaci o karakteristikama zemljišta sa SoilGrids platforme Hengl et al. 2017 predstavljaju visokorezolucijske prostorne podatke zasnovane na mašinskom učenju, koji pružaju globalne informacije o sastavu i kvalitetu zemljišta na različitim dubinama. Ovi podaci uključuju ključne parametre kao što su tekstura zemljišta (pesak, glina, ilovača), organska materija, pH vrednost, sadržaj ugljenika, gustina zemljišta i kapacitet zadržavanja vode. Korišćenjem SoilGrids podataka u kombinaciji sa vektorskim parcelama i podacima monitora prinosa, moguće je identifikovati kako karakteristike zemljišta utiču na varijabilnost prinosa. Takođe, ovi podaci omogućavaju preciznije preklapanje sa satelitskim snimcima i kreiranje prostornih modela za optimizaciju poljoprivredne proizvodnje. Integracija SoilGrids informacija sa podacima sa terena i vremenskim podacima doprinosi razvoju naprednih modela mašinskog učenja za predikciju prinosa i donošenje informisanih agronomskih odluka.

SoilGrids je globalna baza podataka o zemljištu razvijena od strane Međunarodnog instituta za analizu primenjenih sistema (World Soil Information - ISRIC, [Hengl et al. 2017]). Ova baza koristi napredne metode mašinskog učenja za generisanje visoko rezolucionih mapa zemljišnih svojstava na globalnom nivou. SoilGrids omogućava istraživačima, poljoprivrednicima i donosiocima odluka pristup detaljnim informacijama o karakteristikama zemljišta, koje su ključne za razumevanje i upravljanje prirodnim resursima.

Glavni parametri SoilGrids baze podataka uključuju:

- **Rezolucija**: Podaci su dostupni u rezoluciji od 250 m x 250 m, čime omogućavaju analize zemljišta na globalnom i regionalnom nivou.
- **Zemljišna svojstava**: SoilGrids pruža informacije o više zemljišnih parametara, uključujući teksturu zemljišta (procenat peska, mulja i gline), gustinu zemljišta, sadržaj organske materije, pH vrednost, kapacitet zadržavanja vode i druge hemijske i fizičke karakteristike.
- **Metodologija zasnovana na podacima**: SoilGrids koristi mašinsko učenje i podatke iz globalnih pedoloških baza kao što su Soil Information System (WoSIS) kako bi kreirao modele predikcije.

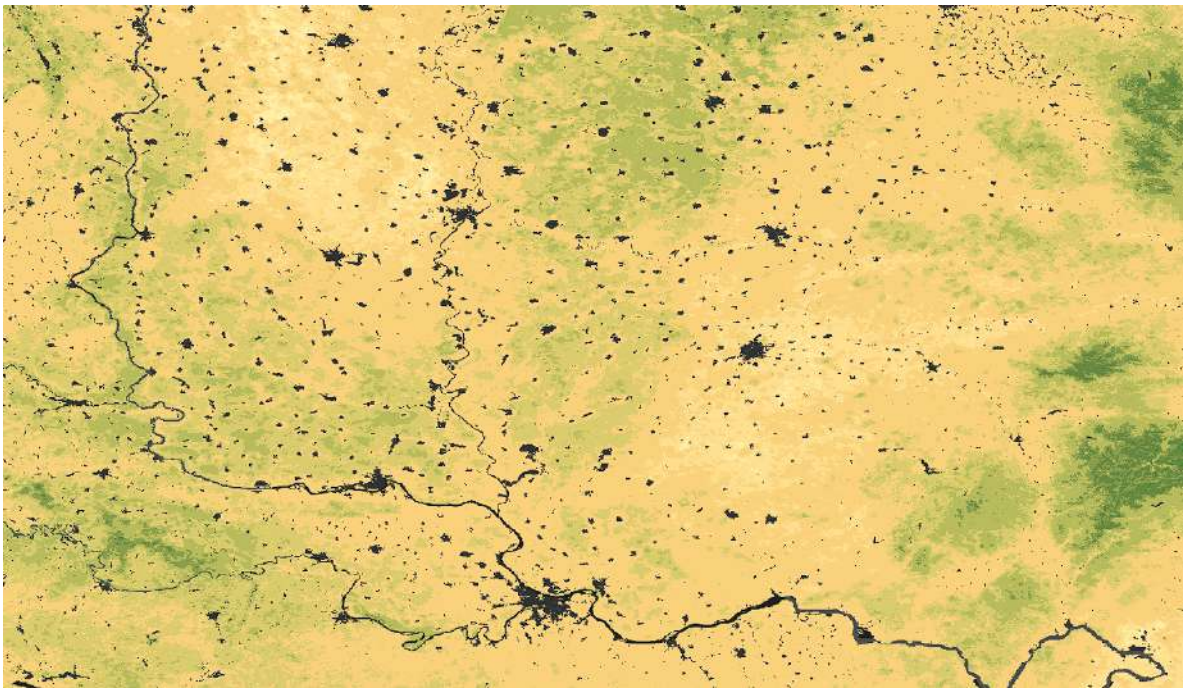
- **Otvoreni podaci**: Podaci su besplatno dostupni istraživačima i praktičarima putem REST API-ja i GIS softverskih alata.

Baza SoilGrids se kreira kombinovanjem postojećih merenja iz različitih izvora sa pomoćnim promenljivama kao što su satelitski snimci, digitalni modeli terena i klimatski podaci. Glavna metoda predikcije koristi regresione ansambl tehnike poput *Random Forest* algoritma za interpolaciju i mapiranje zemljišnih karakteristika.

Podaci u SoilGrids predstavljeni su na sedam standardnih dubinskih slojeva: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm

Dubinske sekcije omogućavaju analizu varijacija zemljišnih svojstava u profilu, što je ključno za modeliranje interakcija između zemljišta i biljaka, kao i za simulacije hidroloških i ekoloških procesa.

SoilGrids je značajan alat u oblastima poljoprivrede, ekologije i klimatskih istraživanja. Korišćenjem ovih podataka, naučnici i stručnjaci mogu bolje razumeti prostorne i vremenske varijacije zemljišta, predviđati produktivnost zemljišta, planirati održive prakse upravljanja i razvijati preciznu poljoprivredu.



Slika 3.3: Prostorna raspodela organskog ugljenika u zemljištu (SOC stock) u površinskom sloju od 0 do 30 cm, prema podacima SoilGrids. Zelenije nijanse označavaju područja sa većom koncetracijom organskog ugljenika, dok žute nijanse predstavljaju područja sa nižim vrednostima.

Najnovija verzija SoilGrids podataka može se preuzeti i koristiti putem online

platforme, a dostupni su i alati za integraciju sa GIS softverima kao što su QGIS i ArcGIS. SoilGrids podaci su kompatibilni sa različitim formatima, uključujući raster i API pristup za analitičke svrhe.

Za obuku algoritama mašinskog učenja koriste se podaci zemljiših karakteristika sa platforme SoilGrids, pri čemu su odabrani sledeći parametri:

- **bdob** – Gustina zemljišta (engl. bulk density), izražena u kg/m^3, koja pokazuje zbijenost zemljišta i njegovu poroznost.
- **cec** – Kationski zamenski kapacitet (engl. cation exchange capacity), izražen u cmol/kg, koji ukazuje na sposobnost zemljišta da zadrži i razmenjuje hranljive materije.
- **cfvo** – Količina grubih fragmenata u zemljištu (engl. coarse fragments volume fraction), izražena u %, koja označava udeo čestica većih od 2 mm.
- **clay** – Udeo gline u zemljištu (engl. clay fraction), izražen u %, koji utiče na zadržavanje vode i plodnost zemljišta.
- **nitrogen** – Sadržaj ukupnog azota (engl. total nitrogen), izražen u g/kg, ključan za plodnost i biološku aktivnost zemljišta.
- **ocd** – Sadržaj organskog ugljenika u zemljištu (engl. organic carbon density), izražen u g/m^3, koji je važan za skladištenje ugljenika i plodnost zemljišta.
- **ocs** – Zaliha organskog ugljenika u zemljištu (engl. organic carbon stock), izražena u t/ha, koja pokazuje kapacitet zemljišta za sekvestraciju ugljenika.
- **phh2o** – pH vrednost zemljišta u vodi (engl. pH in H_2O), koja pokazuje kiselost ili baznost zemljišta i utiče na dostupnost hranljivih materija.
- **sand** – Udeo peska u zemljištu (engl. sand fraction), izražen u %, koji utiče na drenažu i aeraciju zemljišta.
- **silt** – Udeo mulja u zemljištu (engl. silt fraction), izražen u %, koji utiče na sposobnost zemljišta da zadrži vodu i hranljive materije.
- **soc** – Sadržaj organskog ugljenika (engl. soil organic carbon), izražen u g/kg, koji doprinosi plodnosti zemljišta i skladištenju ugljenika.

Ovi parametri su ključni za analizu plodnosti zemljišta, kapaciteta za zadržavanje vode, hemijskih karakteristika i ukupnog kvaliteta zemljišta. Podaci se mogu koristiti za

modeliranje ekoloških procesa, procenu degradacije zemljišta i optimizaciju poljoprivrednih praksi.

3.5 Digitalni elevacioni model

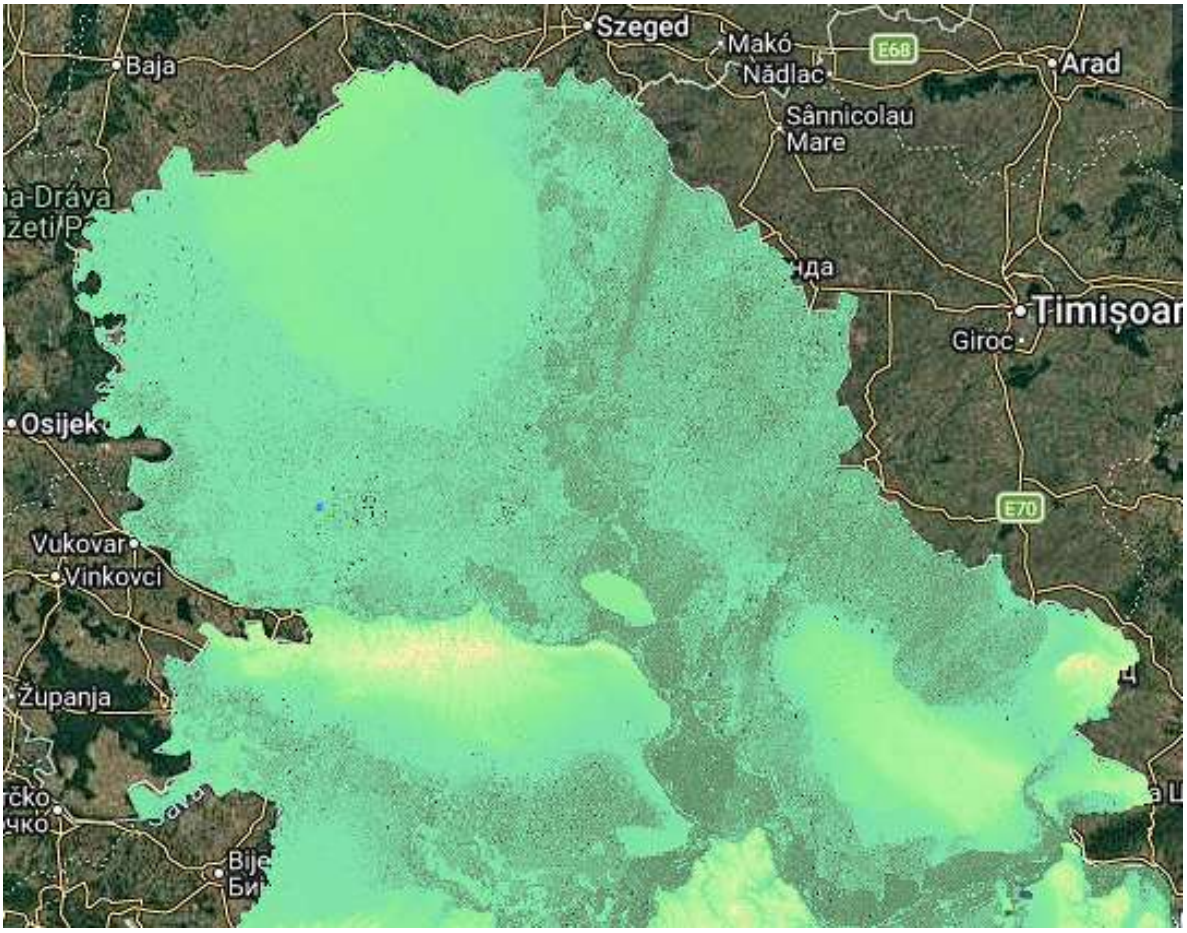
Digitalni elevacioni model (DEM) (Farr, Rosen, et al. 2007) predstavlja rasterizovani prikaz nadmorske visine terena, koji omogućava detaljnu analizu topografskih karakteristika poljoprivrednih parcela. U okviru ove disertacije koristi se globalni DEM proizvod visoke prostorne rezolucije (30 m), generisan interferometrijskom radarskom tehnikom, sa vertikalnom tačnošću reda nekoliko metara. Ovi podaci su reprojektovani i resamplovani na zajedničku prostornu rezoluciju sa satelitskim snimcima Sentinel-2 misije (10 m), čime je omogućena njihova direktna integracija u jedinstveni prostorni referentni okvir.

DEM podaci omogućavaju izvođenje sekundarnih topografskih parametara, kao što su nagib terena, ekspozicija i indeksi drenaze, koji imaju značajan uticaj na prostornu distribuciju vlage, procese erozije i varijabilnost prinosa. Korišćenjem DEM-a u kombinaciji sa vektorskim granicama parcela, podacima sa monitora prinosa i multispektralnim satelitskim snimcima, omogućena je precizna analiza uticaja mikroreljefa na prostorne razlike u produktivnosti useva.

Pored toga, digitalni elevacioni modeli koriste se za modeliranje kretanja površinskih i plitkih podzemnih voda, optimizaciju sistema za navodnjavanje i odvodnjavanje, kao i za identifikaciju zona sa povećanim rizikom od zadržavanja vode ili ispiranja hranljivih materija. Integracija DEM podataka sa ostalim prostornim slojevima doprinosi preciznijem modeliranju agroekoloških faktora i unapređenju poljoprivrednih praksi zasnovanih na prostornim podacima.

3.6 Meteo podaci

Podaci sa ERA5 Climate Data Store (Hersbach et al. 2020) obuhvataju dugoročne meteorološke zapise. Meteo podaci uključuju ključne parametre poput temperature vazduha, količine padavina, relativne vlažnosti, brzine i pravca vetra, sunčevog zračenja i evapotranspiracije, što omogućava detaljnu analizu uticaja klimatskih faktora na poljoprivrednu proizvodnju. Kombinovanjem ERA5 podataka sa vektorskim parcelama, podacima monitora prinosa i satelitskim snimcima, moguće je preciznije modelirati vremenske uticaje na rast i razvoj useva, kao i predikciju prinosa u različitim klimatskim scenarijima. Takođe, upotreba meteo podataka je od suštinskog značaja za razvoj modela mašinskog učenja koji analiziraju dugoročne klimatske trendove i njihov uticaj



Slika 3.4: Digitalni elevacioni model (DEM) područja AP Vojvodine prostorne rezolucije 10 m. Zelene nijanse predstavljaju područja niže nadmorske visine, dok žute i braon nijanse označavaju više terene.

na poljoprivredu, omogućavajući donosiocima odluka da optimizuju strategije upravljanja resursima i smanje rizike povezane sa ekstremnim vremenskim uslovima.

Za analizu meteoroloških karakteristika koriste se podaci iz ERA5 skupa podataka, pri čemu su odabrani sledeći parametri:

- **d2m** – Tačka rose na 2 m visine (engl. 2m dewpoint temperature), izražena u K (Kelvinima), koja pokazuje temperaturu pri kojoj vazduh postaje zasićen vodom. Ovaj parametar se koristi za analizu vlažnosti vazduha i isparavanja. Vrednosti su izračunate kao mesečni prosek.
- **t2m** – Temperatura vazduha na 2 m visine (engl. 2m temperature), izražena u K, koja pokazuje površinsku temperaturu vazduha. Ovaj parametar je ključan za analizu klimatskih uslova i njihov uticaj na ekosisteme. Vrednosti su izračunate kao mesečni prosek.

- **slhf** – Tok latentne toplote (engl. surface latent heat flux), izražen u W/m^2, koji pokazuje količinu energije oslobođene ili apsorbovane tokom isparavanja ili kondenzacije vode. Ovaj parametar je važan za procenu energetskog bilansa površine. Vrednosti su izračunate kao mesečni prosek.

- **lai_hv** – Indeks lisne površine za visoku vegetaciju (engl. leaf area index for high vegetation), koji pokazuje gustinu vegetacije i njenu sposobnost fotosinteze. Ovaj indeks je značajan za modeliranje biljnog rasta i ciklusa ugljenika. Vrednosti su izračunate kao mesečni prosek.

- **lai_lv** – Indeks lisne površine za nisku vegetaciju (engl. leaf area index for low vegetation), koji opisuje pokrivenost tla vegetacijom u nižim slojevima. Ovaj parametar je koristan za analize pokrivenosti zemljišta i evapotranspiracije. Vrednosti su izračunate kao mesečni prosek.

- **tp** – Ukupne padavine (engl. total precipitation), izražene u m, koje predstavljaju ukupnu količinu padavina tokom meseca. Ovaj parametar je ključan za analizu vodnog bilansa i hidroloških procesa. Vrednosti su izračunate kao mesečne sume.

- **e** – Isparavanje (engl. evaporation), izraženo u m, koje pokazuje količinu vode koja je isparila sa površine tla i vegetacije. Ovaj parametar se koristi za procenu hidrološke ravnoteže i vodenih resursa. Vrednosti su izračunate kao mesečne sume.

- **pev** – Potencijalna evapotranspiracija (engl. potential evaporation), izražena u m, koja pokazuje maksimalnu moguću količinu isparavanja pri datim vremenskim uslovima. Ovaj parametar je važan za modeliranje vodnog bilansa i planiranje navodnjavanja. Vrednosti su izračunate kao mesečne sume.

3.7 Podaci o vrstama poljoprivrednih kultura

Podaci o vrstama poljoprivrednih kultura predstavljaju jedan od ključnih podataka u sistemima precizne poljoprivrede, jer omogućavaju identifikaciju i prostorno razgraničenje različitih useva na poljoprivrednim parcelama. Podaci o vrstama useva se najčešće generišu primenom metoda daljinske detekcije, korišćenjem multispektralnih satelitskih snimaka visoke prostorne rezolucije, u kombinaciji sa terenskim podacima za za treniranje modela mažinskog učenja.

Podaci o tipovima kultura prikupljeni su terenskim radom uz korišćenje aplikacije FieldApp, Slika 3.5. Upotrebom ove aplikacije omogućeno je precizno georeferenciranje prikupljenih podataka o tipovima kultura na poljoprivrednim parcelama. U svrhu

validacije terenskih podataka, kroz MkSens aplikaciju formirana je i baza fotografija useva, koja je korišćena za dodatnu proveru ispravnosti oznaka kultura.



Slika 3.5: Korisnički interfejs mobilne aplikacije FieldApp za terensko prikupljanje podataka o poljoprivrednim kulturama. Aplikacija omogućava unos podataka o usevu, predusevu, sorti ili hibridu, datumu setve, fenofazi, opštem stanju biljke i prisustvu bolesti, korova i štetočina. Prikupljeni podaci mogu se koristiti kao referentne informacije za validaciju satelitskih analiza i modela mašinskog učenja.

Pomoću softverskog alata QGIS geolokacije snimljenih fotografija transformisane su u vektorske tačke, odnosno georeferencirane pozicije snimanja fotografija. Na osnovu oznaka tipa kulture i prostornih pozicija prikupljenih tačaka, u istom softverskom okruženju formirani su poligoni na parcelama kojima su odgovarajuće vektorske tačke pripadale. Na taj način formirana je baza podataka o tipovima kultura, koja je dalje korišćena za obuku klasifikacionih algoritama mašinskog učenja.

Studija sprovedena na području Vojvodine (severni deo Srbije) Crnojević et al. 2014

pokazala je da sinergijska upotreba satelitskih snimaka omogućava pouzdanu klasifikaciju malih poljoprivrednih parcela, čime se značajno unapređuje mapiranje useva. Za potrebe daljih analiza u doktorskoj disertaciji korišćene su mape klasifikacije kultura karakterističnih za poljoprivrednu proizvodnju u Srbiji, uključujući kukuruz, pšenicu, suncokret, soju i šećernu repu. Klasifikacione mape razvijene su na osnovu multitemporalnih sateliskih snimaka Sentinel-2 misije, uz primenu nadgledanih klasifikacionih algoritama.

Mape klasifikacije useva imaju višestruku primenu u narednim interacijama predikcionih modela. Mape se koriste kao ulazni podaci za modele predikcije prinosa, za detekciju anomalija u razvoju vegetacije, kao i za podršku pri donošenju odluka u planiranju poljoprivredne proizvđnje, uključujući optimizaciju đubrenja i navodnjavanja. Pored toga, klasifikacione mape imaju značajnu ulogu u administrativnim i kontrolnim procesima, kao što su verifikacija subvencija i procena poštovanja plodoreda, jer omogućavaju nezavisnu i objektivnu proveru podataka iz registara poljoprivrednih gazdinstava.

3.8 Podaci o datumima štete

Za potrebe analize uticaja ekstremnih vremenskih događaja na poljoprivrednu proizvodnju formiran je skup podataka o datumima i lokacijama nastanka šteta. Skup podataka o datumima štete predstavlja pomoćni referentni izvor informacija koji je korišćen za vremensko i prostorno povezivanje registrovanih nepogoda sa satelitski izvedenim indikatorima stanja useva, pre svega u okviru analiza zasnovanih na SAPI pristupu.

Baza podataka je strukturirana po godinama i obuhvata period od 2016. do 2025. godine. Za svaku godinu formiran je poseban list, pri čemu svaki zapis sadrži četiri osnovna atributa: datum događaja, oblast na koju se događaj odnosi, tip nepogode i komentar. Na ovaj način omogućeno je da se za svaki registrovani događaj evidentiraju ne samo vreme pojave, već i osnovni prostorni kontekst i opis vrste nepogode. Ukupan skup podataka obuhvata 80 pojedinačnih zapisa raspoređenih kroz desetogodišnji period, pri čemu je broj evidentiranih događaja najveći u 2025. i 2024. godini, a najmanji u 2017. godini. Ovakva raspodela ukazuje da baza nema karakter homogenog zvaničnog statističkog registra, već predstavlja radni skup događaja prikupljenih i sistematizovanih za potrebe istraživanja.

Najčešće evidentirani tipovi nepogoda u bazi su led i oluja, dok su u manjoj meri zastupljeni i zapisi o poplavama, vetru, jakim pljuskovima, nevremenu i superćelijskim olujama. U velikom broju slučajeva tip događaja je naveden kao kombinacija više pojava,

na primer *oluja, led* ili *oluja, led, vetar*, što ukazuje da pojedinačni zapisi često opisuju složene meteorološke epizode sa više istovremenih štetnih faktora. Takva struktura podataka je značajna jer omogućava fleksibilnu interpretaciju, ali istovremeno ukazuje da baza nije strogo standardizovana u terminološkom smislu.

Prostorni atribut *oblast* sadrži nazive naselja, opština, gradova ili širih regiona u kojima je prijavljena šteta. Prostorna preciznost zapisa nije ujednačena: u nekim slučajevima navedena je pojedinačna lokacija, dok se u drugim slučajevima zapisi odnose na šire teritorijalne celine, kao što su Vojvodina, centralna Srbija ili više gradova i opština navedenih u istom redu. Zbog toga se ovaj skup podataka ne može posmatrati kao precizan geoprostorni inventar šteta na nivou parcele, već kao indikativni izvor informacija o vremenu i približnoj oblasti nastanka nepogode. U tom smislu, njegova osnovna vrednost ogleda se u tome što omogućava usmeravanje satelitske analize na odgovarajuće datume i regione, a ne u direktnom određivanju tačne granice oštećenja.

Atribut *komentar* korišćen je za dodatne napomene kada su one bile dostupne. U pojedinim zapisima komentar sadrži informaciju o uvedenoj vanrednoj situaciji, mogućem odstupanju u tačnom datumu događaja ili dodatni opis posledica nepogode. Iako komentari nisu prisutni za sve događaje, oni mogu biti korisni u interpretaciji pojedinačnih slučajeva, naročito kada je potrebno proveriti pouzdanost ili kontekst evidencije.

Za potrebe ove disertacije podaci nisu korišćeni kao samostalan sloj za direktno kartiranje šteta, već kao pomoćni vremensko-prostorni referentni skup za integraciju sa satelitskim vremenskim serijama. Konkretno, registrovani datumi i lokacije nepogoda povezivani su sa promenama vegetacionog signala izvedenim iz Sentinel-2 podataka i SAPI indeksa, kako bi se identifikovale zone u kojima je došlo do odstupanja u odnosu na očekivani sezonski razvoj useva. Na taj način baza događaja služila je kao polazni okvir za izdvajanje kritičnih vremenskih trenutaka i prostornih celina koje zahtevaju detaljniju analizu.

Važno je naglasiti da ovakav skup podataka ima određena ograničenja. Pre svega, zapisi nisu potpuno uniformni u pogledu formata datuma, naziva lokacija i terminologije za tip nepogode. U bazi su prisutne sitne razlike u pisanju istih ili sličnih pojava, kao i zapisi različite prostorne detaljnosti. Takođe, u nekim slučajevima datum može predstavljati dan prijave ili medijski zabeleženog događaja, a ne nužno tačan trenutak nastanka oštećenja na usevu. Zbog toga ove podatke treba tumačiti kao indikativan referentni sloj, a ne kao strogo standardizovan meteorološki ili osiguravajući registar šteta.

Uprkos navedenim ograničenjima, formirana baza datuma i lokacija štete predstavlja značajan deo metodološkog okvira istraživanja. Njena osnovna uloga je da obezbedi vezu

između stvarnih registrovanih nepogoda i satelitski uočenih promena u stanju vegetacije. Time je omogućeno da se analiza šteta ne zasniva isključivo na apstraktnim anomalijama u vremenskim serijama, već da bude povezana sa konkretnim događajima iz realnog proizvodnog i meteorološkog konteksta.

3.9 Podaci o navodnjavanim parcelama

Za potrebe analize navodnjavanja u okviru ove disertacije korišćen je referentni skup podataka o navodnjavanim parcelama u Vojvodini, razvijen u prethodnim istraživanjima Radulović, Brdar, et al. 2023; Radulović, Marković, et al. 2025. Ovaj skup podataka predstavlja jednu od najvažnijih referentnih osnova za proučavanje prostornog rasporeda navodnjavanja u regionu, jer obezbeđuje georeferencirane informacije o lokaciji parcela, tipu useva i prisustvu infrastrukture za navodnjavanje. U metodološkom smislu, pomenuti podaci korišćeni su kao referentni sloj za interpretaciju i proveru prostornih obrazaca dobijenih analizom satelitskih vremenskih serija, odnosno za procenu verovatnoće navodnjavanja na osnovu SAPI pristupa.

Početnu osnovu čini spomenuto istraživanje iz 2023. godine, u kome je razvijen pristup za detekciju navodnjavanja primenom metoda mašinskog učenja nad Sentinel-2 podacima za područje Vojvodine. U toj studiji korišćeni su satelitski snimci i terenski prikupljeni referentni podaci za tri najzastupljenije navodnjavane kulture: kukuruz, soju i šećernu repu, tokom perioda 2020–2022. godine. Ovaj rad je pokazao da se višedatumski Sentinel-2 podaci mogu uspešno koristiti za razlikovanje navodnjavanih i nenavodnjavanih parcela i predstavlja metodološku osnovu za kasnije formiranje proširenog geoprostornog skupa podataka o navodnjavanju.

Na toj osnovi je u radu Radulović i sar. iz 2025. godine objavljen prvi geoprostorni skup podataka o navodnjavanim parcelama u Vojvodini za period 2020–2024. godine. Skup podataka obuhvata ukupno 1256 navodnjavanih parcela i odnosi se na četiri glavne kulture od interesa: kukuruz, soju, šećernu repu i pšenicu. Podaci su organizovani kao vektorski slojevi u shapefile formatu, sa prostornim referentnim sistemom EPSG:32634, i dostupni su kao georeferenciran godišnji skup podataka za svaku sezonu u analiziranom petogodišnjem periodu. Time je obezbeđen standardizovan i javno dostupan referentni izvor za istraživanja koja se bave mapiranjem i modelovanjem navodnjavanja u Vojvodini.

Prikupljanje podataka sprovedeno je kroz pet terenskih kampanja tokom perioda 2020–2024, uglavnom u maju i junu, odnosno u delu sezone kada se prisustvo sistema za navodnjavanje može efikasno proveravati na terenu. Proces formiranja baze sastojao se iz tri osnovna koraka. U prvoj fazi izvršena je priprema terenskih kampanja korišćenjem

vizuelne interpretacije satelitskih snimaka i identifikacije velikih parcela za koje se pretpostavljalo da su pod navodnjavanjem. U drugoj fazi sprovedeno je terensko prikupljanje podataka uz pomoć posebno razvijene mobilne aplikacije za georeferencirano fotografisanje i beleženje dodatnih atributa. Tokom obilaska parcela evidentirani su tip useva i prisustvo infrastrukture za navodnjavanje. Parcele na kojima su zabeleženi sistemi kao što su pivot, linearni sistemi označene su kao navodnjavane, dok su parcele bez takve infrastrukture klasifikovane kao nenavodnjavane. U trećoj fazi sirovi terenski podaci su integrisani u GIS okruženju, gde su za svaku parcelu formirani poligoni sa pridruženim atributima o kulturi i statusu navodnjavanja.

Vremenska raspodela prikupljenih podataka pokazuje da je baza dopunjavana sukcesivno po godinama. Prema opisu skupa podataka, evidentirano je 110 navodnjavanih parcela u 2020. godini, 295 u 2021, 365 u 2022, 194 u 2023. i 292 u 2024. godini. Takva struktura ukazuje da skup podataka nije rezultat jedne jedinstvene kampanje, već kontinualnog višegodišnjeg prikupljanja i proširivanja baze referentnih parcela. Upravo ta višegodišnja komponenta čini ovaj skup podataka posebno vrednim, jer omogućava analizu stabilnosti i promena prostornih obrazaca navodnjavanja kroz više sezona.

U okviru disertacije podaci o navodnjavanim parcelama nisu korišćeni kao direktan izlazni proizvod, već kao referentni i interpretativni sloj za analizu prostornih obrazaca izvedenih iz satelitskih podataka. Konkretno, korišćeni su za proveru i tumačenje zona sa stabilno povišenim vrednostima SAPI indeksa, pod pretpostavkom da parcele sa očuvanim i izraženijim vegetacionim signalom kroz više sezona imaju veću verovatnoću da su pod uticajem dodatnog vodosnabdevanja. Na taj način je omogućeno da se satelitski izvedene procene navodnjavanja uporede sa nezavisnim terenski prikupljenim informacijama o prisustvu sistema za navodnjavanje.

Iako ovaj skup podataka predstavlja izuzetno vredan referentni izvor, potrebno je naglasiti i njegova ograničenja. Pre svega, skup podataka je fokusiran na parcele kod kojih je vizuelno i terenski potvrđeno prisustvo infrastrukture za navodnjavanje, što znači da ne predstavlja potpun inventar svih poljoprivrednih površina u regionu. Takođe, prostorna zastupljenost uzoraka zavisi od organizacije terenskih kampanja i dostupnosti informacija na terenu. Pored toga, prisustvo sistema za navodnjavanje ne mora nužno značiti da je navodnjavanje aktivno korišćeno u svakoj sezoni i na svakom delu parcele. Zbog toga se ovi podaci u ovoj disertaciji tumače kao referentni pokazatelj verovatnog prisustva navodnjavanja, a ne kao apsolutna i potpuna evidencija stvarno realizovanih količina ili termina navodnjavanja.

Uprkos navedenim ograničenjima, skup podataka o navodnjavanim parcelama predstavlja jednu od najvažnijih osnova za istraživanje navodnjavanja u Vojvodini. Njegova najveća vrednost ogleda se u tome što obezbeđuje georeferenciran, višegodišnji i terenski



Slika 3.6: Prostorni prikaz merenog sadržaja proteina unutar poljoprivredne parcele. Tamnije zelene nijanse predstavljaju više vrednosti procentualnog sadržaja proteina, dok svetlije zelene i bele nijanse označavaju niže vrednosti. Tačke predstavljaju pojedinačna merenja prikupljena duž putanje kretanja mehanizacije.

potvrđen referentni sloj koji omogućava razvoj, validaciju i interpretaciju modela zasnovanih na daljinskoj detekciji. U tom smislu, ovaj skup podataka predstavlja ključnu vezu između terenskih opažanja i satelitski izvedenih indikatora upravljanja vodom u poljoprivredi.

3.10 Podaci o proteinskom sastavu

Za potrebe analize kvalitativnih karakteristika useva korišćen je skup referentnih podataka o proteinskom sastavu zrna, prikupljen iz više komplementarnih izvora. Ovi podaci obuhvataju kako prostorno eksplicitna tačkasta merenja, tako i tabelarne rezultate ogleda i lokalitetskih analiza. Na taj način obezbeđen je skup podataka koji omogućava sagledavanje proteinskog sastava i na detaljnom prostornom nivou unutar parcela, i na nivou oglednih ili proizvodnih celina.

Prostorno eksplicitni deo skupa sastoji se od velikog broja georeferenciranih tačaka u kojima je za svaku lokaciju evidentirana vrednost proteinskog sastava. U pojedinim slučajevima, pored proteina, bili su dostupni i dodatni kvalitativni parametri, kao što su vlaga i sadržaj ulja, što je omogućilo potpuniji uvid u kvalitet zrna. Najveći deo ovih podataka odnosi se na soju, pri čemu su vrednosti proteina pokazivale izraženu prostornu varijabilnost unutar posmatranih parcela.

Pored tačkastih merenja, korišćeni su i rezultati ogleda i lokalitetskih analiza u kojima je proteinski sastav zabeležen zajedno sa prinosom, vlagom, sadržajem ulja, datumom žetve i osnovnim informacijama o lokalitetu ili tretmanu. Ovakvi podaci su imali manji broj merenja u odnosu na georeferencirane podatke, ali su posedovali bogatiji agronomski kontekst i omogućavali su povezivanje kvaliteta zrna sa širim uslovima proizvodnje.

Posmatrano u celini, podaci o proteinskom sastavu obuhvatali su više nivoa detaljnosti i više tipova merenja. Jedan deo skupa bio je namenjen detaljnoj prostornoj analizi raspodele proteina unutar parcela, dok je drugi deo pružao informacije o proteinskom sastavu na nivou ogleda, lokaliteta ili proizvodnih varijanti. Takva struktura omogućila je fleksibilnu upotrebu podataka u zavisnosti od cilja analize i tipa modela.

U okviru ove disertacije podaci o proteinskom sastavu korišćeni su kao referentni skup za razvoj i evaluaciju modela predikcije kvaliteta useva. Njihova osnovna uloga bila je da obezbede vezu između satelitski izvedenih indeksa stanja vegetacije i stvarno izmerenih kvalitativnih osobina zrna.

3.11 Zbirni pregled i uloga skupova podataka

Satelitski podaci predstavljaju zajedničku osnovu svih modela, dok podaci sa monitora prinosa, laboratorijska merenja proteina, evidencije kultura, olujnih događaja i navodnjavanja imaju ulogu ciljnih ili referentnih promenljivih. SoilGrids, digitalni model elevacije i meteorološki podaci predstavljaju pomoćne agroekološke prediktore. Pri prikazivanju svakog eksperimenta posebno su navedeni vremenski obuhvat, način

prostornog povezivanja i jedinica analize.

Zbirni pregled svih skupova podataka, njihovih prostornih i vremenskih karakteristika, kao i njihove namene u okviru razvijenog sistema, dat je u Tabeli 3.5.

Tabela 3.5: Zbirni pregled skupova podataka korišćenih u disertaciji

Podatak	**Izvor**	**Period**	**Rezolucija**	**Broj scena, uzoraka ili parcela**	**Namena**
Sentinel-2 multispektralni snimci	ESA / Copernicus	2018–2025	10–60 m	*344 snimaka*	Klasifikacija useva, predikcija prinosa, izračunavanje SAPI indeksa, analiza štetnih događaja, analiza navodnjavanja i zoniranje parcela
Sentinel-1 SAR snimci	ESA / Copernicus	*2018–2025*	10 m	*172 snimaka*	Rekonstrukcija optičkih satelitskih snimaka i dopuna vremenskih serija
Podaci sa monitora prinosa	*Privatna agropreduzeća*	2018–2019	Tačkasta merenja, agregirana na mrežu 10 m × 10 m	*60 parcela*	Formiranje referentnih vrednosti prinosa, obuka i validacija modela za predikciju prinosa
Podaci o proteinskom sadržaju useva	*Privatna agropreduzeća*	*2018 i 2020*	Laboratorijski ili terenski uzorci	*36 parcela*	Klasifikacija proteinskog sadržaja useva i procena generalizacije između različitih izvora podataka
Podaci o navodnjavanim parcelama	*Zenodo/naučni rad*	2020–2022	Vektorski poligoni parcela	*1256 parcela*	Analiza povezanosti poznatih navodnjavanih površina sa SAPI vrednostima
Podaci o olujnim i štetnim događajima	*Interna baza*	*2016-2025*	Lokacije i datumi događaja	*86 događaja*	Analiza anomalija produktivnosti prostorno i vremenski povezanih sa evidentiranim štetnim događajima
Podaci o klasama poljoprivrednih kultura	*Interna baza*	2016–2025	Vektorski poligoni parcela i/ili raster rezolucije 10 m	*20000 parcela*	Obuka, validacija i generisanje mapa klasifikacije poljoprivrednih kultura
SoilGrids podaci	ISRIC – World Soil Information	-	250 m	*10 parametara*	Pomoćna obeležja za predikciju prinosa, analizu produktivnosti i delineaciju homogenih zona
Digitalni model elevacije	*NASA*	-	*30 m*	Jedan rasterski sloj	Izračunavanje topografskih karakteristika i dopuna ulaznih obeležja modela
Meteorološki podaci	ERA5 / ERA5-Land	2018–2025	*9 km*	*Akumulirani dnevni podaci*	Dopuna satelitskih i zemljišnih obeležja za predikciju prinosa i analizu produktivnosti useva
Poligoni poljoprivrednih parcela	*Interna baza*	*2016-2025*	Vektorski poligoni	*20000 parcela*	Prostorno ograničavanje analiza, parcelna validacija, agregacija rezultata i delineacija upravljačkih zona

Poglavlje 4

Opšti metodološki okvir i priprema podataka

Sadržaj

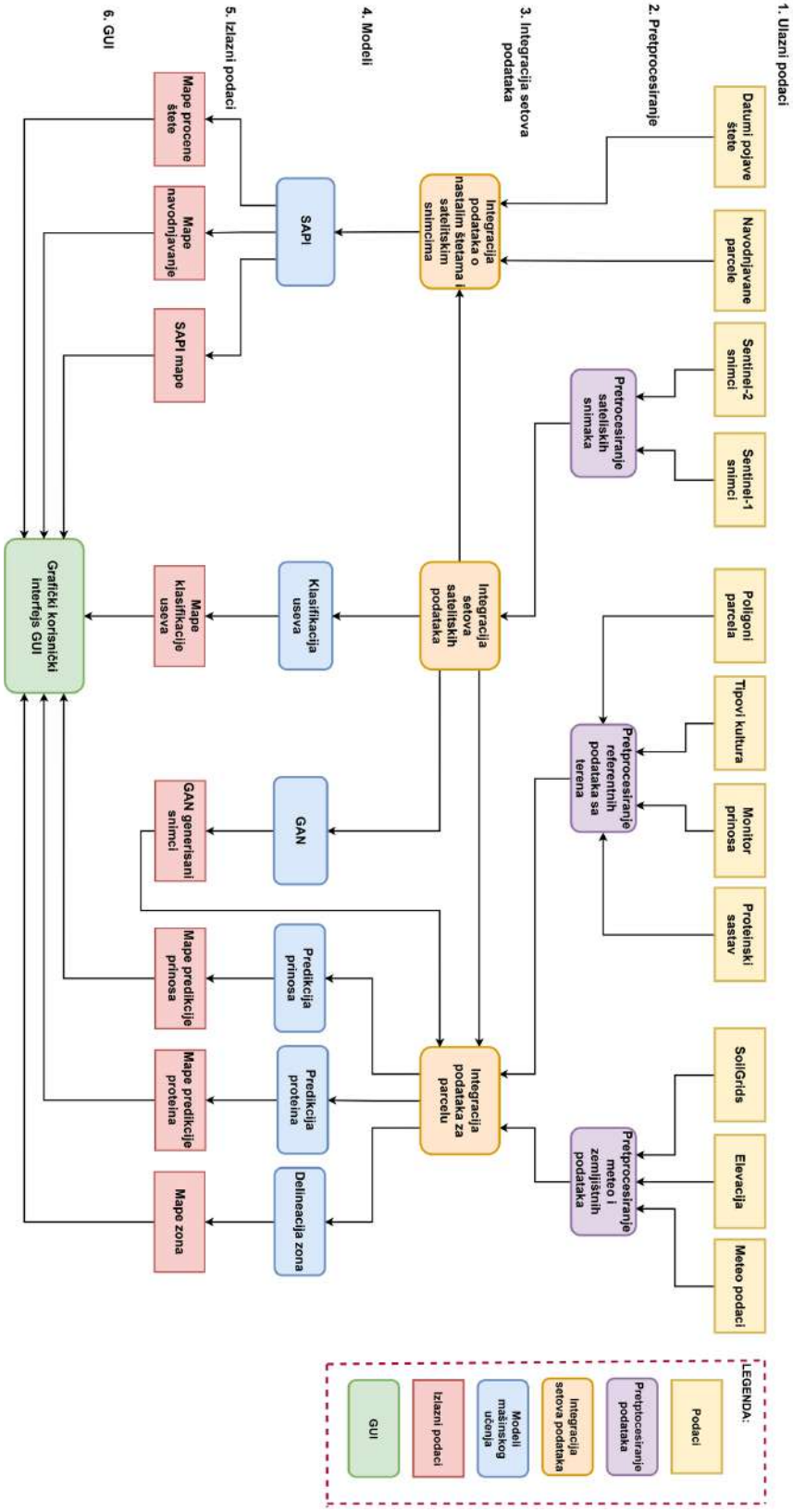
4.1 Arhitektura i metodološki okvir predloženog sistema 58
4.1.1 Ulazni podaci sistema 61
4.1.2 Predobrada podataka 62
4.1.3 Integracija skupova podataka 63
4.1.4 Modeli sistema 64
4.1.5 Izlazni podaci sistema 66
4.1.6 Grafički korisnički interfejs 67
4.2 Sistem za klasifikaciju useva 67
4.3 Sistem za predikciju prinosa i proteinskog sastava useva 68
4.4 Sistem za generisanje nedostajućih satelitskih podataka 69
4.5 Sistem za zoniranje parcele 70
4.6 Obrada vektorskih podataka 71
4.6.1 Obrada podataka sa monitora prinosa 71
4.7 Predobrada satelitskih slika 75
4.7.1 Izračunavanje vegetacionih indeksa 75
4.7.2 Predobrada satelitskih slika za CNN 78
4.7.3 Generisanje obeležja pomoću GAN 79

4.1 Arhitektura i metodološki okvir predloženog sistema

Dijagram na Slici 4.1 prikazuje arhitekturu predloženog sistema za monitoring poljoprivrednih površina, analizu stanja useva i generisanje operativnih izlaza zasnovanih na kombinaciji satelitskih, terenskih, zemljišnih i meteoroloških podataka. Sistem je koncipiran

kao višeslojni metodološki okvir u kome se heterogeni izvori podataka najpre obrađuju i prostorno-vremenski usklađuju, zatim integrišu u tematske skupove podataka prilagođene pojedinačnim zadacima modelovanja, a potom koriste kao ulaz u više modela mašinskog učenja. Krajnji rezultat sistema predstavljaju tematske mape i izlazi namenjeni operativnoj upotrebi kroz grafički korisnički interfejs.

Predloženi sistem nije zasnovan na jednom jedinstvenom modelu, već na međusobno povezanoj arhitekturi u kojoj različiti tipovi podataka i modeli imaju komplementarne uloge. U tom smislu, sistem objedinjuje procenu šteta, procenu verovatnoće navodnjavanja, klasifikaciju useva, delineaciju zona unutar parcela, predikciju prinosa i predikciju sadržaja proteina, pri čemu svi ovi zadaci dele zajedničku osnovu u vidu harmonizovanih satelitskih i pomoćnih skupova podataka.



Slika 4.1: Dijagram toka obrade podataka za analizu poljoprivrednih useva korišćenjem satelitskih podataka i metoda mašinskog učenja.

4.1.1 Ulazni podaci sistema

Ulazni sloj sistema čini skup međusobno komplementarnih podataka koji omogućavaju analizu useva iz više perspektiva: prostorne, vremenske, biofizičke i proizvodne. U skladu sa arhitekturom prikazanom na dijagramu, ulazni podaci obuhvataju sledeće grupe:

- datume i lokacije pojave štete,
- podatke o navodnjavanim parcelama,
- Sentinel-2 optičke snimke,
- Sentinel-1 radarske snimke,
- poligone parcela,
- podatke o tipovima kultura,
- podatke sa monitora prinosa,
- podatke o proteinskom sastavu,
- SoilGrids podatke,
- elevaciju,
- meteorološke podatke.

Sentinel-2 snimci predstavljaju osnovni izvor optičkih multispektralnih informacija o stanju vegetacije, refleksionim osobinama useva i njihovoj sezonskoj dinamici. Sentinel-1 radarski podaci obezbeđuju dodatnu, atmosferski nezavisnu informaciju o strukturi i stanju površine, čime se postiže veća robustnost sistema, posebno u periodima oblačnosti ili nedostatka kvalitetnih optičkih opservacija.

Poligoni parcela predstavljaju osnovni prostorni referentni okvir sistema. Oni omogućavaju agregaciju pikselnih informacija na nivou proizvodnih jedinica, formiranje atributa na nivou parcele i povezivanje satelitskih opservacija sa terenskim podacima. Podaci o tipovima kultura koriste se prvenstveno kao referentni skup za treniranje i validaciju modela klasifikacije useva, ali i kao pomoćni sloj za maskiranje i interpretaciju drugih izlaza.

Podaci sa monitora prinosa i podaci o proteinskom sastavu imaju ulogu ciljnih promenljivih u regresionim zadacima. Monitor prinosa obezbeđuje prostorno eksplicitnu referencu o količini prinosa, dok proteinski sastav predstavlja pokazatelj kvalitativnih

karakteristika proizvodnje. Ovi podaci povezuju satelitski signal sa stvarnim agronomskim ishodima na terenu.

SoilGrids podaci, elevacija i meteorološki podaci predstavljaju pomoćne prediktorske slojeve koji opisuju zemljišne, topografske i klimatske uslove proizvodnje. Njihovo uključivanje omogućava da modeli ne zavise isključivo od daljinske detekcije, već da uzmu u obzir i širi agroekološki kontekst.

Posebnu grupu ulaznih podataka čine datumi i lokacije pojave štete, kao i informacije o navodnjavanim parcelama. Ovi podaci imaju specifičnu ulogu u sistemu: ne služe kao standardni prediktori u svim modelima, već kao pomoćni ili referentni slojevi za interpretaciju satelitskih vremenskih serija i generisanje specifičnih izlaza, kao što su mape procene štete i mape navodnjavanja.

4.1.2 Predobrada podataka

S obzirom na to da ulazni podaci potiču iz različitih izvora i imaju različitu prostornu, vremensku i semantičku strukturu, sistem uključuje tri osnovna pretprocesna toka: pretprocesiranje satelitskih snimaka, pretprocesiranje referentnih podataka sa terena i pretprocesiranje meteoroloških i zemljišnih podataka.

4.1.2.1 Predobrada satelitskih snimaka

Predobrada satelitskih snimaka obuhvata standardne korake pripreme Sentinel-1 i Sentinel-2 podataka za dalju analizu. U ovu fazu spadaju geometrijsko i prostorno usklađivanje scena, koordinatnih sistem, prostorne rezolucije, vremensko organizovanje opservacija i priprema vremenskih serija sateliskih snimaka. Kod optičkih Sentinel-2 podataka poseban značaj ima identifikacija i uklanjanje nepouzdanih piksela, prvenstveno onih zahvaćenih oblacima i senkama.

Kako bi se smanjio problem nedostajućih ili nepouzdanih optičkih opservacija, u arhitekturu sistema uključen je i model zasnovan na generativnim adversarijalnim mrežama (GAN). Uloga GAN modula nije da predstavlja krajnji analitički izlaz, već da omogući dopunu, transformaciju ili rekonstrukciju satelitskih podataka, čime se formira konzistentniji satelitski skup pogodan za dalju analizu. Na taj način GAN predstavlja podršku integraciji satelitskih skupova podataka, posebno u uslovima oblačnosti ili nepotpunih vremenskih serija.

4.1.2.2 Predobrada referentnih podataka sa terena

Predobrada referentnih i terenskih podataka obuhvata obradu poligona parcela, tipova kultura, monitora prinosa i proteinskog sastava. U ovoj fazi vrši se proveravanje geometrijske ispravnosti, usklađivanje koordinatnih sistema, čišćenje atributnih tabela, uklanjanje nekonzistentnih ili nepotpunih zapisa i priprema slojeva za kasniju agregaciju na nivou parcele.

Podaci o tipovima kultura pripremaju se kao referentni skup za klasifikaciju useva. Podaci sa monitora prinosa i o proteinskom sastavu se prostorno povezuju sa parcelama i pripremaju za regresione modele. Na taj način referentni terenski slojevi postaju kompatibilni sa satelitskim i pomoćnim podacima i mogu se koristiti u zajedničkom toku obrade.

4.1.2.3 Predobrada meteoroloških i zemljišnih podataka

Meteorološki i zemljišni podaci prolaze zaseban tok pretprocesiranja. U ovoj fazi vrši se selekcija relevantnih promenljivih, prostorna interpolacija ili agregacija kada je to potrebno, usklađivanje sa prostornim jedinicama analize i vremenska sinhronizacija sa satelitskim podacima i terenskim informacijama. Cilj ovog koraka jeste da se SoilGrids slojevi, elevacija i meteorološki parametri prevedu u oblik pogodan za objedinjavanje sa ostalim izvorima podataka.

4.1.3 Integracija skupova podataka

Centralni deo sistema čini faza integracije skupova podataka, u kojoj se pretprocesirani ulazni izvori organizuju u tri tematski različita integrisana skupa. Ova podela je važna jer svi zadaci u sistemu ne koriste iste ulazne informacije, već su skupovi podataka formirani u skladu sa prirodom konkretnog problema.

4.1.3.1 Integracija podataka o nastalim štetama

Prvi integrisani skup nastaje povezivanjem datuma i lokacija pojave štete sa satelitskim vremenskim serijama. Ova integracija omogućava formiranje posebnog analitičkog skupa podataka namenjenog SAPI modulu. U ovom toku obrade satelitske opservacije se povezuju sa informacijama o prijavljenim događajima, čime se omogućava analiza odstupanja vegetacionog signala u odnosu na očekivani sezonski razvoj.

Ovakav skup podataka koristi se za tri međusobno povezana izlaza: mape procene štete, mape navodnjavanja i SAPI mape. Time SAPI nije ograničen samo na detekciju

oštećenja, već služi kao opšti modul za analizu promena i prostorne produktivnosti u odnosu na spoljne događaje i režim upravljanja parcelama.

4.1.3.2 Integracija skupova satelitskih podataka

Drugi integrisani skup predstavlja objedinjeni satelitski skup nastao spajanjem i harmonizacijom Sentinel-1, Sentinel-2 i i rekonstruisanih satelitskih snimaka. Ovaj skup je osnova za zadatke u kojima je primarni nosilac informacije upravo satelitski signal kroz vremensku seriju.

Iz ovog integrisanog satelitskog skupa izdvajaju se atributi za klasifikaciju useva, ali i za dalju upotrebu u ostalim modelima. Na dijagramu je prikazano da ovaj skup ne predstavlja samo ulaz za klasifikaciju useva, već i međukorak koji dalje doprinosi formiranju parcelno orijentisanog integrisanog skupa podataka.

4.1.3.3 Integracija podataka za parcelu

Treći integrisani skup formira se na nivou parcele i predstavlja ključni atributni sloj za regresione modele i delineaciju zona. U ovom koraku objedinjeni satelitski podaci povezuju se sa pretprocesiranim terenskim, zemljišnim, topografskim i meteorološkim podacima. Rezultat je jedinstveni skup atributa u kome svaka parcela, ili deo parcele, dobija odgovarajući skup deskriptora koji obuhvataju spektralno stanje useva, uslove zemljišta, reljef, vremenske uslove i raspoložive referentne terenske informacije.

Ovaj nivo integracije je suštinski jer omogućava prelaz sa pojedinačnih izvora podataka na objedinjeni model ulaza koji se može koristiti u više zadataka modelovanja. On predstavlja glavnu vezu između satelitskog toka obrade i agronomskih/regresionih zadataka.

4.1.4 Modeli sistema

Nakon formiranja odgovarajućih integrisanih skupova, sistem primenjuje više modela mašinskog učenja i analitičkih procedura, pri čemu svaki model ima jasno definisanu ulogu.

4.1.4.1 Satelitski indeks poljoprivredne produktivnosti (engl. Satellite Agricultural Productivity Index - SAPI)

SAPI modul koristi integrisani skup podataka o štetama i satelitskim snimcima za analizu sezonske dinamike vegetacionog signala i detekciju odstupanja u odnosu na očekivani razvoj useva. Na osnovu ovog modula generišu se:

- SAPI mape,
- SAPI mape procene navodnjavanja,
- SAPI mape procene štete.

Na taj način SAPI predstavlja centralni indeksni i analitički mehanizam za interpretaciju promena u produktivnosti i stanju useva, kako u kontekstu nepogoda, tako i u kontekstu režima procene navodnjavanja parcela.

4.1.4.2 Klasifikacija useva

Model klasifikacije useva koristi integrisani skup satelitskih podataka za određivanje tipova kultura na analiziranom području. U ovom slučaju ulaz čine prvenstveno višedatumski satelitski atributi, dok tipovi kultura iz referentnih podataka služe kao osnova za nadgledano treniranje i validaciju modela. Izlaz ovog modula predstavljaju mape klasifikacije useva, koje imaju važnu ulogu kako kao samostalan proizvod, tako i kao pomoćni sloj u ostalim analizama.

4.1.4.3 Generisanje sateliskih snimaka

Generativni modul ima specifičnu pomoćnu ulogu u sistemu. On nije primarno namenjen direktnom generisanju završnih mapa, već unapređenju kvaliteta i kontinuiteta satelitskih podataka. Kroz rekonstrukciju ili transformaciju satelitskih snimaka ovaj modul doprinosi stabilnijoj integraciji satelitskih skupova i povećava raspoloživost ulaznih podataka za dalje modele. U tom smislu, generativni modul je infrastrukturna komponenta sistema koja podržava ostale analitičke tokove.

4.1.4.4 Predikcija prinosa

Model za predikciju prinosa koristi integrisani skup podataka na nivou parcele. U ovom slučaju ulazni atributi obuhvataju satelitske deskriptore, zemljišne i topografske parametre, meteorološke promenljive i referentne podatke o proizvodnji. Cilj modela jeste da na osnovu ovih ulaza proceni prostornu raspodelu prinosa, odnosno da generiše mape predikcije prinosa.

4.1.4.5 Predikcija proteina

Slično prethodnom modelu, modul za predikciju proteina koristi integrisane podatke na nivou parcele kako bi procenio prostornu raspodelu proteinskog sastava useva. Ovaj zadatak proširuje sistem sa kvantitativne procene proizvodnje na kvalitativnu procenu

prinosa, što je od posebnog značaja u proizvodnim sistemima gde kvalitet ima podjednak značaj kao i količina.

4.1.4.6 Delineacija zona

Model za delineaciju zona koristi integrisane podatke na nivou parcele za izdvajanje unutrašnje prostorne heterogenosti unutar poljoprivrednih parcela. Cilj ovog modula jeste da identifikuje prostorno homogene zone koje mogu imati različit proizvodni potencijal, stanje vegetacije ili potrebe za upravljanjem. Kao izlaz generišu se mape zona, koje predstavljaju osnovu za preciznu poljoprivredu i diferencirano upravljanje parcelama.

4.1.5 Izlazni podaci sistema

Krajnji proizvodi sistema su organizovani kao skup tematskih izlaza koji proizlaze iz pojedinačnih modela. Prema arhitekturi prikazanoj na dijagramu, sistem generiše sledeće izlazne proizvode:

- mape procene štete,
- mape navodnjavanja,
- SAPI mape,
- mape klasifikacije useva,
- mape predikcije prinosa,
- mape predikcije proteina,
- mape zona.

Svi ovi proizvodi predstavljaju prostorno eksplicitne rezultate obrade i kompatibilni su sa standardnim GIS formatima i daljom analizom. Važno je naglasiti da su izlazi međusobno komplementarni: mape klasifikacije useva opisuju raspored kultura, SAPI proizvodi opisuju stanje i promene u vegetacionom razvoju, regresione karte daju kvantitativne i kvalitativne procene proizvodnje, dok mape zona predstavljaju osnovu za operativno upravljanje.

4.1.6 Grafički korisnički interfejs

Završni sloj sistema predstavlja grafički korisnički interfejs (GUI), u koji se integrišu svi glavni izlazni proizvodi. GUI ima funkciju operativnog sloja sistema, odnosno služi kao veza između analitičkog dela metodologije i njene praktične primene. Kroz ovaj interfejs korisnik može pregledati i upoređivati mape procene štete, navodnjavanja, SAPI vrednosti, klasifikacije useva, predikcije prinosa, predikcije proteina i delineacije zona.

Na taj način predloženi sistem ne ostaje samo na nivou metodološkog okvira za obradu podataka, već predstavlja osanovu za razvoj operativne platforme za podršku odlučivanju u poljoprivredi. Integracijom svih izlaza u jedinstveno korisničko okruženje omogućena je njihova interpretacija, poređenje i primena u realnim uslovima upravljanja poljoprivrednom proizvodnjom.

4.2 Sistem za klasifikaciju useva

Sistem za klasifikaciju useva predstavlja jednu od ključnih komponenti predloženog metodološkog okvira, jer obezbeđuje osnovne semantičke informacije neophodne za primenu naprednijih analitičkih i prediktivnih algoritama. Identifikacija tipa kulture na nivou poljoprivrednih parcela predstavlja preduslov za formiranje pouzdanih labela, koje se dalje koriste u procesima modelovanja prinosa, kvaliteta useva, detekcije oštećenja i definisanja zona upravljanja.

Klasifikacija useva se zasniva na analizi multitemporalnih i multispektralnih satelitskih podataka, pre svega sa Sentinel-2 misije, uz integraciju dodatnih pomoćnih informacija, kao što su karakteristike zemljišta i prostorni kontekst parcela. Primenom nadgledanih metoda mašinskog učenja omogućava se automatsko prepoznavanje tipova useva na velikim prostornim obuhvatima, uz zadržavanje visoke prostorne rezolucije i sezonske dinamike vegetacije.

Dobijene mape klasifikacije useva imaju višestruku ulogu u okviru sistema. Pre svega, one služe kao referentni sloj za selekciju parcela od interesa i filtriranje ulaznih podataka za dalje analize. Pored toga, rezultati klasifikacije predstavljaju osnovu za treniranje specijalizovanih regresionih modela, budući da se odnos između satelitskih indikatora, zemljišnih parametara i drugih promenljivih značajno razlikuje u zavisnosti od tipa kulture. Na taj način se omogućava razvoj specifičnih modela, čime se povećava tačnost i robusnost predikcija.

4.3 Sistem za predikciju prinosa i proteinskog sastava useva

Sistem za predikciju prinosa i proteinskog sastava useva razvijen u okviru ove disertacije predstavlja složen i integrisan metodološki okvir, koji kombinuje podatke daljinske detekcije, terenska merenja i napredne metode mašinskog učenja.

Osnovu sistema čini satelitski skup podataka, formiran primenom generativnih adversarijalnih mreža (GAN), koje se koriste za rekonstrukciju i homogenizaciju vremenskih serija optičkih snimaka. Na ovaj način se obezbeđuje kontinualna dostupnost relevantnih satelitskih informacija tokom cele vegetacione sezone, čime se značajno smanjuje negativan uticaj oblačnosti i nedostatka podataka. Dobijeni satelitski skup podataka predstavlja ulaz za dalju analizu i modelovanje prinosa.

U narednom koraku, primenom sistema za klasifikaciju useva, identifikuju se regioni i parcele od interesa, pri čemu se za svaku kulturu formira poseban skup podataka. Ovakav pristup omogućava razvoj specifičnog regresionog modela za svaku kulturu, imajući u vidu da se odnos između satelitskih indikatora, zemljišnih osobina i prinosa značajno razlikuje između različitih tipova useva. Za svaku kulturu od interesa trenira se zaseban regresioni model, čime se povećava tačnost i generalizacija predikcija.

Posebna pažnja posvećena je načinu podele podataka tokom obuke i validacije modela. Kako bi se izbegla prostorna i informacijska pristrasnost, svi pikseli koji pripadaju istoj poljoprivrednoj parceli grupišu se u isti podskup prilikom podele na skupove za obuku i testiranje. Na taj način se sprečava da podaci sa iste parcele budu istovremeno prisutni u različitim particijama, čime se obezbeđuje realistična procena performansi modela i njegova sposobnost generalizacije na nove parcele.

Ulazne ciljne promenljive za obuku regresionih modela dobijaju se iz vektorskih poligona prikupljenih pomoću monitora prinosa, koji sadrže informacije o količini prikupljenog useva na nivou parcela. Prostornim preklapanjem ovih poligona sa multispektralnim satelitskim snimcima Sentinel-2 misije vrši se distribucija vrednosti prinosa na nivo piksela, čime se omogućava formiranje trening skupa za prostorno eksplicitno modelovanje. Pored satelitskih obeležja, baza podataka uključuje i dodatne informacije o mehaničkim i hemijskim svojstvima zemljišta, preuzete iz odgovarajućih baza podataka.

Formirani integrisani skup podataka koristi se za obuku regresionih modela za predikciju prinosa i proteinskog sastava useva. Kao krajnji rezultat sistema generišu se prostorne mape prinosa i kvaliteta useva u prostornoj rezoluciji od 10 m, koje omogućavaju detaljnu analizu unutar-parcelne varijabilnosti i predstavljaju osnovu za donošenje odluka u okviru sistema precizne poljoprivrede.

4.4 Sistem za generisanje nedostajućih satelitskih podataka

Duboke neuronske mreže (DNN) beleže sve širu primenu u oblasti informacionih tehnologija, sa posebnim značajem u digitalnoj i preciznoj poljoprivredi You et al. 2020. Jedan od ključnih izazova u primeni satelitskih multispektralnih podataka predstavlja prisustvo oblačnosti, koja dovodi do gubitka velikog dela informacija i značajnog smanjenja dostupnosti validnih podataka tokom vegetacione sezone. U cilju prevazilaženja ovog problema, u okviru ove disertacije razvijen je sistem za generisanje nedostajućih podataka zasnovan na generativnim adversarijalnim mrežama (GAN), koji omogućava rekonstrukciju degradiranih ili nedostajućih piksela na multispektralnim snimcima Meraner et al. 2020.

Primena GAN modela u ovom kontekstu omogućava značajno proširenje dostupne baze podataka, smanjenje potrebe za odbacivanjem snimaka sa delimičnom oblačnošću i povećanje vremenske serije satelitskih posmatranja. Dodatno, ovakav pristup bi omogućio primenu modela za klasifikaciju useva i predikciju prinosa u narednim poljoprivrednim sezonama, bez potrebe za ponovnom obukom modela (engl. transfer learning). Time se obezbeđuje kontinuitet ulaznih podataka i stabilnost obeležja, nezavisno od sezonskih varijacija u dostupnosti optičkih snimaka.

Generativna adversarijalna mreža sastoji se od dve neuronske mreže koje se istovremeno obučavaju u okviru adversarijalnog procesa. Adversarijalni proces podrazumeva takmičarski način obuke u kome generator nastoji da proizvede što realističnije podatke, dok diskriminator pokušava da razlikuje generisane podatke od stvarnih. Tokom iterativne obuke, generator postepeno uči da generiše sve uverljivije primere, dok diskriminator istovremeno unapređuje sposobnost njihove detekcije. Prvu mrežu čini generator, čiji je zadatak da nauči raspodelu podataka na osnovu trening skupa i generiše sintetičke, ali realistične uzorke. Drugu mrežu čini diskriminator, koji ima ulogu da razlikuje originalne podatke od podataka generisanih od strane generatora. Cilj obuke generatora je da proizvede takve podatke koji maksimizuju verovatnoću greške diskriminatora, dok diskriminator teži suprotnom cilju, čime se postiže stabilna ravnoteža između dve mreže.

U predloženom sistemu generator je implementiran kao dekonvoluciona neuronska mreža zasnovana na U-Net arhitekturi, dok je diskriminator realizovan kao konvoluciona neuronska mreža tipa PatchGAN Lei Wang et al. 2019. Ovakva arhitektura omogućava očuvanje lokalne prostorne strukture i fine teksturalne detalje u procesu rekonstrukcije satelitskih snimaka.

Naučni doprinos ove disertacije ogleda se u primeni GAN modela za rekonstrukciju

multispektralnih Sentinel-2 snimaka na osnovu radarskih (SAR) snimaka sa Sentinel-1 satelita. Predloženi pristup koristi dva osnovna radarska kanala Sentinel-1 snimaka, uz dodatno formiranje kombinovanog kanala, kako bi se omogućila rekonstrukcija multispektralne slike sa dvanaest kanala u vidljivom, bliskom infracrvenom i infracrvenom spektru. Prednost korišćenja SAR podataka ogleda se u njihovoj nezavisnosti od atmosferskih uslova, čime se omogućava pouzdana rekonstrukcija i u uslovima potpune oblačnosti.

Početna faza realizacije sistema obuhvata selekciju odgovarajućih Sentinel-1 i Sentinel-2 snimaka koji pokrivaju region od interesa. Faza predobrade uključuje usklađivanje koordinatnih sistema, korekciju prostorne rezolucije i prostorno poravnanje snimaka. Nad SAR snimcima dodatno se primenjuju postupci terenske korekcije, radiometrijske kalibracije i filtriranja spekl šuma. Za obe vrste snimaka vrši se normalizacija obeležja, nakon čega se podaci dele na podregione manjih dimenzija, prilagođene ulaznom sloju neuronske mreže.

Detekcija oblačnosti i degradiranih piksela na multispektralnim Sentinel-2 snimcima sprovodi se korišćenjem dodatnog maskirnog kanala koji je sastavni deo standardnog proizvoda. Segmenti snimaka koji sadrže piksele prekrivene oblacima izdvajaju se u poseban skup podataka namenjen rekonstrukciji. GAN model se trenira isključivo na snimcima bez oblačnosti, dok se rekonstrukcija degradiranih segmenata vrši iterativno, u skladu sa prostornom raspodelom nedostajućih piksela.

Evaluacija kvaliteta rekonstrukcije sprovodi se na nivou pojedinačnih kanala korišćenjem metrika RMSE (engl. Root Mean Square Error) i SSIM (engl. Structural Similarity Index Measure), čime se obuhvataju i numerička tačnost i očuvanje strukturnih karakteristika slike. Rekonstruisani multispektralni snimci omogućavaju izračunavanje dodatnih ulaznih obeležja, kao što su vegetacioni indeksi, koji se dalje koriste u obuci modela mašinskog učenja za klasifikaciju useva, zoniranje parcela i predikciju prinosa.

4.5 Sistem za zoniranje parcele

Zemljište unutar iste poljoprivredne parcele često pokazuje izraženu prostornu heterogenost u pogledu fizičkih, hemijskih i bioloških osobina, što se direktno odražava na varijabilnost prinosa i efikasnost primenjenih agrotehničkih mera. Umesto uniformne primene agrotehničkih mera na celoj parceli, podela parcele na približno homogene zone omogućava implementaciju principa precizne poljoprivrede, kao što je varijabilna primena mineralnih đubriva, čime se postiže povećanje prinosa uz istovremeno smanjenje troškova proizvodnje i negativnog uticaja na životnu sredinu Basso et al. 2016.

Formiranje zona upravljanja predstavlja preduslov za sprovođenje sistema preporuke đubrenja i zasniva se na analizi satelitskih snimaka visoke prostorne rezolucije. U okviru predloženog sistema koriste se multispektralni snimci Sentinel-2 satelita, na osnovu kojih se izračunavaju vegetacioni indeksi i indeksi golog zemljišta, koji reflektuju prostornu varijabilnost stanja vegetacije i proizvodnog potencijala zemljišta. Ovi indeksi se koriste kao ulazna obeležja za proces zoniranja unutar parcele.

Zoniranje, odnosno grupisanje piksela sa sličnim karakteristikama, realizuje se primenom K-Means algoritma klasterizacije Hartigan and Wong 1979, pri čemu se broj klastera definiše kao ulazni parametar u skladu sa agronomskim zahtevima i prostornom kompleksnošću parcele. Rezultat početne klasterizacije dodatno se unapređuje primenom postupaka prostorne homogenizacije. Konkretno, dobijene oznake klastera se filtriraju median filtrom, čime se postiže veća prostorna konzistentnost zona, uklanjanje izolovanih piksela i formiranje regularnijih granica između zona.

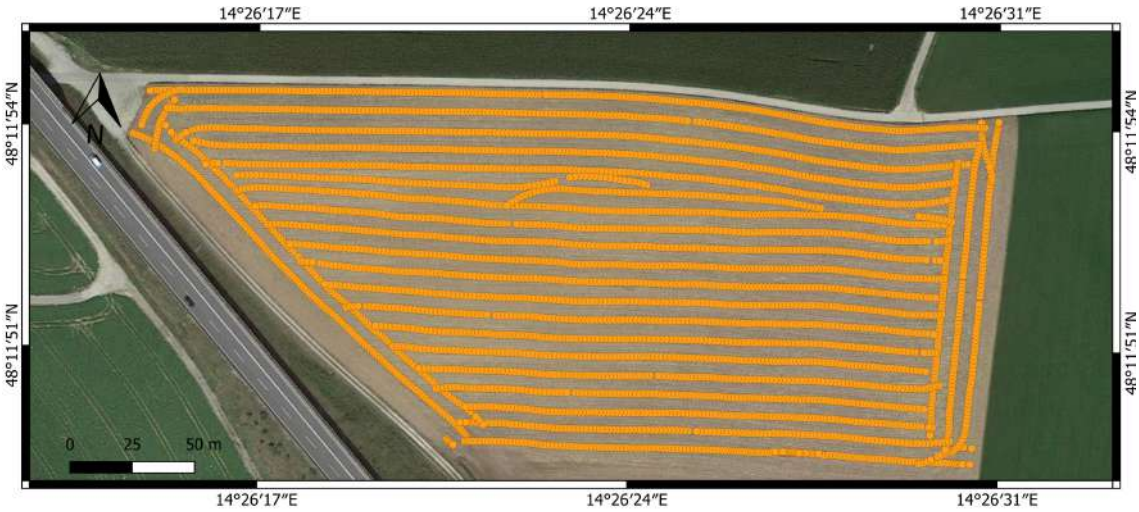
Nakon formiranja zona upravljanja, sledeći modul sistema obuhvata generisanje tačaka za uzorkovanje zemljišta, pri čemu se uzimaju u obzir prostorne granice definisanih zona. Inicijalne tačke za uzorkovanje određuju se kao centri mase pojedinačnih zona. Kako bi se obezbedilo da tačke uzorkovanja budu dovoljno udaljene od granica zona i potencijalnih graničnih anomalija, primenjuje se iterativni proces morfološke erozije. Ovim postupkom se dobijaju redukovane zone unutar kojih se definišu konačne, reprezentativne tačke za uzorkovanje zemljišta po zoni.

Podaci prikupljeni laboratorijskom analizom uzoraka zemljišta koriste se kao ulazna obeležja u algoritmu za preporuku đubrenja. Algoritam zasniva svoje odluke na unapred definisanim pragovima optimalnih koncentracija hraniva, pri čemu se za svaku zonu izračunava potrebna količina azota koja je neophodna za postizanje ciljnog nivoa plodnosti. Na ovaj način sistem omogućava prostorno diferenciranu i agronomski utemeljenu primenu đubriva, čime se ostvaruje efikasnije upravljanje resursima na nivou parcele.

4.6 Obrada vektorskih podataka

4.6.1 Obrada podataka sa monitora prinosa

Predložena PPI metoda Pejak et al. 2022 zasniva se na međusobnom povezivanju podataka monitora prinosa i satelitskih podataka. Podaci o prinosu sa kombajna dolaze u formi georeferenciranih tačkastih merenja (Slika 4.2). Merenja su vršena u redovnim vremenskim intervalima, ali pri promenljivoj brzini kombajna, što rezultira nejednako raspoređenim tačkama unutar redova. Pored toga, zbog nepravilnog oblika njiva, prolazi kombajna odstupaju od pravolinijske putanje.



Slika 4.2: Prikaz podataka sa monitora prinosa gde svaka tačka predstavlja merenu vrednost prinosa.

Originalni podaci sa uređaja za praćenje prinosa sadrže informacije o geografskoj dužini (x), geografskoj širini (y), širini zahvata (sw) (m), uglu (a) (rad), pređenoj udaljenosti (d) (m) i prinosu (y) (kg). Svi podaci korišćeni za rekonstrukciju zahvata prohoda kombajna i očitavanja senzora monitora prinosa prikazani su na Slici 4.3. Zeleni kvadrat označava Sentinel-2 piksel, dok ljubičasti pravougaonik predstavlja oblast koju je kombajn požnjeo u jednom intervalu, tj. oblast koja odgovara jednom merenju prinosa. Tačka P_c predstavlja položaj kombajna na kraju intervala skeniranja. Kombajn se kreće u smeru žute strelice i prešao je udaljenost d. Tačke P_1 do P_4, koje predstavljaju temena poligona, izražene su u sledećem obliku.

$$P_i = [x_i, y_i], \qquad i \in \{1, 2, 3, 4\} \tag{4.1}$$

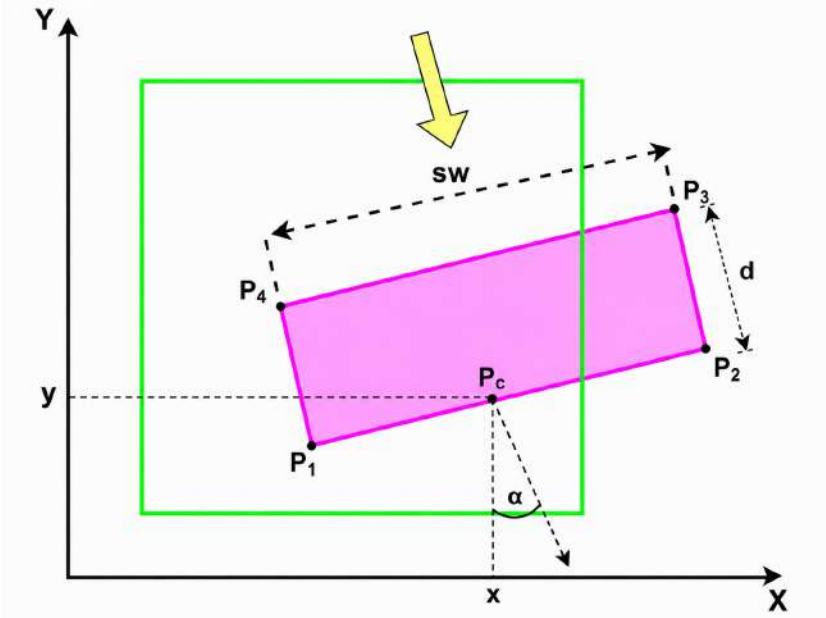
Njihove koordinate se izračunavaju korišćenjem prethodno navedenih parametara na sledeći način:

$$P_1 = \left[x - \frac{sw}{2}\cos(\alpha), y - \frac{sw}{2}\sin(\alpha)\right] \tag{4.2}$$

$$P_2 = \left[x + \frac{sw}{2}\cos(\alpha), y + \frac{sw}{2}\sin(\alpha)\right] \tag{4.3}$$

$$P_3 = \left[x + \frac{sw}{2}\cos(\alpha) - d\sin(\alpha), y + \frac{sw}{2}\sin(\alpha) + d\cos(\alpha)\right] \tag{4.4}$$

$$P_4 = \left[x - \frac{sw}{2}\cos(\alpha) - d\sin(\alpha), y - \frac{sw}{2}\sin(\alpha) + d\cos(\alpha)\right] \tag{4.5}$$



Slika 4.3: Grafički prikaz preklapanja poligona i piksela za PPI metodu.

Kako bi se dobila ukupna količina prinosa za određeni Sentinel-2 piksel, bilo je potrebno sabrati doprinos svakog senzorskog očitavanja po svakom pikselu. Postupak uključuje nekoliko koraka procesiranja. Prvo, za svako merenje prinosa crta se pravougaonik pomoću četiri temena. On odgovara površini sa koje je požnjevena kultura (Slika 4.4). Širina pravougaonika se beleži pomoću senzora kombajna i izvlači iz shapefile-a, zajedno sa orijentacijom kretanja.



Slika 4.4: Poligoni koji odgovaraju površini sa koje je soja požnjevena za svako pojedinačno merenje prinosa.

Zatim se preko toga postavlja mreža Sentinel-2 piksela, kao na Slici 4.5. Satelitska slika je isečena prema veličini parcele korišćenjem shapefile-a sa granicama parcele. Za svaki Sentinel-2 piksel, prinos je izračunat kao preklapanje između piksela i svih preklapajućih poligona prinosa. Na taj način je izvedena mapa prinosa za mrežu Sentinel-2 piksela, omogućavajući nesmetanu fuziju podataka. Procenat preklapanja (O_{p}) između poligona monitora prinosa (A_{ymp}) i poligona Sentinel-2 piksela (A_{pp}) izražava se na sledeći način:

$$O_{\mathrm{p}} = \frac{A_{\mathrm{ymp}}}{A_{\mathrm{pp}}} \tag{4.6}$$

Konačni izlaz modela izražen je kao količina prinosa u kg po pikselu ($\mathrm{Y_{pp}}$).

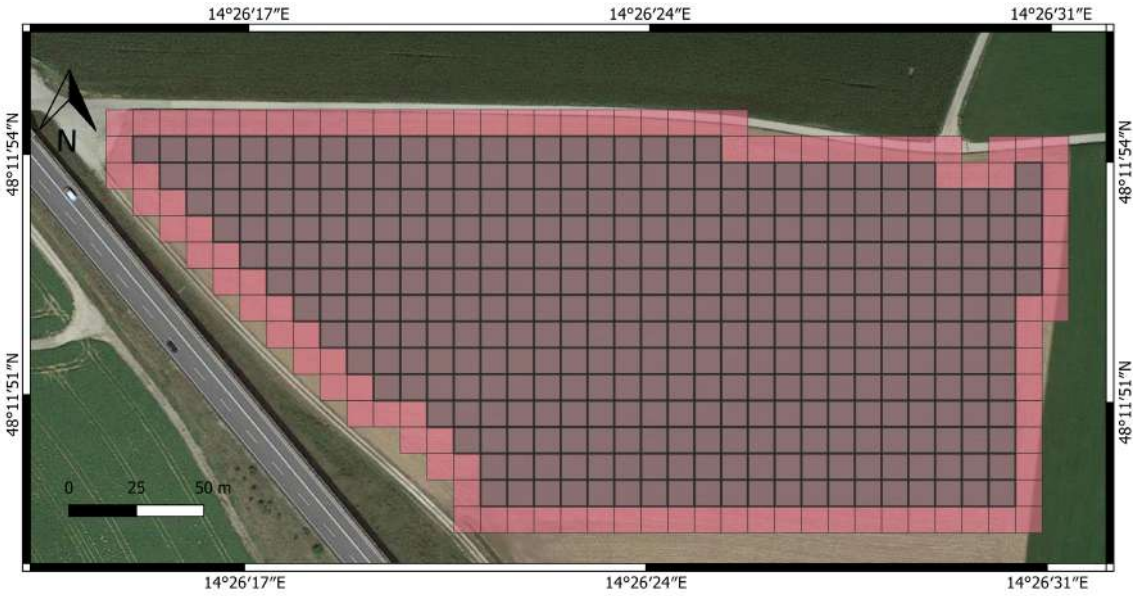
$$Y_{\mathrm{pp}} = \sum_{j=1}^{n} Y_{\mathrm{ymp}}(n) \cdot O_{\mathrm{p}}(n) \tag{4.7}$$

gde je n broj poligona koji se preklapaju sa pikselom, a Y_{ymp} doprinos poligona monitora prinosa satelitskom pikselu.

Jedini problemi su uočeni kod graničnih piksela, koji, pored useva, sadrže i puteve, zemlju, šume i druge delove zemljišta koji ne pripadaju parceli (Slika 4.6). Stoga su granični pikseli izostavljeni iz daljih analiza. Ovaj postupak je smanjio skup podataka uklanjanjem kontaminiranih i nepravilnih podataka.



Slika 4.5: Mreža Sentinel-2 piksela (sivi kvadranti) postavljena preko poligona prinosa (magenta).



Slika 4.6: Grafički prikaz mreže Sentinel-2 piksela. Granični pikseli isključeni iz analize označeni su crvenom bojom, dok su pikseli za dalju obradu označeni sivom bojom.

4.7 Predobrada satelitskih slika

4.7.1 Izračunavanje vegetacionih indeksa

Vegetacioni indeksi se izračunavaju različitim kombinacijama multispektralnih kanala satelita. Ovi indeksi su našli široku primenu u aplikacijama daljinske detekcije kao veoma pouzdan izvor informacija. Vegetativni indeksi mogu poslužiti kao indikatori metaboličkih procesa biljaka. Mnogo istraživanja dokazuje da vegetacioni indeksi mogu pružiti veoma korisne informacije kada je u pitanju predikcija prinosa. U ovom radu postavlja se hipoteza da li se ovi indeksi mogu koristiti za analizu podataka dobijenih pomoću metoda daljinske detekcije za klasifikaciju kultura. Biljke apsorbuju radijaciju u PAR (engl. photosyntetic active radiation) spektralnom regionu. Energiju koja je apsorbovana u ovom regionu biljke koriste u procesu fotosinteze. Ćelije lišća su evoluirale na takav način da rasipaju solarnu radijaciju iz blisko infracrvenog spektra. Ova energija predstavlja polovinu energije Sunčevog zračenja. Razlog je taj da energija koju ima foton na talasnim dužinama većim od 700nm nije dovoljna da bi se sintetizovao organski molekul. Apsorpcija ovih talasnih dužina dovela bi samo do pregrevanja biljaka koja dovodi do oštećenja tkiva biljaka. Na osnovu ovog se može zaključiti da će biljka biti relativno tamna u PAR delu spektra i svetla u blisko infracrvenom delu spektra. Za razliku od biljaka, oblaci i sneg su svetli u celom opsegu vidljivog zračenja, ali su tamni u bliskoinfracrvenom svetlu.

Vegetacioni indeksi koji su korišćeni kao dodatna obeležja u radu klasifikatora su: NDVI (engl. Normalized Difference Vegetation Index), EVI (engl. Enhanced Vegetation Index), EVI2 (engl. Enhanced Vegetation Index 2), ARVI (engl. Atmospherically Resistant Vegetation Index), GAVI (engl. engl. Green-Adjusted Vegetation Index), SAVI (engl. Soil-Adjusted Vegetation Index), VDVI (engl. Visible-Band Difference Vegetation Index), NDMI (engl. Normalized Difference Moisture Index), NLI (engl.Non-linear Index), NLI2 (engl. Non-linear Index 2), MNLI (engl. Modified Non-linear Index), MNLI2 (engl. Modified Non-linear Index 2).

Vegetacioni indeksi predstavljaju kvantitativne pokazatelje koji se koriste za analizu vegetacije na osnovu reflektovanih vrednosti svetlosti u različitim spektralnim opsezima. Oni se najčešće izračunavaju iz podataka dobijenih satelitskim, avionskim ili dronovskim snimcima i koriste se za procenu stanja vegetacije, njene gustine, fotosintetske aktivnosti i stresa usled suše ili bolesti.

- Normalized Difference Vegetation Index *(NDVI)* je jedan od najčešće korišćenih vegetacionih indeksa, koji kombinuje dve spektralne trake kako bi procenio gustinu zelene vegetacije i zdravlje biljaka Rouse et al. 1974; Pettorelli 2013.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{4.8}$$

- Enhanced Vegetation Index (EVI) je razvijen kako bi poboljšao osetljivost na vegetacioni signal u područjima sa gustom vegetacijom i unapredio praćenje vegetacije smanjenjem uticaja pozadinskog signala krošnje i atmosferskih uticaja A. Huete et al. 1997; Jiang et al. 2008.

$$EVI = 2.5 \cdot \frac{NIR - RED}{NIR + 6 \cdot RED - 7.5 \cdot BLUE + 1} \tag{4.9}$$

- Atmospherically Resistant Vegetation Index ($ARVI$) obezbeđuje proces samokorekcije kako bi korigovao radijaciju za atmosferski efekat na RED traku Kaufman and Tanre 1992.

$$ARVI = \frac{NIR - (RED - 1.7 \cdot (BLUE - RED))}{NIR + (RED - 1.7 \cdot (BLUE - RED))} \tag{4.10}$$

- Soil-Adjusted Vegetation Index ($SAVI$) je predstavljen kako bi se minimizirali uticaji osvetljenosti tla na spektralne vegetacijske indekse Alfredo R Huete 1988.

$$SAVI = \frac{NIR - RED}{NIR + RED + 0.5} \cdot (1 + L) \tag{4.11}$$

  U formuli L predstavlja faktor korekcije refleksije tla.

- Normalized Difference Vegetation Index Red-edge ($NDVIRE$) je modifikacija $NDVI$ indeksa u kojoj je RED traka zamenjena RedEdge trakom Anatoly A Gitelson and Mark N Merzlyak 1996; Tillack et al. 2014.

$$NDVIRE = \frac{NIR - RedEdge}{NIR + RedEdge} \tag{4.12}$$

- Visible Atmospherically Resistant Index ($VARI$) se koristi za procenu udela vegetacije sa niskom osetljivošću na atmosferske efekte Stow et al. 2005.

$$VARI = \frac{GREEN - RED}{GREEN + RED - BLUE} \tag{4.13}$$

- Normalized Difference Water Index ($NDWI$) se koristi za određivanje sadržaja vode u vegetaciji Gao 1996.

$$NDWI = \frac{NIR - SWIR2}{NIR + SWIR2} \tag{4.14}$$

- Modified Normalized Difference Water Index ($MNDWI$) je modifikovana verzija NDWI indeksa i takođe se koristi za detekciju sadržaja vode Xu 2006.

$$MNDWI = \frac{GREEN - SWIR2}{GREEN + SWIR2} \tag{4.15}$$

- Visible-Band Difference Vegetation Index ($VDVI$) igra važnu ulogu u izdvajanju vegetacionih informacija koristeći samo vidljive spektralne trake i koristi se za procenu stepena pokrivenosti vegetacijom Fuentes et al. 2012.

$$VDVI = \frac{2 \cdot GREEN - RED - BLUE}{2 \cdot GREEN + RED + BLUE} \tag{4.16}$$

- Non-linear Index (NLI) je razvijen korišćenjem interakcije između optičkog zračenja i biljne krošnje, kao i nekih rezultata analitičkih modela. Ovaj indeks može minimizirati efekte ometajućih faktora, azimuta gledanja i osvetljenosti tla Goel and Qin 1994.

$$NLI = \frac{NIR^2 - RED}{NIR^2 + RED} \tag{4.17}$$

- Modified Non-linear Index ($MNLI$) je modifikacija NLI indeksa sa dodatnim faktorom smanjenja uticaja tla Feng et al. 2019.

$$MNLI = \frac{(NIR^2 - RED) \cdot (1 + 0.5)}{NIR^2 + RED + 0.5} \tag{4.18}$$

- Normalised Multi-Band Drought Index ($NMDI$) je predložen za praćenje vlage u tlu i vegetaciji putem satelitskih podataka Lingli Wang and Qu 2007. Koristi se za detekciju suše.

$$NMDI = \frac{NIR2 - (SWIR2 - SWIR3)}{NIR2 + (SWIR2 - SWIR3)} \tag{4.19}$$

- Green Leaf Index (GLI) je važan pokazatelj fotosinteze krošnje, evapotranspiracije i kompeticije između biljnih kultura i korova Denison and Russotti 1997.

$$GLI = \frac{(GREEN - RED)(GREEN - BLUE)}{(2 \cdot GREEN) + RED + BLUE} \tag{4.20}$$

- Excess Green (ExG) vegetacioni indeks koristi se za izdvajanje skoro binarne slike intenziteta koja ocrtava oblast biljke od interesa Woebbecke et al. 1995.

$$ExG = 2 \cdot GREEN - RED - BLUE \tag{4.21}$$

- Color Index of Vegetation Extraction ($CIVE$) je kreiran da odvoji i naglasi zeleni biljni deo u odnosu na pozadinu Kataoka et al. 2003.

$$CIVE = 0.441 \cdot RED - 0.811 \cdot GREEN + 0.385 \cdot BLUE + 18.78745 \tag{4.22}$$

- Automated Water Extraction Index ($AWEI$) povećava kontrast između vode i drugih tamnih površina. Cilj AWEI indeksa je da maksimizira razlikovanje vodenih i nevodenih piksela primenom različitih koeficijenata Feyisa et al. 2014.

$$AWEI = 4 \cdot (GREEN - SWIR2) - (0.25 \cdot NIR + 2.75 \cdot SWIR3) \tag{4.23}$$

- Green-Red Vegetation Index ($GRVI$) je procenjen kao fenološki indikator zasnovan na višegodišnjim posmatranjima refleksije spektra i fenologije Motohka et al. 2010.

$$GRVI = \frac{GREEN - RED}{GREEN + RED} \tag{4.24}$$

4.7.2 Predobrada satelitskih slika za CNN

Predobrada satelitskih slika za konvolucione neuronske mreže (CNN) predstavlja ključni korak u obradi podataka daljinske detekcije, jer omogućava da se visokodimenzionalni i prostorno veliki ulazni podaci prilagode arhitekturi modela. Budući da CNN modeli zahtevaju ulaz fiksnih dimenzija (npr. 128×128 ili 256×256 piksela), velike satelitske scene se dele na manje segmente tzv. isečke (engl. patch), koji predstavljaju male

isečke slike sa zadržanim spektralnim kanalima (npr. RGB, NIR, NDVI). Ovi pečevi se generišu kliznim prozorom ili u regularnoj mreži sa preklapanjem kako bi se osigurala pokrivenost ivica i očuvala prostorna informacija. Svaki patch se može etiketirati na osnovu dominantne klase unutar tog isečka (npr. tip useva), čime se omogućava trening modela za klasifikaciju ili segmentaciju. Ova metoda ne samo da standardizuje ulaz u neuronsku mrežu, već i značajno uvećava broj uzoraka, što je posebno korisno kada je količina ručno anotiranih podataka ograničena. Tako pripremljeni pečevi omogućavaju efikasnu obuku CNN modela na složenim satelitskim scenama sa visokom prostornom rezolucijom.

4.7.3 Generisanje obeležja pomoću GAN

Generisanje obeležja u okviru ove disertacije zasniva se na razvoju i obuci generativnog modela za rekonstrukciju multispektralnih satelitskih snimaka degradiranih prisustvom oblačnosti. Cilj ovog postupka jeste dopuna vremenskih serija Sentinel-2 snimaka i obezbeđivanje konzistentnog skupa ulaznih obeležja za dalje zadatke klasifikacije useva, zoniranja parcela i predikcije prinosa. U tu svrhu primenjuje se generativni algortmi poput GAN, koji omogućava učenje složenih nelinearnih preslikavanja između radarskih i optičkih satelitskih podataka, uz eksplicitno modeliranje regiona prekrivenih oblacima.

4.7.3.1 Formulacija problema

Neka je $\mathbf{x}_{15} \in \mathbb{R}^{H\times W\times 15}$ ulazni tenzor koji sadrži radarske kanale Sentinel-1 (VV, VH), multispektralne kanale Sentinel-2 (B1–B12), kao i SCL (Scene Classification Layer). Neka je $\mathbf{y} \in \mathbb{R}^{H\times W\times 12}$ referentni Sentinel-2 snimak bez oblačnosti. U eksperimentalnom režimu obuke koristi se konkatenovani ulaz

$$\mathbf{x}_{27} = [\mathbf{x}_{15}; \mathbf{y}] \in \mathbb{R}^{H\times W\times 27}, \tag{4.25}$$

na osnovu kog generator G predviđa rekonstruisani multispektralni izlaz

$$\hat{\mathbf{y}} = G(\mathbf{x}_{27}), \tag{4.26}$$

pri čemu se optimizacija posebno usmerava na regione degradirane oblačnošću. U operativnoj primeni, referentni multispektralni snimak bez oblačnosti za isti datum nije direktno dostupan. Zbog toga se kao dodatni *vodič* (engl. guide) koristi prvi naredni vremenski susedni Sentinel-2 snimak bez oblačnosti, tj. najraniji datum $t^+ > t$ za koji je scena dovoljno čista. Na taj način, ulaz u generator u trenutku t formira se kao $\mathbf{x}_{27}(t) = [\mathbf{x}_{15}(t); \mathbf{y}(t^+)]$, gde $\mathbf{x}_{15}(t)$ sadrži podatke degradirane oblačnošću za datum

t, dok $\mathbf{y}(t^+)$ obezbeđuje spektralni i fenološki kontekst iz najbližeg dostupnog čistog posmatranja. Ovakva postavka omogućava operativno popunjavanje oblačnih regiona uz očuvanje konzistentnosti vremenske serije, uz napomenu da razlika $t^+ - t$ može uticati na tačnost rekonstrukcije u periodima brzih fenoloških promena.

Važno je razlikovati referentni snimak bez oblačnosti za isti datum t, $\mathbf{y}(t)$, i vodič (engl. guide) snimak $\mathbf{y}(t^+)$ koji potiče iz prvog narednog dovoljno čistog posmatranja. Referentni snimak $\mathbf{y}(t)$ koristi se isključivo tokom obuke i kvantitativne evaluacije, kako bi se funkcije gubitka mogle definisati u odnosu na poznatu „ground-truth“ sliku i kako bi se performanse mogle objektivno izmeriti.

4.7.3.2 Formiranje maske oblačnosti

Maska oblačnosti $\mathbf{M} \in \{0,1\}^{H \times W}$ generiše se iz SCL kanala Sentinel 2 snimaka mapiranjem klasa koje odgovaraju oblacima i senkama (3, 8, 9 i 10). Radi robustnosti, maska se proširuje blagom dilatacijom. Vrednosti $M(p) = 1$ označavaju piksele pod oblakom, dok $M(p) = 0$ označava čiste piksele.

4.7.3.3 Funkcija cilja

Ukupna funkcija gubitka generatora definisana je kao kombinacija rekonstrukcionih, regularizacionih i adversarijalnih članova:

$$\mathcal{L}_G = \lambda_{L1}\mathcal{L}^w_{L1} + \lambda_{SSIM}\mathcal{L}_{SSIM} + \lambda_{BL}\mathcal{L}_{BLUR} + \lambda_{ID}\mathcal{L}_{ID} + \lambda_{ADV}\mathcal{L}_{ADV} + \lambda_{FM}\mathcal{L}_{FM} + \lambda_{COL}\mathcal{L}_{COL}. \tag{4.27}$$

Ponderisani L1 gubitak. Rekonstrukcija se optimizuje ponderisanim L1 gubitkom koji favorizuje oblačne regione, ali zadržava nadzor i nad čistim pikselima:

$$\mathcal{L}^w_{L1} = \frac{\sum_{p\in\Omega} w(p)\,\|\hat{\mathbf{y}}(p) - \mathbf{y}(p)\|}{\sum_{p\in\Omega} w(p)}, \qquad w(p) = w_c\,M(p) + w_{cl}\,(1 - M(p)), \tag{4.28}$$

gde su w_c i w_{cl} težine za oblačne i čiste piksele, Ω skup svih piksela slike nad kojima se računa funkcija gubitka, p označava pojedinačni piksel, $\hat{\mathbf{y}}(p)$ predikovana vrednost modela u pikselu p, dok je $\mathbf{y}(p)$ referentna vrednost bez oblaka.

Structural Similarity Index Measure (SSIM) je mera strukturlne sličnosti. Koristi se za poređenje dve slike tako što ne meri samo razliku po pikselima, već poredi osvetljenje, kontrast i strukturu slike. Vrednost SSIM je najčešće u opsegu od 0 do 1, gde 1 znači da su slike veoma slične ili identične.

SSIM između rekonstruisane slike $\hat{\mathbf{y}}$ i referentne slike $\mathbf{y}$ definiše se kao:

$$SSIM(\hat{\mathbf{y}}, \mathbf{y}) = \frac{(2\mu_{\hat{y}}\mu_y + C_1)(2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}}^2 + \mu_y^2 + C_1)(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2)} \tag{4.29}$$

gde $\mu_{\hat{y}}$ i μ_y predstavljaju srednje vrednosti rekonstruisane i referentne slike, $\sigma_{\hat{y}}^2$ i σ_y^2 njihove varijanse, dok $\sigma_{\hat{y}y}$ označava kovarijansu između rekonstruisane i referentne slike. Konstante C_1 i C_2 uvode se radi numeričke stabilnosti, odnosno da bi se izbegle nestabilnosti u slučaju kada su vrednosti u imeniocu veoma male.

Na osnovu SSIM mere definiše se SSIM gubitak:

$$\mathcal{L}_{SSIM} = 1 - SSIM(\hat{\mathbf{y}}, \mathbf{y}) \tag{4.30}$$

Blur gubitak (niskofrekventna konzistentnost). Radi stabilizacije globalnog kolorita i niskofrekventnih komponenti, primenjuje se L1 gubitak nad zamućenim verzijama slika, gde $\mathcal{B}(\cdot)$ označava niskopropusno filtriranje sa uniformnim kernelom (engl. box blur):

$$\mathcal{L}_{BLUR} = \|\mathcal{B}(\hat{\mathbf{y}}) - \mathcal{B}(\mathbf{y})\|. \tag{4.31}$$

Identity gubitak nad čistim pikselima. Da bi se sprečile neželjene izmene u regionima bez oblačnosti, generator se penalizuje ako odstupa od oblačnog Sentinel-2 ulaza $\mathbf{x}^{S2}$ na čistim pikselima:

$$\mathcal{L}_{ID} = \frac{\sum_{p\in\Omega}(1 - M(p))\,\|\hat{\mathbf{y}}(p) - \mathbf{x}^{S2}(p)\|}{\sum_{p\in\Omega}(1 - M(p))}. \tag{4.32}$$

Adversarijalni hinge-loss. Diskriminator D se obučava hinge-loss funkcijom:

$$\mathcal{L}_D = \mathbb{E}\Big[\max(0, 1 - D([\mathbf{x}_{27}, \mathbf{y}]))\Big] + \mathbb{E}\Big[\max(0, 1 + D([\mathbf{x}_{27}, \hat{\mathbf{y}}]))\Big], \tag{4.33}$$

dok generator koristi:

$$\mathcal{L}_{ADV} = -\mathbb{E}\Big[D([\mathbf{x}_{27}, \hat{\mathbf{y}}])\Big]. \tag{4.34}$$

Feature matching. Za stabilizaciju adversarijalnog treninga koristi se feature matching na internim aktivacijama diskriminatora:

$$\mathcal{L}_{FM} = \sum_{\ell} \|f_\ell([\mathbf{x}_{27}, \hat{\mathbf{y}}]) - f_\ell([\mathbf{x}_{27}, \mathbf{y}])\|, \tag{4.35}$$

gde f_ℓ označava aktivacije odabranih slojeva diskriminatora.

Usklađivanje spektralnih statistika na oblačnim pikselima. Dodatno se uvodi uslov da predikcija i referenca imaju slične per-band statistike (srednju vrednost i standardnu devijaciju) nad oblačnim pikselima:

$$\mathcal{L}_{COL} = \|\mu_M(\hat{\mathbf{y}}) - \mu_M(\mathbf{y})\|_1 + \|\sigma_M(\hat{\mathbf{y}}) - \sigma_M(\mathbf{y})\|, \tag{4.36}$$

gde $\mu_M(\cdot)$ i $\sigma_M(\cdot)$ predstavljaju srednju vrednost i standardnu devijaciju računatu uz ponderisanje maskom M.

Warm-up i raspored težine adversarijalnog termina. Tokom prvih E_w epoha (warm-up) model se obučava bez adversarijalnog termina ($\lambda_{ADV} = 0$). Nakon toga, λ_{ADV} se postepeno povećava (ramp-up) od početne do maksimalne vrednosti, čime se stabilizuje trening i smanjuje rizik od artefakata u ranim fazama optimizacije.

4.7.3.4 Trening i validacija

Trening se sprovodi na nasumično izabranim pečevima veličine 128×128 piksela, uz primenu prostorne augmentacije (rotacije i preslikavanja). Optimizacija parametara generatora i diskriminatora vrši se Adam optimizatorom u TTUR režimu (engl. Two Time-Scale Update Rule). Validacija se sprovodi isključivo na uzorcima koji sadrže oblačne regione, pri čemu se izračunavaju cloud-specific L1, globalni L1 i SSIM metrički pokazatelji. Tokom treninga čuvaju se kontrolne tačke modela, uključujući poslednju i najbolju verziju modela definisanu minimalnom greškom u oblačnim regionima.

U okviru obuke primenjen je TTUR režim, pri čemu se generator G optimizuje većom stopom učenja u odnosu na diskriminator D ($\mathrm{LR}_G = 2 \cdot 10^{-4}$, $\mathrm{LR}_D = 1 \cdot 10^{-4}$), kako bi se obezbedila stabilnija dinamika adversarijalnog treniranja i smanjila verovatnoća dominacije diskriminatora u ranim fazama optimizacije. Validaciona procedura koristi ciljano uzorkovanje pečeva, gde se prihvataju samo oni pečevi koji sadrže minimalni udeo piksela prekrivenih oblačnošću, definisan pragom `VAL_MIN_MASK_FRAC`, čime se validacija eksplicitno usmerava na procenu sposobnosti modela da rekonstruiše degradirane regione, a ne da reprodukuje već postojeće čiste piksele. Tokom obuke čuvaju se kontrolne tačke modela (*last* i *best*), pri čemu se najbolji model selektuje na osnovu minimizacije metrike $L1_{cloud}$ na validacionom skupu, odnosno prosečne $L1$ greške izračunate isključivo nad pikselima označenim maskom oblačnosti.

Krajnji rezultat ovog procesa jeste generativni model sposoban da pouzdano rekonstruiše multispektralne Sentinel-2 snimke, očuva spektralnu verodostojnost i prostorne detalje, i obezbedi visokokvalitetna ulazna obeležja za dalje faze analize u okviru sistema precizne poljoprivrede.

4.7.3.5 Razdvajanje faza obuke i inferencije

Implementacija generativnog modela za rekonstrukciju satelitskih snimaka koncipirana je kroz jasno razdvajanje dve funkcionalno različite faze: (i) faze obuke (engl. training) i (ii) faze primene modela (engl. inference). Ovakvo razdvajanje predstavlja standardnu praksu u razvoju sistema zasnovanih na dubokom učenju i omogućava veću modularnost, reproducibilnost i operativnu primenljivost predloženog rešenja.

Skripta za obuku modela (training skripta) zadužena je za kompletan proces učenja parametara generativnog i diskriminativnog modela. Njene funkcionalnosti obuhvataju učitavanje i parsiranje trening skupa, formiranje ulazno–izlaznih parova, generisanje maski oblačnosti, ekstrakciju pečeva, primenu prostorne augmentacije, računanje višekomponentne funkcije greške, optimizaciju parametara modela, periodičnu validaciju i čuvanje kontrolnih tačaka (engl. checkpoints). Ova faza je računski najzahtevnija i koristi se isključivo tokom razvoja i evaluacije modela.

Nasuprot tome, skripta za inferenciju koristi prethodno obučeni model u režimu bez učenja i namenjena je operativnoj primeni sistema. Njena uloga je rekonstrukcija oblačnih ili degradiranih multispektralnih Sentinel-2 snimaka nad novim, prethodno neviđenim podacima. Inference skripta implementira deterministički tok obrade koji obuhvata učitavanje satelitskih snimaka, prostorno poravnanje, segmentaciju na podregione, primenu generatora nad oblačnim regionima i spajanje rekonstruisanih segmenata u celovitu multispektralnu sliku.

Razdvajanje trening i inference skripti omogućava višestruke prednosti. Prvo, obezbeđuje potpunu reproduktivnost naučnih eksperimenata, jer se faza obuke odvija u kontrolisanom okruženju sa fiksnim skupom hiperparametara. Drugo, omogućava efikasnu primenu modela u operativnim sistemima, gde nije neophodna prisutnost diskriminatora niti dodatnih komponenti potrebnih tokom treninga. Treće, ovakav dizajn omogućava ponovnu upotrebu istog obučnog modela u više vegetacionih sezona i nad različitim regionima, bez potrebe za ponovnom obukom, čime se značajno smanjuju računski zahtevi i vreme obrade.

4.7.3.6 Inferencija i evaluacija modela

Neka je t datum sa oblačnošću, a t^+ prvi naredni datum sa dovoljno čistom scenom. Tada se vodič definiše kao

$$t^+ = \min\{\tau \,:\, \tau > t \;\wedge\; q(\tau) \geq \theta\}, \tag{4.37}$$

gde $q(\tau)$ predstavlja meru čistoće scene (npr. udeo čistih piksela iz SCL), a θ prag kvaliteta. Ulaz generatora u trenutku t postaje

$$\mathbf{x}_{27}(t) = [\mathbf{x}_{15}(t); \mathbf{y}(t^+)], \qquad \hat{\mathbf{y}}(t) = G(\mathbf{x}_{27}(t)). \tag{4.38}$$

Konačni izlazni snimak dobija se spajanjem predikcije i originalnog oblačnog snimka pomoću maske oblačnosti:

$$\mathbf{y}_{out}(t) = \mathbf{M}(t) \odot \hat{\mathbf{y}}(t) + \Big(1 - \mathbf{M}(t)\Big) \odot \mathbf{x}^{S2}(t), \tag{4.39}$$

čime se generator primenjuje isključivo na piksele degradirane oblačnošću.

Nakon završetka obuke, model se primenjuje u režimu inferencije koristeći isključivo generator G, pri čemu su parametri fiksirani, a računanje gradijenata isključeno. U ovom delu razmatra se evaluacioni režim inferencije nad uparenim parovima snimaka, koji omogućava direktno poređenje rekonstruisanih i referentnih multispektralnih slika i kvantitativnu procenu performansi modela. Inferencija je implementirana kao evaluaciona skripta nad skupom uparenih parova $\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^{N}$, gde $\mathbf{x}$ predstavlja ulazni tenzor (multisenzorski podaci sa maskom oblačnosti), a $\mathbf{y}$ referentni multispektralni snimak bez oblačnosti.

Patch-wise inferencija i spajanje preklapajućih pečeva. S obzirom na veliku prostornu dimenziju satelitskih scena, inferencija se sprovodi nad preklapajućim pečevima dimenzije $T \times T$ sa korakom $s = T - o$, gde o označava preklapanje. Za svaki peč (i, j) dobija se predikcija $\hat{\mathbf{y}}_{ij} = G(\mathbf{x}_{27,ij}(t))$. Konačna slika dobija se ponderisanim spajanjem pečeva pomoću težinske maske $\mathbf{W}$ (Hann prozor), čime se izbegavaju artefakti na granicama pečeva:

$$\hat{\mathbf{Y}}(p) = \frac{\sum\limits_{(i,j)\in\mathcal{T}(p)} \mathbf{W}_{ij}(p)\, \hat{\mathbf{y}}_{ij}(p)}{\sum\limits_{(i,j)\in\mathcal{T}(p)} \mathbf{W}_{ij}(p) + \epsilon}, \tag{4.40}$$

gde p označava piksel, $\mathcal{T}(p)$ skup svih pečeva koji pokrivaju piksel p, a ϵ je mala konstanta radi numeričke stabilnosti.

Test-time augmentation (TTA). U cilju povećanja stabilnosti predikcije, koristi se test-time augmentacija (rotacije i preslikavanja). Neka $\mathcal{A} = \{a_k\}_{k=1}^{K}$ predstavlja skup transformacija, a a_k^{-1} odgovarajuće inverzne transformacije. Tada se konačna predikcija dobija prosečenjem:

$$\hat{\mathbf{y}}(t) = \frac{1}{K} \sum_{k=1}^{K} a_k^{-1}\Big(G\Big(a_k(\mathbf{x}_{27}(t))\Big)\Big). \tag{4.41}$$

Opciona linearna per-band korekcija na čistim pikselima. Kao dodatnu nakdnadnu obradu, moguće je izvršiti po kanalima linearnu korekciju spektralnih vrednosti na pikselima bez oblačnosti. Za svaki kanal b procenjuju se koeficijenti (α_b, β_b) metodom najmanjih kvadrata nad skupom čistih piksela Ω_{clear}:

$$(\alpha_b, \beta_b) = \arg\min_{\alpha,\beta} \sum_{p \in \Omega_{clear}} \Big(\alpha\, \hat{y}_b(p) + \beta - x_b^{S2}(p)\Big)^2, \tag{4.42}$$

nakon čega se predikcija koriguje kao $\hat{y}_b^{corr}(p) = \mathrm{clip}(\alpha_b \hat{y}_b(p) + \beta_b, 0, 1)$. Ovaj korak se koristi isključivo kao opcioni modul za poboljšanje globalne konzistentnosti kolorita, bez uticaja na osnovni tok modela.

Metrike. Kvalitet rekonstrukcije procenjuje se korišćenjem globalne $L1$ greške i $L1$ greške u oblačnim regionima definisanim maskom $\mathbf{M}$:

$$L1_{all}(t) = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|\hat{\mathbf{y}}(t,p) - \mathbf{y}(t,p)\|\,, \tag{4.43}$$

$$L1_{cloud}(t) = \frac{1}{|\Omega_c(t)|} \sum_{p \in \Omega} M(t,p)\, \|\hat{\mathbf{y}}(t,p) - \mathbf{y}(t,p)\|\,, \qquad |\Omega_c(t)| = \sum_{p \in \Omega} M(t,p). \tag{4.44}$$

Pored toga, računa se i strukturna sličnost $SSIM(\hat{\mathbf{y}}, \mathbf{y})$, kako bi se ocenilo očuvanje lokalnih tekstura i prostora.

Poglavlje 5

Rekonstrukcija nedostajućih satelitskih opservacija

Sadržaj

5.1 GAN generisane satelitske slike 86
5.1.1 Formiranje skupa podataka za obuku GAN 86
5.1.2 Upotreba GAN modela u svrhu generisanja optičkih satelitskih snimaka 88
5.1.3 Rezultati rekonstrukcije Sentinel-2 snimaka 88

5.1 GAN generisane satelitske slike

5.1.1 Formiranje skupa podataka za obuku GAN

Skup podataka za obuku i evaluaciju generativnog modela formiran je u vidu uparenih satelitskih snimaka, pri čemu svaki par sadrži jedan ulazni snimak koji sadrži prisustvom oblačnosti i jedan odgovarajući referentni čisti snimak. Na ovaj način formiran je skup od ukupno 42 uparena uzorka, raspoređena kroz više sezona i godina (2023, 2024 i 2025). Ovakva organizacija podataka omogućava obuku modela na vremenski serijama satelitskih snimaka koje obuhvataju različitim fenološke faze biljaka i atmosferske uslove.

Formiranje uparenih skupova podataka za obuku i evaluaciju generativnog modela izvršeno je automatskim povezivanjem oblačnih i čistih Sentinel-2 snimaka na osnovu vremenske bliskosti datuma akvizicije. Polazni skup činile su dve odvojene kolekcije snimaka: (i) skup oblačnih scena i (ii) skup referentnih scena bez oblačnosti. Za svaku analiziranu godinu ove kolekcije bile su organizovane u posebne skupove podataka da bi se na kraju za potrebno obuke modela objedinile u zajednički skup.

Postupak uparivanja zasnivao se na izdvajanju datuma akvizicije iz naziva fajlova. Datum je identifikovan pomoću obrasca YYYYMMDD, koji je automatski prepoznat unutar naziva svakog snimka. Nakon toga, svi oblačni i čisti snimci su indeksirani i

sortirani hronološki prema datumu akvizicije.

Za svaki oblačni snimak tražen je vremenski najbliži čisti snimak iz referentnog skupa. Pretraga je realizovana nad sortiranim nizom datuma, pri čemu su kao kandidati posmatrani najbliži prethodni i najbliži naredni datum u referentnom skupu. Od ova dva kandidata biran je onaj čiji je apsolutni vremenski razmak u odnosu na oblačni snimak minimalan. Na taj način je za svaki oblačni snimak definisan jedinstveni referentni par.

Kako bi se izbeglo uparivanje vremenski suviše udaljenih scena, uvedeno je ograničenje maksimalnog dozvoljenog vremenskog pomaka, definisano parametrom koji definiše maksimalni broj dana. U konkretnom eksperimentu korišćena je vrednost od $\pm$ 15 dana, što znači da je oblačni snimak mogao biti uparen samo sa onim čistim snimkom čiji je datum akvizicije udaljen najviše 15 dana. Ukoliko nijedan referentni snimak nije zadovoljavao ovaj uslov, takva scena nije uključena u dalju obradu.

U slučaju da je za isti referentni datum postojalo više čistih snimaka, za uparivanje je korišćen prvi snimak iz liste pridružene tom datumu. Ovakav redosled izabran je kako bi bio usklađen sa daljim koracima obuke modela, u kojima se oblačni snimak koristi kao ulaz, a čisti snimak kao ciljna referenca. Snimak za koji ne postoji odgovarajaću referentni snimak bez oblačnosti se izostavlja i dalje analize.

Predloženi postupak uparivanja omogućava formiranje vremenski konzistentnog skupa podataka za rekonstrukciju optičkih satelitskih scena, pri čemu se minimizuje vremenski razmak između ulazne i referentne slike. Na ovaj način obezbeđuje se da uparene scene budu fenološki i sezonski što sličnije, čime se poboljšavaju uslovi za obuku generativnog modela i smanjuje rizik od uvođenja veštačkih razlika koje nisu posledica oblačnosti, već prirodne dinamike vegetacije.

U okviru ove disertacije, GAN rekonstruisane satelitske slike korišćene su kao dopunski izvor informacija za datume kod kojih su originalne Sentinel-2 opservacije bile degradirane prisustvom oblaka. Njihova prediktivna korisnost ispitana je zasebno na skupu rekonstruisanih oblačnih datuma, kao i u kombinaciji sa dostupnim realnim snimcima bez značajne oblačnosti. Ulazni podaci preuzeti su korišćenjem platforme Google Earth Engine (GEE), pri čemu su korišćeni multispektralni Sentinel-2 snimci i radarski Sentinel-1 snimci za odgovarajuće regione od interesa.

Kao osnovni ulaz u generativni model korišćeni su upareni skupovi podataka koji obuhvataju: (i) Sentinel-1 radarske kanale VV i VH, (ii) Sentinel-2 multispektralne kanale B1–B12, kao i (iii) Scene Classification Layer (SCL), koji je korišćen za identifikaciju oblačnih i čistih regiona.

Na ovaj način formiran je ulazni tenzor koji sadrži multisenzorske informacije o istoj sceni, pri čemu radarski podaci obezbeđuju dodatni signal nezavisan od atmosferskih uslova.

5.1.2 Upotreba GAN modela u svrhu generisanja optičkih satelitskih snimaka

Za potrebe eksperimenata formirana su tri režima ulaznih podataka: (i) realni Sentinel-2 snimci bez značajne oblačnosti, (ii) GAN rekonstruisani snimci za datume kod kojih su originalne Sentinel-2 opservacije bile degradirane oblačnošću i (iii) kombinovani skup.

U kombinovanom režimu formirana je jedinstvena vremenska serija u kojoj je za datume sa raspoloživim čistim Sentinel-2 snimkom korišćen originalni snimak, dok je za datume degradirane oblačnošću korišćena odgovarajuća GAN rekonstrukcija. Na taj način isti datum nije uključen dva puta.

U sezoni 2018. bilo je dostupno 12 čistih realnih i 5 oblačnih datuma, dok je u sezoni 2019. bilo dostupno 7 čistih realnih i 10 oblačnih datuma. GAN režim zato je obuhvatao 5 rekonstruisanih datuma u 2018. i 10 rekonstruisanih datuma u 2019. godini, dok je kombinovani režim u obe godine obuhvatao ukupno 17 datuma.

Metodološki, generativni model zasnovan je na GAN arhitekturi Goodfellow et al. 2014 sa generatorom tipa U-Net Ronneberger et al. 2015 i diskriminatorom tipa PatchGAN Isola et al. 2017. GAN model se sastoji od dve neuronske mreže. Jedna mreža čini generator, a druga diskriminator. Generator u GAN modelu služi da, na osnovu dostupnih ulaznih multispektralnih podataka, rekonstruiše multispektralni Sentinel-2 snimak u regionima degradiranim prisustvom oblaka, dok je diskriminator imao zadatak da razlikuje rekonstruisane od originalnih referentnih snimaka. Tokom obuke modela korišćena je *cloud-aware* funkcija cilja, koja kombinuje rekonstrukcione i strukturne greške sa adversarijalnim učenjem, kako bi se istovremeno očuvala spektralna verodostojnost i prostorna konzistentnost rekonstruisanih slika.

Evaluacija kvaliteta generisanih satelitskih snimaka vršena je korišćenjem metrika Structural Similarity Index Measure (SSIM) i Root Mean Square Error (RMSE), računatih u odnosu na referentne čiste Sentinel-2 snimke. SSIM je korišćen za procenu očuvanosti strukturalnih i teksturalnih karakteristika rekonstruisanih slika, dok je RMSE korišćen za kvantifikaciju apsolutnog odstupanja između generisanih i referentnih vrednosti po pikselu. Pored globalne greške, posebna pažnja posvećena je evaluaciji u oblačnim regionima, s obzirom na to da upravo ti delovi predstavljaju primarni cilj rekonstrukcije.

5.1.3 Rezultati rekonstrukcije Sentinel-2 snimaka

Rezultati evaluacije GAN rekonstrukcije Sentinel-2 snimaka za sezonu 2018 i 2019 pokazuju da model postiže stabilan nivo performansi, uz uočljive razlike između dve

vremenske serije. U tabeli za 2019. godinu prosečna vrednost metrike $RMSE_{_ceo_snimak}$ iznosi 0.0165, prosečna vrednost $RMSE_{_oblak}$ 0.0124, dok prosečni SSIM dostiže 0.9151. Ove vrednosti ukazuju na visoku strukturnu sličnost između rekonstruisanih i referentnih slika, kao i na relativno malu apsolutnu grešku rekonstrukcije, posebno u oblačnim regionima. Dobijeni SSIM veći od 0.91 potvrđuje da model uspešno očuvava prostorne i teksturalne karakteristike scene, dok niske vrednosti $RMSE$ greške ukazuju na dobru spektralnu konzistentnost rekonstruisanih piksela.

Tabela 5.1: Rezultati evaluacije GAN rekonstrukcije Sentinel-2 snimaka po datumima za sezonu 2018

Datum	**RMSE_ceo_snimak**	**RMSE_oblak**	**SSIM**
2018-05-10	0.035609	0.038207	0.824446
2018-05-30	0.018342	0.016066	0.904701
2018-06-04	0.017797	0.014020	0.903834
2018-07-14	0.025381	0.027365	0.879937
2018-07-24	0.026078	0.031985	0.881016
Prosek	**0.024641**	**0.025529**	**0.878787**

Tabela 5.2: Rezultati evaluacije GAN rekonstrukcije Sentinel-2 snimaka po datumima

Datum	**RMSE_ceo_snimak**	**RMSE_oblak**	**SSIM**
2019-05-10	0.015868	0.013413	0.912894
2019-05-20	0.014547	0.011269	0.913561
2019-05-25	0.014315	0.011501	0.917270
2019-06-04	0.016191	0.013426	0.913867
2019-06-29	0.016843	0.009720	0.913286
2019-07-09	0.016551	0.013552	0.916110
2019-07-19	0.015970	0.010378	0.916129
2019-07-24	0.015730	0.012296	0.915430
2019-08-23	0.016672	0.014541	0.922180
2019-09-27	0.022397	0.013467	0.910202
Prosek	**0.016508**	**0.012356**	**0.915093**

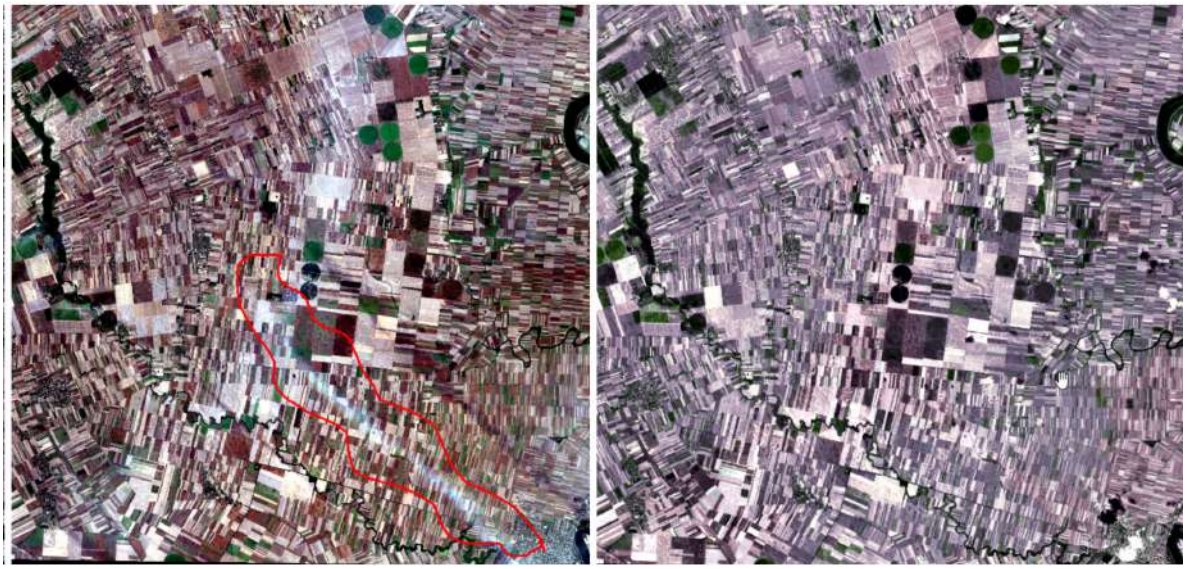
Nasuprot tome, rezultati za sezonu 2018. pokazuju slabije performanse u odnosu na 2019. godinu. Prosečna vrednost $RMSE_{_ceo_snimak}$ iznosi 0.0246, prosečna vrednost $RMSE_{_oblak}$ 0.0255, dok prosečni SSIM dostiže 0.8788. Iako i ove vrednosti potvrđuju da model ostvaruje upotrebljiv nivo rekonstrukcije, uočljivo je da su greške veće, a

strukturna sličnost niža nego u 2019. godini. Posebno se izdvaja datum 2018-05-10, za koji su zabeležene najveće vrednosti $RMSE$ greške i najniži SSIM, što ukazuje na složeniju scenu za rekonstrukciju, verovatno usled većeg stepena degradacije, nepovoljnijih fenoloških uslova ili većeg odstupanja između radarskog i optičkog signala u tom periodu.

Poređenjem dve sezone može se zaključiti da GAN model postiže bolje rezultate u 2019. godini, kako u pogledu spektralne tačnosti, tako i u pogledu očuvanja strukturnih karakteristika slike. U proseku, 2019. godina pokazuje niže vrednosti i za $RMSE_{_ceo_snimak}$ i za $RMSE_{_oblak}$, kao i višu vrednost SSIM, što ukazuje na stabilniju i kvalitetniju rekonstrukciju. Ovakva razlika može biti posledica povoljnijih ulaznih podataka, manjeg stepena degradacije scena, boljeg vremenskog usklađivanja Sentinel-1 i Sentinel-2 snimaka, ili jednostavno manje složenog fenološkog i atmosferskog konteksta u izabranim datumima.

Ukupno posmatrano, obe tabele potvrđuju da generativni model može uspešno da dopuni vremenske serije optičkih satelitskih podataka u uslovima oblačnosti, pri čemu se kvalitet rekonstrukcije razlikuje u zavisnosti od sezone i datuma posmatranja. Dobijeni rezultati pokazuju da GAN rekonstruisane slike mogu predstavljati dopunski izvor informacija u vremenskim serijama satelitskih posmatranja, pri čemu se postiže zadovoljavajuća strukturna sličnost sa referentnim snimcima i prihvatljiv nivo rekonstrukcione greške. Njihova korisnost u daljim zadacima dodatno je ispitana kroz modele predikcije prinosa, korišćenjem realnih, GAN rekonstruisanih i kombinovanih ulaznih podataka.

Dobijeni rezultati pokazuju da GAN generisane slike mogu uspešno da dopune vremenske serije satelitskih posmatranja, pri čemu se postiže zadovoljavajuća strukturna sličnost sa referentnim snimcima i prihvatljiv nivo rekonstrukcione greške. Na taj način generisane slike ne predstavljaju zamenu za originalna posmatranja, već dopunski izvor informacija koji povećava gustinu vremenske serije i poboljšava uslove za obuku modela mašinskog učenja u zadacima predikcije prinosa.



Slika 5.1: Vizuelno poređenje originalnog Sentinel-2 snimka (slika levo) i rekonstruisanog snimka (slika desno 06.08.2025.) iz test skupa podataka. Oblast označena crvenom bojom je region koji je bio zahvaćen oblacima

Poglavlje 6

Klasifikacija useva

Sadržaj

6.1 Klasifikacija useva 92
6.1.1 Ulazni podaci modela 94
6.1.2 Obuka klasifikatora 95
6.1.3 Generisanje finalnih mapa klasifikacije useva 96
6.1.4 Metodološke prednosti i ograničenja 98

6.1 Klasifikacija useva

Mape klasifikacije poljoprivrednih useva zasnovane su na vremenskim serijama multispektralnih Sentinel-2 satelitskih snimaka, koji obezbeđuju visoku prostornu (10–20 m) i vremensku rezoluciju pogodnu za praćenje fenološkog razvoja useva tokom vegetacione sezone. Iz originalnih spektralnih kanala izračunat je skup standardnih vegetacionih indeksa, koji opisuju biofizičko stanje i dinamiku vegetacije, čime je unapređena diskriminacija različitih tipova useva u prostoru i vremenu.

Referentni (ground truth) podaci prikupljani su u periodu od 2016. do 2025. godine korišćenjem mobilne aplikacije *FieldMap*, razvijene od strane Instituta BioSens. Prikupljeni podaci sadrže preciznu geolokaciju poljoprivrednih parcela, kao i informaciju o tipu gajene kulture za svaku posmatranu godinu. Dug vremenski period prikupljanja podataka omogućio je obuhvatan prikaz međugodišnje varijabilnosti poljoprivredne proizvodnje na području istraživanja i predstavljao je osnovu za izgradnju robusnog skupa za obuku i evaluaciju klasifikacionog modela.

Poljoprivredno zemljište korišćeno za finalnu klasifikaciju izdvojeno je na osnovu CORINE Land Cover European Environment Agency 2020 klasifikacije iz 2018. godine, čime je obezbeđeno da se analiza sprovodi isključivo nad relevantnim zemljišnim pokrivačem.

Klasifikacija useva sprovedena je primenom RF algoritma, ansambl metode zasnovane na velikom broju stabala odlučivanja, koja se pokazala kao pouzdana i efikasna u

obradi visoko-dimenzionalnih i heterogenih podataka daljinske detekcije. RF algoritam omogućava modelovanje nelinearnih odnosa između ulaznih obeležja i ciljnih klasa, uz smanjenu osetljivost na šum i pretreniranje, što ga čini pogodnim za operativne sisteme mapiranja poljoprivrednih kultura.

Ulazna obeležja modela činile su vrednosti spektralnih kanala Sentinel-2 snimaka i izvedeni vegetacioni indeksi, organizovani u obliku vremenskih serija na nivou piksela. Tokom procesa ekstrakcije podataka posebna pažnja posvećena je očuvanju veze između svakog piksela i poljoprivredne parcele kojoj pripada, kako bi se omogućila korektna prostorna validacija i interpretacija rezultata.

Nakon obuke modela, klasifikacija je izvršena nad celokupnim slojem poljoprivrednog zemljišta Autonomne Pokrajine Vojvodine, pri čemu je generisana tematska mapa prostorne rezolucije od 10 m, u skladu sa nativnom rezolucijom Sentinel-2 podataka.

Evaluacija performansi klasifikacionog modela sprovedena je primenom 10-k fold unakrsne validacije, sa ciljem dobijanja pouzdane i stabilne procene tačnosti. Kako bi se ublažio uticaj prostorne autokorelacije i izbeglo precenjivanje performansi modela, podela podataka na podskupove (foldove) izvršena je na nivou poljoprivrednih parcela, a ne na nivou pojedinačnih piksela. Na ovaj način je osigurano da svi pikseli koji pripadaju istoj parceli budu uključeni isključivo u jedan fold, čime je očuvana veza između piksela i odgovarajućeg polja tokom celokupnog procesa obuke i validacije. Ovakav pristup sprečava situacije u kojima bi se pikseli iz iste parcele mogli istovremeno pojaviti i u skupu za obuku i u skupu za testiranje, što bi dovelo do nerealno visokih procena tačnosti modela.

Finalna mapa klasifikacije obuhvata sedam tematskih klasa: kukuruz, pšenicu, soju, šećernu repu, suncokret, uljanu repicu i klasu ostalo, koja uključuje sve preostale kulture i heterogene poljoprivredne površine koje nisu obuhvaćene dominantnim klasama. Ovakva tematska podela omogućava detaljnu analizu prostorne raspodele glavnih ratarskih kultura u Vojvodini i predstavlja osnovu za dalje analize u okviru sistema za praćenje poljoprivrede zasnovanih na podacima daljinske detekcije.

U ovom radu razvijen je postupak klasifikacije useva zasnovan na višedatumskoj analizi Sentinel-2 optičkih podataka, pri čemu se klasifikacija vrši na nivou piksela uz korišćenje referentnih podataka o kulturama na nivou parcela. Metodologija je implementirana u programskom jeziku Python i obuhvata više međusobno povezanih koraka: pripremu trening skupa iz vektorskih referentnih podataka, ekstrakciju višedatumskih spektralnih atributa iz Sentinel-2 scena, konstrukciju deskriptora na nivou piksela, obuku klasifikatora, validaciju kroz parcelno-stratifikovanu kros-validaciju, prostorno eksplicitnu klasifikaciju svih poljoprivrednih površina i naknadnu postklasifikacionu

filtraciju rezultata. Implementacija je organizovana kroz centralni skript za klasifikaciju i skup pomoćnih modula za rad sa rasterima, geometrijama, Sentinel-2 metapodacima i filtracijom klasifikacionih mapa.

6.1.1 Ulazni podaci modela

Osnovni izvor daljinski osmatranih podataka činile su Sentinel-2 Level-2A scene, organizovane po putanjama i granulama (engl. tiles), pri čemu je obrada bila usmerena na granule koje pokrivaju područje Vojvodine. U kodu je definisan skup korišćenih spektralnih kanala B01–B12, pri čemu su kao operativni ulaz korišćeni svi raspoloživi optički kanali, uključujući 10 m, 20 m i 60 m prostorne rezolucije. Metapodaci o scenama i granulama čitaju se iz Sentinel-2 pomoćnih fajlova, pri čemu se za svaku scenu određuju vreme snimanja, orbitni broj, identifikator granule, geometrija obuhvata, broj redova i kolona, kao i georeferenciranje u prostoru. Na taj način obezbeđena je dosledna prostorno-vremenska organizacija svih ulaznih scena pre ekstrakcije spektralnih vrednosti.

U okviru obrade satelitskih scena sprovedeno je učitavanje pojedinačnih spektralnih kanala, njihovo povezivanje sa odgovarajućim prostornim koordinatama i izračunavanje vegetacionih indeksa. Posebna pažnja posvećena je radu sa georeferenciranim rasterima, kako bi se omogućilo tačno izdvajanje vrednosti piksela na osnovu prostornih koordinata. Ovaj postupak bio je značajan pri obradi mozaika, maski i drugih pomoćnih slojeva. Takođe su implementirani postupci za konverziju između prostornih i pikselskih koordinata, formiranje referentnih rastera, rasterizaciju vektorskih geometrija i čuvanje rezultata u GeoTIFF formatu.

Referentni podaci o kulturama zadati su u vektorskom formatu, pri čemu svaka parcela poseduje identifikator parcele i oznaku klase useva. U prvom koraku vektorski sloj se rasterizuje na prostornu mrežu usklađenu sa Sentinel-2 podacima. Funkcija za pripremu maski generiše binarnu masku obuhvata parcela, kao i rastere sa klasnom oznakom i identifikatorom parcele za svaki piksel unutar poljoprivrednih površina. Nakon rasterizacije izdvajaju se sve validne piksel-lokacije unutar parcela, njihove prostorne koordinate, oznake klase i pripadnost konkretnoj parceli. Ovi podaci se zatim čuvaju u komprimovanom obliku kao osnova za dalju ekstrakciju satelitskih atributa i obuku klasifikatora.

Ovakav pristup ima dve važne metodološke posledice. Prvo, svi pikseli unutar referentnih parcela postaju potencijalni uzorci za obuku, čime se dobija veoma gust skup trening podataka. Drugo, zadržava se informacija o pripadnosti piksela parceli, što omogućava da se kasnije validacija ne vrši nasumično na nivou piksela, već kontrolisano

na nivou celih parcela. Time se smanjuje opasnost od precenjivanja tačnosti usled prostorne autokorelacije između susednih piksela iste parcele.

Za svaku referentnu piksel-lokaciju iz trening skupa ekstrahuju se vrednosti svih Sentinel-2 kanala kroz celu vremensku seriju raspoloživih scena. U operativnoj implementaciji to znači da se za svaki datum snimanja i za svaki od 12 kanala čita odgovarajuća reflektansa u datim koordinatama. Postupak je organizovan tako da se scene najpre grupišu po putanji i datumu snimanja, a zatim se za svaku relevantnu granulu učitavaju svi optički kanali i vrši indeksiranje u koordinatama trening piksela. Rezultat je matrica dimenzija $N \times (12 \times T)$, gde je N broj piksela, a T broj datuma snimanja. Na ovaj način svaki piksel biva opisan ne samo trenutnim spektralnim potpisom, već kompletnom sezonskom dinamikom.

Pored direktnih spektralnih vrednosti, u kodu je implementirana i funkcija za računanje većeg broja vegetacionih i biofizičkih indeksa iz više kanala i više datuma. Među njima su NDVI, NDVI izveden iz red-edge kanala, EVI, VARI, SAVI, ARVI, GAVI, VDVI, NDMI, NDWI i više nelinearnih indeksa. Ovi indeksi se računaju za sve raspoložive termine i konkateniraju sa originalnim spektarskim atributima kada je izabran eksperiment sa dodatnim indeksima. Time se ulazni atributni prostor proširuje sa čistih spektralnih vrednosti na kombinaciju reflektansi i izvedenih deskriptora koji opisuju vegetaciono stanje, sadržaj vode i odgovor useva u različitim delovima spektra.

Važan element implementacije jeste i kontrola nepotpunih vremenskih serija. U kodu se koristi broj validnih spektralnih opservacija po pikselu, a za dalju obuku zadržavaju se uglavnom oni pikseli koji imaju gotovo maksimalan broj dostupnih merenja. Time se smanjuje prisustvo nedostajućih vrednosti i obezbeđuje homogeniji ulazni skup za klasifikaciju.

6.1.2 Obuka klasifikatora

Za klasifikaciju useva korišćen je algoritam RF, implementiran kroz biblioteku *scikit-learn*. U skripti je definisan klasifikator sa 100 stabala odlučivanja, bez ograničenja dubine stabla, uz korišćenje balansiranih težina klasa kako bi se ublažio uticaj klasne neuravnoteženosti. Upotreba RF metodološki je opravdan iz više razloga. Ovaj algoritam je robustan na visoko-dimenzione ulazne prostore, može da modeluje nelinearne odnose između spektralno-vremenskih atributa i klasne pripadnosti, relativno je otporan na šum i ne zahteva strogu pretpostavku o raspodeli ulaznih promenljivih. To je naročito važno u višedatumskim satelitskim nizovima, gde različite kulture mogu imati slične vrednosti u pojedinačnim terminima, ali se razlikuju po celokupnoj sezonskoj trajektoriji.

Za svrhu obuke modela korišćena je procedura koja najpre grupiše sve piksele prema

parcelama i klasama, zatim parcele raspodeljuje u K podskupova. Algoritam dodatno balansira podskupove prema broju piksela parcele kako bi u svakom podskupu bio podjednak broj piksela. Razlog za ovu dodatnu optimizaciju jeste činjenica da različite parcele imaju različite površine, pa bi nasumična podela po broju parcela mogla dovesti do velike neravnoteže u broju piksela između validacionih podskupova.

U praksi to znači da nijedna parcela ne učestvuje istovremeno i u trening i u test skupu. Na ovaj način procena tačnosti više odgovara realnoj generalizaciji modela, umesto da bude veštački uvećana usled činjenice da pikseli iste parcele imaju veoma sličan spektralni odziv. Za svaki podskup trenira se poseban RF model, vrši predikcija nad test skupom i formira matrica konfuzije. Na osnovu zbirne matrice konfuzione računaju se standardni metrički pokazatelji, uključujući tačnost po klasi, osetljivost, specifičnost, preciznost i F-meru.

6.1.3 Generisanje finalnih mapa klasifikacije useva

Nakon obuke modela, klasifikator se primenjuje na sve poljoprivredne površine unutar područja interesa. U tu svrhu koristi se maska poljoprivrednog zemljišta koja ograničava klasifikaciju na relevantne piksele. Područje obrade se deli po Sentinel-2 granulama i po particijama kada je broj piksela prevelik za direktnu obradu u memoriji. Za svaki piksel unutar maske ponovo se ekstrahuje kompletna višedatumska spektarska vremenska serija, po potrebi se računaju dodatni indeksi, a zatim se primenjuje prethodno obučeni RF model.

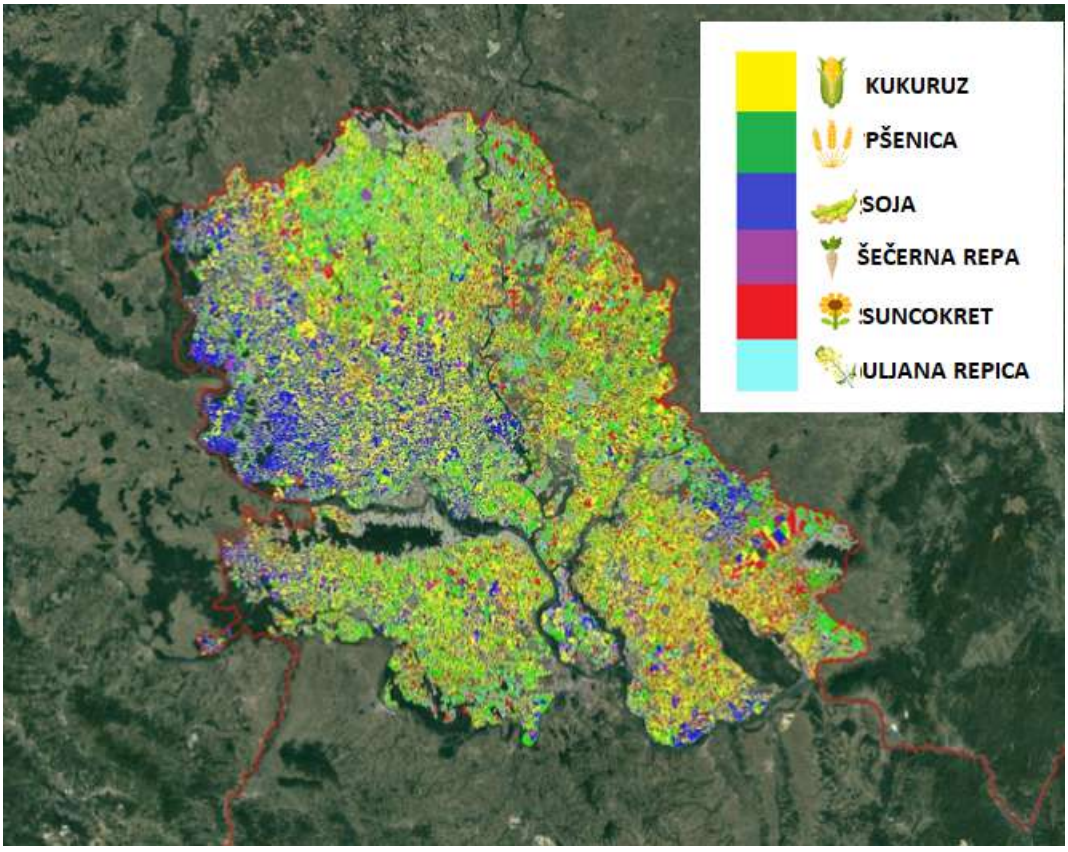
Pored diskretne oznake klase, model računa i verovatnoće pripadnosti svim klasama. Ove verovatnoće se upisuju kao posebni izlazni rasteri, što omogućava kasniju procenu sigurnosti klasifikacije i eventualno kombinovanje više modela ili eksperimenata. Rezultat klasifikacije je za svaku granulu geoprostorni raster sa oznakom dominantne klase useva i odgovarajućim rasterima za verovatnoću svake klase.

Početni rezultat piksel klasifikacije može sadržati lokalni šum, izolovane pogrešno klasifikovane piksele i geometrijski nerealne obrasce unutar parcela. Zbog toga je implementiran poseban postklasifikacioni postupak filtracije. U modulu filtraciju definisana je funkcija prostorne filtracije zasnovana na transformaciji klasifikacione mape u koordinatni sistem poravnat sa orijentacijom parcele. Nakon rotacije, za svaku klasu se računa lokalna mapa verovatnoće u pokretnom prozoru, korišćenjem integralne slike radi ubrzanja proračuna. Potom se svakom pikselu dodeljuje klasa sa najvećom lokalnom podrškom, a rezultat se vraća u originalni koordinatni sistem.

Pored toga, u zasebnom modulu implementirana je dodatna filtracija po parcelama, pri čemu se klasifikaciona mapa preseca sa katastarskim parcelama iz prostorne baze

podataka. Time se dobija mogućnost da se klasifikacija unutar svake parcele dodatno koriguje i homogenizuje u skladu sa parcelnom geometrijom. Ovakav pristup je naročito značajan u operativnim primenama, jer realne parcele uglavnom imaju dominantnu kulturu, pa se lokalni piksel-šum može efikasno ukloniti korišćenjem prostornog konteksta i geometrije parcele.

Nakon klasifikacije i filtracije pojedinačnih granula, rezultati se spajaju u mozaik za celo područje istraživanja. U tu svrhu koriste se pomoćne funkcije za mozaiciranje rastera, koje vode računa o prostornom preklapanju i pravilnom prepisivanju vrednosti u zajednički izlazni raster. Finalni rezultat predstavlja prostorno konzistentnu klasifikacionu mapu useva za celu analiziranu oblast. Dodatno, implementirane su i funkcije za konverziju raster rezultata u vektorski oblik i za izdvajanje pojedinačnih kultura u zasebne shapefile slojeve, kao i alati za računanje statistika po administrativnim jedinicama.



Slika 6.1: Grafički prikaz klasifikacije 6 najzastupljenijih useva u Vojvodini.

TPR – True Positive Rate (osetljivost), FNR – False Negative Rate, FPR – False Positive Rate, TNR – True Negative Rate (specifičnost), PREC – preciznost, FDR – False Discovery Rate, ACC – tačnost, INACC – netačnost, F – F1-metra.

Tabela 6.1: Performanse klasifikacije po klasama useva za 2019. godinu (TPR, FNR, FPR, TNR, PREC, FDR, ACC, INACC, F-score)

Usev	TPR	FNR	FPR	TNR	PREC	FDR	ACC	INACC	F
kukuruz	0.9131	0.0869	0.0279	0.9721	0.8507	0.1493	0.9633	0.0367	0.8808
pšenica	0.9634	0.0366	0.0396	0.9604	0.8908	0.1092	0.9612	0.0388	0.9256
soja	0.7883	0.2117	0.0117	0.9883	0.8470	0.1530	0.9730	0.0270	0.8166
šećerna repa	0.9470	0.0530	0.0078	0.9922	0.9729	0.0271	0.9818	0.0182	0.9598
suncokret	0.7895	0.2105	0.0153	0.9847	0.8534	0.1466	0.9650	0.0350	0.8202
uljana repica	0.9247	0.0753	0.0002	0.9998	0.9758	0.0242	0.9991	0.0009	0.9496
ostalo	0.8559	0.1441	0.0181	0.9819	0.9146	0.0854	0.9586	0.0414	0.8842
Prosek	0.8831	0.1169	0.0172	0.9828	0.9007	0.0993	0.9717	0.0283	0.8910

6.1.4 Metodološke prednosti i ograničenja

Razvijeni pristup ima više važnih prednosti. Prvo, zasniva se na punoj vremenskoj seriji Sentinel-2 podataka, čime se iskorišćava fenološka dinamika useva, a ne samo informacija iz jednog datuma. Drugo, referentni uzorci su organizovani na nivou parcela, što omogućava znatno realističniju validaciju nego klasična slučajna podela piksela. Treće, RF klasifikator dobro podnosi visokodimenzione i heterogene ulazne prostore, dok postklasifikaciona filtracija dodatno povećava prostornu konzistentnost izlaza.

Istovremeno, određena ograničenja treba jasno naglasiti. Prvo, ulazni atributni prostor može postati veoma velik, naročito kada se koristi više datuma i više indeksa, što povećava memorijske zahteve i vreme obrade. Drugo, dostupnost i kvalitet klasifikacije direktno zavise od gustine i konzistentnosti referentnih parcelnih podataka. Treće, iako je parcelna kros-validacija metodološki ispravnija od pikselske, ona i dalje zavisi od reprezentativnosti referentnog skupa u odnosu na celu teritoriju. Konačno, deo postprocesinga zasniva se na heurističkim pravilima prostorne homogenizacije, pa se dobijene korekcije moraju tumačiti kao operativno unapređenje rezultata, a ne kao deo osnovnog statističkog modela.

U celini posmatrano, razvijena metodologija predstavlja robustan okvir za klasifikaciju useva na velikim prostornim obuhvatima korišćenjem Sentinel-2 vremenskih serija i parcelnih referentnih podataka. Njena glavna vrednost ogleda se u povezivanju vremenskih serija multispektralnih satelitskih snimaka sa parcelno zasnovanom validacijom i prostornom filtracijom, što omogućava generisanje operativnih mapa useva pogodnih za dalje analize.

Poglavlje 7

Predikcija prinosa i proteinskog sadržaja

Sadržaj

7.1 Predikcija prinosa useva 99
7.1.1 Priprema satelitskih podataka 100
7.1.2 Referentni podaci o prinosu i prostorna priprema uzoraka 100
7.1.3 Konstrukcija vremenskih serija spektralnih atributa 101
7.1.4 Eksperimentalni dizajn: realni, GAN generisani i kombinovani atributi 102
7.1.5 Validacija metodom izostavljanja jedne parcele - Leave-One-Parcel-Out 102
7.1.6 Metodološke prednosti pristupa 103
7.1.7 Finalne mape predikcije prinosa useva 103
7.1.8 Rezultati modela za predikciju prinosa 104
7.1.9 Rezultati predikcije prinosa za 2018. godinu 109
7.1.10 Rezultati za 2019. godinu 112
7.1.11 Poređenje modela i eksperimentalnih režima 116
7.1.12 Interpretacija rezultata u odnosu na vrednosti prinosa 117
7.1.13 Evaluacija generalizacije modela na podacima iz različitih izvora 121
7.2 Predikcija proteinskog sastava useva 123
7.2.1 Opis podataka, transformacija vremenskih serija i konstrukcija ulaznih obeležja 123
7.2.2 Klasifikacija proteinskog sastava 125
7.2.3 Rezultati klasifikacije proteinskog sastava 125

7.1 Predikcija prinosa useva

U cilju procene prinosa useva razvijeni su modeli koji kao ulazne podatke koriste vremenske serije Sentinel-2 snimaka. Metodologija je osmišljena tako da omogući predikciju prinosa na nivou piksela korišćenjem vremenske serije optičkih snimaka, uz

strogu evaluaciju generalizacije modela između različitih parcela. Celokupan postupak sastoji se iz dve glavne faze: (i) pripreme skupa satelitskih ulaznih podataka i (ii) obuke i evaluacije regresionih modela za predikciju prinosa.

7.1.1 Priprema satelitskih podataka

Satelitski podaci organizovani su u obliku višekanalnih rasterskih stackova, pri čemu originalni stack sadrži 15 kanala: dva Sentinel-1 polarizaciona kanala (VV i VH), dvanaest Sentinel-2 optičkih kanala (B1–B12) i sloj klasifikacije scene (SCL). Za potrebe ovog eksperimenta koriste se isključivo Sentinel-2 snimci, dok se odgovorajući Sentinel-1 snimci koriste za GAN model. Za datume sa snimcima bez oblaka koriste se originalni Sentinel-2 snici sa procentom oblacnošti manjim od 10 % dok se GAN rekonstruisanim optičkim snimcima koriste za datume gde je oblačnost veća. Na taj način dobijen je skup ulaznih podataka sa konzistentnom strukturom od 12 optičkih kanala po datumu, nezavisno od toga da li je scena originalni snimak ili rekonstrukcija snimaka sa oblačnim datumima.

Novi rasteri napravljeni pomoću GAN modela zadržavaju georeferenciranje i koordinatni sistem originalnih snimaka, čime je obezbeđena puna prostorna kompatibilnost između izvornih i rekonstruisanih scena. Rezultat ove faze je standardizovan skup 12 Sentinel-2 kanala koji predstavlja osnovu za dalje uzorkovanje i konstrukciju atributa za predikciju prinosa.

7.1.2 Referentni podaci o prinosu i prostorna priprema uzoraka

Kao referentni podaci korišćen je vektorski sloj sa geometrijama uzoraka i vrednostima prinosa useva. Definisana su dva ključna atributa: kolona sa vrednostima prinosa i kolona sa identifikatorom parcele, koja služi za grupisanje uzoraka tokom validacije. Geometrije uzoraka prestavljena je vektorskom zapisom u vidu georeferenciranih struktura u vidu tačaka. Time se svaki referentni zapis svodi na jednu prostornu lokaciju u kojoj se iz rastera uzimaju vrednosti vremenske serije.

Pre ekstrakcije atributa vrši se osnovno čišćenje referentnog skupa. Izbacuju se zapisi bez validne vrednosti prinosa, geometrije bez prostorne definicije i uzorci sa prinosom jednakim nuli. Pored toga, uveden je i globalni filtracioni korak zasnovan na kvantilima raspodele prinosa, pri čemu se zadržavaju samo uzorci unutar intervala između 10. i 90. percentila. Time se uklanjaju ekstremno niski i ekstremno visoki prinosi koji mogu predstavljati netačne vrednosti prinosa ili merne greške. Ovaj korak ima za cilj da stabilizuje regresioni model i smanji njegovu osetljivost na krajnje ekstremne vrednosti.

Pošto se satelitski rasteri i referentni vektorski sloj mogu nalaziti u različitim koordinatnim sistemima, pre uzorkovanja se koordinatni sistem vektorskih podataka usklađuje sa koordinatnim sistemom rastera. Nakon toga se za svaki referentni zapis formira lista koordinata u kojima će biti preuzeti satelitski podaci. Uz samu vrednost prinosa čuva se i oznaka parcele, što će kasnije biti ključno za validacioni korak.

7.1.3 Konstrukcija vremenskih serija spektralnih atributa

Osnovna ideja modela jeste da se svaka referentna lokacija opiše ne jednim satelitskim snimkom, već celokupnom vremenskom serijom optičkih snimaka kroz sezonu. U tu svrhu za svaki dostupni datum snimanja iz odgovarajućeg 12–kanalnog rastera uzimaju se vrednosti Sentinel-2 kanala B1–B12 u koordinatama uzoraka. Datumi su organizovani tako da se najpre identifikuju svi dostupni datumi bez oblaka, a zatim svi datumi iz skupa oblačnih scena, a potom se formira objedinjena lista svih datuma za koje postoje realni čisti Sentinel-2 snimci i GAN generisani snimci.

Pored dvanaest originalnih spektralnih kanala, za svaki datum izračunat je i skup vegetacionih indeksa. Indeksi su izvedeni iz standardnih Sentinel-2 kanala i uvedeni su kako bi se dodatno opisali aspekti vegetacione aktivnosti, hlorofilnog statusa, vodnog sadržaja i stresa biljaka. Na primer, NDVI, GNDVI i EVI primarno opisuju intenzitet vegetacije i fotosintetsku aktivnost, NDRE i CIRE dodatno naglašavaju promene u blisko infracrvenom domenu povezane sa sadržajem hlorofila, dok NDWI nosi informaciju o vodnom statusu i stresu vegetacije. Na ovaj način svaki datum nije predstavljen samo reflektansama, već proširenim skupom izvedenih vegetacionih indeksa.

Podaci sa monitora prinosa se pomoću prethodno opisane PPI metode iz tačkaste strukture prevode u poligone. Ovim postupkom se omogućava da se podaci o prinosu useva sa monitora prinosa lakše preklope sa raster podacima odnosno satelitskim snimcima. Na taj način se definišu atributi za svaku pojedinačnu vrednost useva. Za modele mašinskog učenja atributi predstavljaju ulazne podake, a prinos useva je ciljana varijabla.

Konačni vektor obeležja za jedan uzorak formira se spajanjem svih atributa po datumima, odnosno 12 spektralnih kanala i 10 izvedenih indeksa. Ukupan broj ulaznih atributa zavisi od broja raspoloživih datuma i jednak je proizvodu broja datuma i broja atributa. Ovakva konstrukcija omogućava da se obuhvati dinamika rasta useva kroz sezonu i omogućava modelu da razlikuje uzorke ne samo na osnovu pojedinačnog stanja, već na osnovu kompletnog sezonskog razvoja.

7.1.4 Eksperimentalni dizajn: realni, GAN generisani i kombinovani atributi

Jedna od ključnih karakteristika implementacije jeste podela ulaznih podataka na tri eksperimentalna režima. Prvi režim obuhvata isključivo atribute izvedene iz originalnih Sentinel-2 snimaka bez značajne oblačnosti. Drugi režim obuhvata isključivo GAN rekonstruisane snimke za datume kod kojih su originalne optičke opservacije bile degradirane prisustvom oblaka.

Treći, kombinovani režim formiran je objedinjavanjem čistih realnih datuma i GAN rekonstrukcija oblačnih datuma. Za svaki datum korišćena je samo jedna verzija optičke opservacije, tako da isti datum nije istovremeno predstavljen realnim i GAN rekonstruisanim snimkom. Na taj način moguće je nezavisno proceniti koliki doprinos u predikciji prinosa imaju originalni satelitski datumi, koliki doprinos imaju rekonstruisane scene i da li njihova kombinacija dovodi do poboljšanja performansi modela.

Ovaj eksperimentalni dizajn je posebno značajan u kontekstu rada sa oblačnim snimcima i generativnim modelima za popunjavanje nedostajućih optičkih informacija. Umesto da se GAN podaci koriste samo kao vizuelna dopuna, u ovom pristupu oni se tretiraju kao ulazni atributi i direktno se koriste u regresionim modelima za predikcije prinosa. Time se evaluacija rekonstrukcije podiže na viši nivo, jer se meri njen stvarni doprinos u realnoj primeni.

Za svaki od tri eksperimentalna skupa atributa trenira se poseban model zasnovan samo na realnim datumima, model zasnovan samo na GAN datumima i model zasnovan na kompletnom atributnom skupu. Time se omogućava direktno poređenje performansi između različitih izvora informacija, kako na trening, tako i na test podacima.

7.1.5 Validacija metodom izostavljanja jedne parcele - Leave-One-Parcel-Out

Umesto nasumične podele uzoraka na trening i test skup, primenjena je validacija metodom izostavljanja jedne parcele (engl. Leave-One-Parcel-Out – LOPO), pri čemu svaka parcela služi kao potpuno izdvojeni test skup, dok se model trenira na svim preostalim parcelama. Ovakav pristup je znatno stroži i realističniji od klasične nasumične podele na nivou pojedinačnih tačaka ili piksela, jer sprečava da uzorci iz iste parcele istovremeno budu prisutni i u trening i u test delu. Na taj način se izbegava optimistička procena tačnosti usled prostorne autokorelacije i međusobne sličnosti uzoraka unutar iste parcele.

U svakom LOPO koraku izdvaja se jedna test parcela, pri čemu se vrlo male parcele

sa manje od 50 uzoraka (piksela) odbacuju kako bi procena performansi bila stabilnija. Za preostale parcele formiraju se trening i test podskupovi, treniraju svi regresioni modeli i računa se skup evaluacionih metrika.

Performanse modela procenjuju se korišćenjem standardnih regresionih metrika: srednja apsolutna greška (engl. Mean Absolute Error - MAE), srednja kvadratna greška (engl. Mean Squared Error - MSE) i koren srednje kvadratne greške (engl. Root Mean Squared Error - RMSE). Na nivou svakog LOPO koraka beleže se vrednosti MSE, RMSE i MAE za model obučavan na realnim, GAN generisani i kombinovani skupom podataka. Pored pojedinačnih rezultata po parcelama, na kraju se računa i zbirna statistika raspodele performansi, uključujući srednju vrednost, medijanu, kvartile, minimum i maksimum. Takav pristup omogućava da se uspešnost modela ne tumači samo kroz jednu agregatnu vrednost, već kroz celu distribuciju performansi između različitih parcela.

7.1.6 Metodološke prednosti pristupa

Razvijeni pristup ima nekoliko značajnih metodoloških prednosti. Prvo, model koristi Sentinel-2 vremensku seriju, čime se fenološki razvoj useva uvodi kao ključna komponenta u predikciji prinosa. Drugo, kombinovanjem sirovih spektralnih kanala i izvedenih vegetacionih indeksa obezbeđuje se mnogi faktori koji utiču na stanje vegetacije, vodnog režima i stresa useva kroz sezonu. Treće, posebno je važna eksperimentalna podela na realne i GAN datume, jer omogućava da se direktno proceni operativna vrednost generativno rekonstruisanih satelitskih scena u zadatku predikcije prinosa. Četvrto, evaluacija tipa LOPO daje strožu i metodološki korektniju procenu generalizacije nego validacija na nivou pojedinačnih tačaka, što je od naročitog značaja u prostorno zavisnim agroekološkim podacima.

7.1.7 Finalne mape predikcije prinosa useva

Mape prinosa useva generisane su na osnovu vremenskih serija multispektralnih Sentinel-2 satelitskih snimaka, koji omogućavaju praćenje razvoja vegetacije i njenih biofizičkih karakteristika tokom čitave vegetacione sezone. Iz spektralnih kanala Sentinel-2 podataka izračunat je skup relevantnih vegetacionih indeksa, koji su korišćeni kao ulazna obeležja za modele predikcije prinosa.

Referentne (ground truth) podatke predstavljaju podaci sa monitora prinosa prikupljeni tokom žetve za četiri dominantne ratarske kulture: kukuruz, pšenicu, soju i suncokret. Podaci sa monitora prinosa sadrže informacije o ostvarenom prinosu, koje

su agregirane i usklađene sa prostornom rezolucijom Sentinel-2 snimaka. Time je obezbeđena konzistentnost između satelitskih ulaznih podataka i referentnih vrednosti prinosa na nivou piksela.

Kao i u slučaju klasifikacije useva, poljoprivredno zemljište korišćeno za finalnu predikciju prinosa izdvojeno je na osnovu CORINE Land Cover klasifikacije, čime je obezbeđeno da se modeli primenjuju isključivo nad relevantnim poljoprivrednim površinama na teritoriji Autonomne Pokrajine Vojvodine.

Za predikciju prinosa useva korišćeno je više modela: MLR, RF, XGBoost, SVM i LSTM neuronske mreže. Tokom procesa pripreme podataka, posebna pažnja posvećena je očuvanju veze između svakog piksela i poljoprivredne parcele kojoj pripada, kako bi se omogućila korektna prostorna validacija modela i izbeglo curenje informacija između skupa za obuku i skupa za testiranje.

Evaluacija performansi modela za predikciju prinosa sprovedena je korišćenjem strategije LOPO unakrsne validacije. Ovaj pristup podrazumeva da se u svakom koraku validacije kompletna poljoprivredna parcela izostavlja iz skupa za obuku i koristi isključivo za testiranje modela, čime se efikasno eliminiše uticaj prostorne autokorelacije i obezbeđuje realistična procena generalizacione sposobnosti modela.

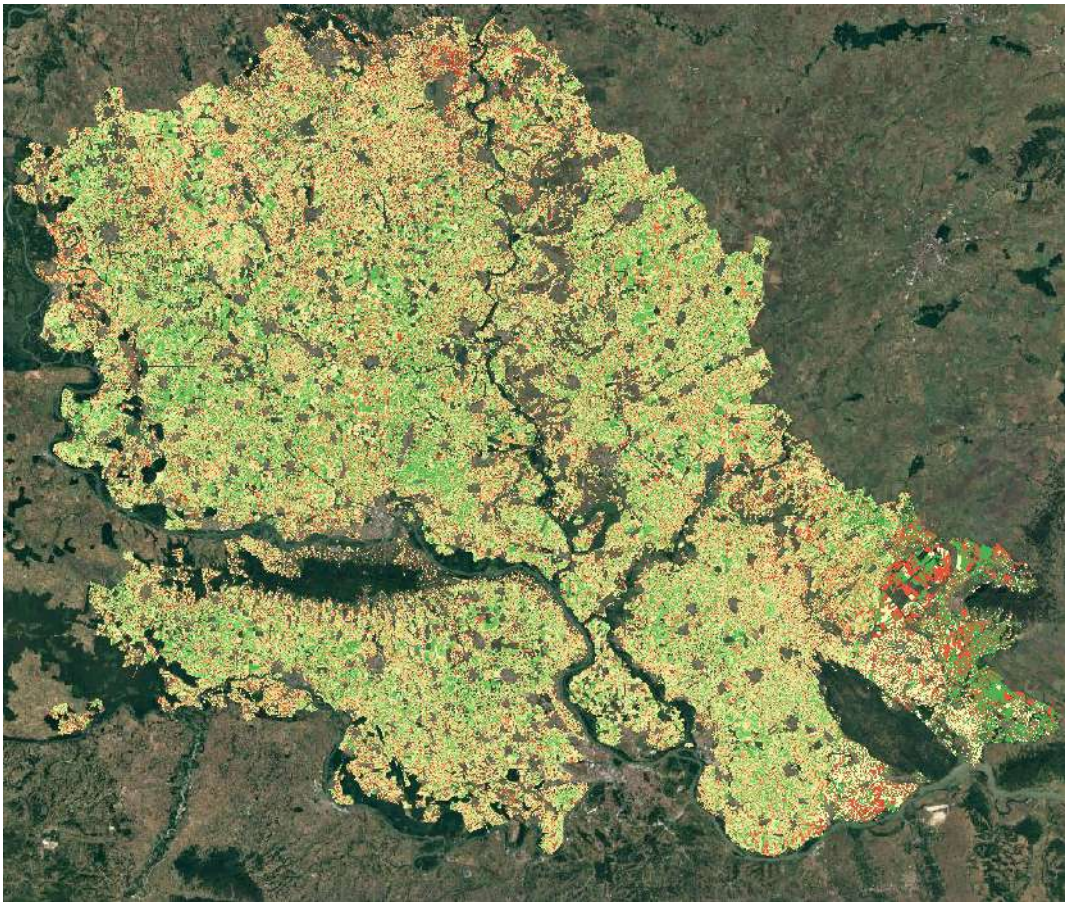
Primena LOPO strategije omogućava da se modeli evaluiraju u uslovima koji verno odražavaju operativni scenario, u kome se predikcija prinosa vrši na parcelama koje nisu bile dostupne tokom obuke. Time se osigurava stabilnost i pouzdanost procene performansi, posebno u kontekstu prostorno eksplicitnih poljoprivrednih podataka.

Nakon obuke modela, predikcija prinosa izvršena je nad celokupnim slojem poljoprivrednog zemljišta Autonomne Pokrajine Vojvodine, pri čemu su generisane kontinuirane mape prinosa u prostornoj rezoluciji od 10 m, u skladu sa nativnom rezolucijom Sentinel-2 podataka.

Finalne mape predikcije prinosa generisane su u prostornoj rezoluciji od 10 m i predstavljaju osnovu za dalju analizu prostorne varijabilnosti prinosa, identifikaciju zona upravljanja i integraciju sa sistemima za podršku odlučivanju u preciznoj poljoprivredi.

7.1.8 Rezultati modela za predikciju prinosa

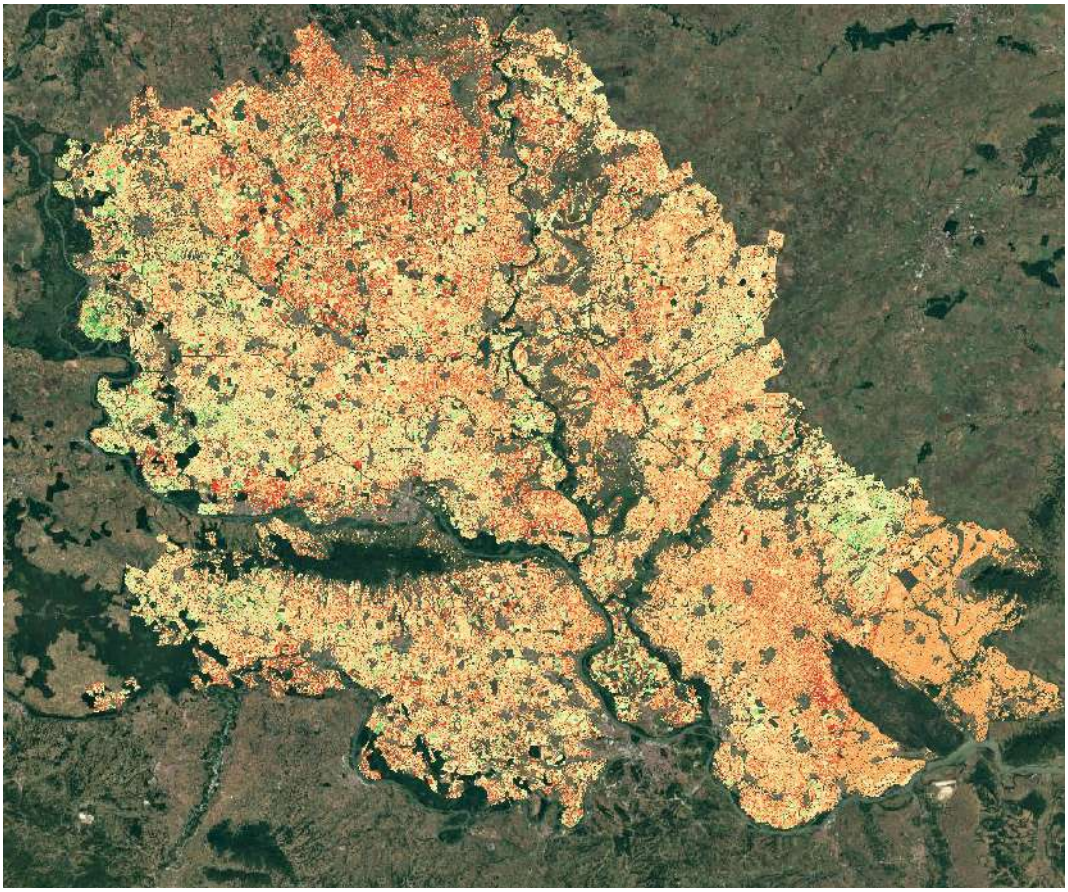
Za pretragu hiperparametara izdvojen je reprezentativni podskup iz sezone 2018. godine, koji je obuhvatio podatke za kukuruz dobijene iz realnih Sentinel-2 snimaka. Optimizacija je sprovedena pomoću paketa Optuna u okviru ugnježdene LOPO unakrsne validacije. U svakoj iteraciji korišćena je posebna unutrašnja podela na nivou parcela, pri čemu izostavljena test parcela nije učestvovala u selekciji obeležja niti u izboru hiperparametara. Pošto je u svakoj LOPO iteraciji dobijena posebna optimalna konfiguracija, konačni fiksni



Slika 7.1: Grafički prikaz predikcije prinosa na nivou piksela za kukuruz, pšenicu, soju i suncokret na teritoriji AP Vojvodine

parametri određeni su na osnovu najčešće biranih kategorijalnih vrednosti i medijana numeričkih vrednosti. Tako određene konfiguracije zatim su bez dodatnog podešavanja primenjene na sve kulture, obe godine i sva tri režima ulaznih podataka.

Broj obeležja predstavlja reprezentativnu vrednost određenu na osnovu medijane konfiguracija izabranih u pojedinačnim LOPO iteracijama. Za neuronsku mrežu medijana je iznosila 120 obeležja, dok je za ostale modele iznosila 200. Kada je ukupan broj raspoloživih obeležja bio manji od navedene vrednosti, korišćena su sva dostupna obeležja. Razlika u broju obeležja zato nije unapred zadata osobina modela, već rezultat Optuna pretrage.



Slika 7.2: Grafički prikaz predikcije prinosa na nivou piksela za kukuruz, pšenicu, soju i suncokret na teritoriji AP Vojvodine, sezona 2019.

Tabela 7.1: Reprezentativni fiksni hiperparametri i korišćene Python biblioteke

Model i broj obeležja	**Biblioteka i implementacija**	**Odabrani hiperparametri**
RF	`scikit-learn`	`n_estimators` = 700
$k = 200$	`RandomForestRegresso`	`max_depth` = 24
		`min_samples_split` = 9
		`min_samples_leaf` = 5
		`max_features` = log2
		`criterion` = squared_error
		`bootstrap` = True
		`max_samples` = 0.7725

Nastavak Tabele 7.1

Model i broj obeležja	**Biblioteka i implementacija**	**Odabrani hiperparametri**
XGBoost $k = 200$	`xgboost` `XGBRegressor`	`n_estimators` = 900 `max_depth` = 5 `learning_rate` = 0.03788 `subsample` = 0.7942 `colsample_bytree` = 0.7187 `min_child_weight` = 8.4465 `gamma` = 0.06488 `reg_alpha` = 7.59×10^{-6} `reg_lambda` = 0.24498
MLR $k = 200$	`scikit-learn` `LinearRegression`	`fit_intercept` = False
SVR $k = 200$	`scikit-learn` `SVR`	`kernel` = rbf `C` = 1.2024 `epsilon` = 0.02178 `gamma` = auto `shrinking` = True `tol` = 1.82×10^{-4} `degree` = 3 `coef0` = 0
MLR_SGD $k = 200$	`scikit-learn` `SGDRegressor`	`loss` = huber `penalty` = l2 `alpha` = 2.73×10^{-6} `learning_rate` = constant `eta0` = 0.00707 `average` = False `epsilon` = 0.17751 `max_iter` = 10000 `tol` = 1×10^{-6}

Nastavak Tabele 7.1

Model i broj obeležja	**Biblioteka i implementacija**	**Odabrani hiperparametri**
NN $k = 120$	`TensorFlow/Keras` `Sequential`	`hidden_layers` = (96, 64) `activation` = ELU `dropout` = 0.20 `l2_strength` = 3.57×10^{-5} `learning_rate` = 0.001073 `batch_size` = 64 `epochs` = 28 `iterations` = 3

Performanse modela za predikciju prinosa procenjene su primenom LOPO unakrsne validacije. U svakoj iteraciji jedna parcela je u potpunosti izostavljena iz skupa za obuku i korišćena isključivo za testiranje, dok su sve preostale parcele korišćene za obuku modela. Ovakav postupak sprečava da prostorno bliski pikseli sa iste parcele istovremeno pripadaju skupu za obuku i skupu za testiranje i omogućava realističniju procenu prenosivosti modela na novu prostornu jedinicu.

Analizirana su tri režima ulaznih podataka. U eksperimentu sa realnim snimcima korišćeni su isključivo raspoloživi čisti Sentinel-2 snimci. U eksperimentu sa GAN rekonstruisanim snimcima korišćena je vremenska serija rekonstruisanih snimaka. U kombinovanom eksperimentu formirana je jedinstvena vremenska serija u kojoj je za datum sa raspoloživim realnim snimkom korišćen originalni snimak, dok je za nedostajući datum korišćena GAN rekonstrukcija. Na taj način isti datum nije uključen dva puta.

Za svaki datum formirana su 22 ulazna obeležja, koja obuhvataju 12 multispektralnih kanala i 10 spektralnih indeksa. U 2018. godini eksperiment sa realnim snimcima sadržao je 12 datuma i 264 izvorna obeležja. GAN eksperiment sadržao je 5 rekonstruisanih oblačnih datuma i 110 početnih obeležja, dok je kombinovani eksperiment obuhvatio 12 realnih i 5 GAN rekonstruisanih datuma, odnosno ukupno 17 datuma i 374 početna obeležja. U 2019. godini realni eksperiment sadržao je 7 datuma i 154 obeležja, GAN eksperiment 10 rekonstruisanih oblačnih datuma i 220 obeležja, dok je kombinovani eksperiment sadržao 7 realnih i 10 GAN rekonstruisanih datuma, odnosno ukupno 17 datuma i 374 početna obeležja.

Testirano je šest regresionih algoritama: RF, XGBoost, MLR, SVR, MLR_SGD i NN. Rezultati su prikazani pomoću srednje kvadratne greške (MSE), korena srednje kvadratne

greške (RMSE) i srednje apsolutne greške (MAE). Niže vrednosti svih navedenih metrika označavaju veću preciznost. U svakoj tabeli podebljana je najniža vrednost RMSE i najniža vrednost MAE za posmatranu godinu i kulturu. Za razliku od prethodnih izvršavanja, svih šest modela završilo je obuku u svim analiziranim kombinacijama, pa u novim tabelama nema rezultata označenih kao divergencija.

7.1.9 Rezultati predikcije prinosa za 2018. godinu

Vrednosti RMSE i MAE izražene su u jedinici t/piksel, dok je MSE izražen u $(\text{t/piksel})^2$. Jedan piksel predstavlja površinu od 10 m × 10 m, pa RMSE i MAE odgovaraju odstupanju prinosa po jediničnom pikselu Sentinel-2 mreže.

7.1.9.1 Predikcija prinosa kukuruza

Tabela 7.2: Rezultati svih modela za predikciju prinosa kukuruza u 2018. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00023	**0.01469**	**0.01153**
	XGBoost	0.00024	0.01521	0.01200
	MLR	0.00086	0.02244	0.01653
	SVR	0.00023	0.01501	0.01185
	MLR_SGD	0.00059	0.02190	0.01599
	NN	0.00029	0.01647	0.01258
GAN rekonstruisani snimci	RF	0.00023	0.01499	0.01180
	XGBoost	0.00025	0.01560	0.01240
	MLR	0.00053	0.01956	0.01300
	SVR	0.00025	0.01535	0.01222
	MLR_SGD	0.00048	0.02036	0.01552
	NN	0.00031	0.01712	0.01308
Kombinovani snimci	RF	0.00023	0.01504	0.01186
	XGBoost	0.00026	0.01578	0.01259
	MLR	0.00047	0.01915	0.01384
	SVR	0.00026	0.01569	0.01257
	MLR_SGD	0.00110	0.02368	0.01452
	NN	0.00029	0.01675	0.01350

Najnižu prosečnu grešku za kukuruz u 2018. godini ostvario je RF nad realnim snimcima, sa RMSE od 0.01469 i MAE od 0.01153. RF je bio najbolji i u GAN režimu (RMSE=0.01499, MAE=0.01180), kao i u ispravljenom kombinovanom režimu koji je obuhvatio sve realne i sve GAN snimke (RMSE=0.01504, MAE=0.01186). Razlike između ova tri rezultata bile su male, ali kombinovanje svih raspoloživih blokova nije dodatno smanjilo grešku. SVR je u realnom režimu ostvario rezultat veoma blizak RF modelu (RMSE=0.01501, MAE=0.01185), dok su MLR i MLR_SGD pokazali veće prosečne greške. Rezultati predstavljaju prosek kroz 11 prostornih LOPO iteracija.

7.1.9.2 Predikcija prinosa soje

Tabela 7.3: Rezultati svih modela za predikciju prinosa soje u 2018. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00003	0.00486	0.00397
	XGBoost	0.00003	0.00538	0.00450
	MLR	0.00036	0.01417	0.01293
	SVR	0.00003	0.00506	0.00414
	MLR_SGD	0.00006	0.00744	0.00614
	NN	0.00003	0.00518	0.00425
GAN rekonstruisani snimci	RF	0.00003	0.00485	0.00395
	XGBoost	0.00003	0.00535	0.00446
	MLR	0.00013	0.01011	0.00892
	SVR	0.00003	0.00501	0.00409
	MLR_SGD	0.00005	0.00675	0.00560
	NN	0.00005	0.00710	0.00611
Kombinovani snimci	RF	0.00003	0.00487	0.00392
	XGBoost	0.00003	0.00534	0.00446
	MLR	0.00011	0.00952	0.00791
	SVR	0.00002	**0.00460**	**0.00366**
	MLR_SGD	0.00015	0.01025	0.00869
	NN	0.00007	0.00791	0.00697

Kod soje je najniža greška ostvarena SVR modelom u kombinovanom režimu, sa RMSE od 0.00460 i MAE od 0.00366. U režimima sa pojedinačnim izvorima najbolji je bio RF, sa RMSE od 0.00485 i MAE od 0.00395 za GAN snimke, odnosno RMSE

od 0.00486 i MAE od 0.00397 za realne snimke. Uključivanje svih realnih i svih GAN blokova smanjilo je RMSE za približno 5% u odnosu na najbolji rezultat pojedinačnog izvora. Ipak, zaključak treba tumačiti sa oprezom jer su za soju bile dostupne samo tri prostorno nezavisne parcele.

7.1.9.3 Predikcija prinosa pšenice

Tabela 7.4: Rezultati svih modela za predikciju prinosa pšenice u 2018. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00023	0.01384	0.01168
	XGBoost	0.00024	0.01443	0.01224
	MLR	0.00020	0.01377	**0.01133**
	SVR	0.00022	**0.01373**	0.01144
	MLR_SGD	0.00022	0.01420	0.01160
	NN	0.00026	0.01390	0.01185
GAN rekonstruisani snimci	RF	0.00026	0.01415	0.01206
	XGBoost	0.00028	0.01523	0.01301
	MLR	0.00032	0.01619	0.01310
	SVR	0.00024	0.01394	0.01170
	MLR_SGD	0.00052	0.02038	0.01651
	NN	0.00023	0.01385	0.01180
Kombinovani snimci	RF	0.00024	0.01374	0.01167
	XGBoost	0.00027	0.01480	0.01253
	MLR	0.00039	0.01684	0.01462
	SVR	0.00023	0.01389	0.01177
	MLR_SGD	0.00037	0.01672	0.01439
	NN	0.00024	0.01441	0.01244

Za pšenicu u 2018. godini najniži RMSE ostvario je SVR nad realnim snimcima (0.01373), dok je najniži MAE ostvario MLR u istom režimu (0.01133). Kombinovani režim dao je veoma blizak rezultat: RF je ostvario RMSE od 0.01374 i MAE od 0.01167. U GAN režimu najniži RMSE ostvarila je neuronska mreža (0.01385), dok je SVR imao najniži MAE (0.01170). Iako kombinovanje svih snimaka nije nadmašilo najbolji rezultat realnog režima, razlika u RMSE-u bila je zanemarljiva. Proseci su dobijeni iz osam LOPO testiranja.

7.1.9.4 Predikcija prinosa suncokreta

Tabela 7.5: Rezultati svih modela za predikciju prinosa suncokreta u 2018. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00003	0.00540	0.00441
	XGBoost	0.00005	0.00681	0.00573
	MLR	0.00015	0.01204	0.00906
	SVR	0.00004	0.00592	0.00485
	MLR_SGD	0.00019	0.01298	0.00963
	NN	0.00003	**0.00533**	**0.00424**
GAN rekonstruisani snimci	RF	0.00003	0.00535	0.00436
	XGBoost	0.00005	0.00677	0.00570
	MLR	0.00076	0.02228	0.01010
	SVR	0.00004	0.00630	0.00523
	MLR_SGD	0.00061	0.01994	0.00905
	NN	0.00031	0.01327	0.00530
Kombinovani snimci	RF	0.00003	0.00537	0.00437
	XGBoost	0.00005	0.00676	0.00569
	MLR	0.00447	0.04595	0.00893
	SVR	0.00005	0.00650	0.00543
	MLR_SGD	0.00506	0.04794	0.00939
	NN	0.00035	0.01378	0.00501

Najpovoljniji rezultat za suncokret u 2018. godini ostvarila je neuronska mreža nad realnim snimcima, sa RMSE od 0.00533 i MAE od 0.00424. RF je bio najbolji model u GAN režimu (RMSE=0.00535, MAE=0.00436) i u kombinovanom režimu (RMSE=0.00537, MAE=0.00437). Razlike između najboljih rezultata bile su male, pa kombinovanje svih snimaka nije donelo praktično poboljšanje. MLR i MLR_SGD bili su posebno nestabilni u visokodimenzionalnom kombinovanom eksperimentu. Procena je zasnovana na samo tri LOPO iteracije.

7.1.10 Rezultati za 2019. godinu

U 2019. godini svi modeli takođe su uspešno završili obuku. Najbolji algoritam zavisio je od kulture i režima ulaznih podataka: MLR je ostvario najmanju ukupnu grešku za

kukuruz, neuronska mreža za soju, dok je SVR bio najuspešniji za pšenicu i suncokret.

7.1.10.1 Predikcija prinosa kukuruza

Tabela 7.6: Rezultati svih modela za predikciju prinosa kukuruza u 2019. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00016	0.01236	0.00967
	XGBoost	0.00018	0.01293	0.01021
	MLR	0.00013	**0.01136**	**0.00876**
	SVR	0.00016	0.01226	0.00964
	MLR_SGD	0.00018	0.01329	0.01057
	NN	0.00014	0.01152	0.00895
GAN rekonstruisani snimci	RF	0.00017	0.01240	0.00968
	XGBoost	0.00019	0.01314	0.01037
	MLR	0.00018	0.01313	0.01043
	SVR	0.00019	0.01312	0.01041
	MLR_SGD	0.00021	0.01415	0.01134
	NN	0.00019	0.01340	0.01070
Kombinovani snimci	RF	0.00017	0.01258	0.00987
	XGBoost	0.00019	0.01338	0.01062
	MLR	0.00018	0.01327	0.01067
	SVR	0.00021	0.01414	0.01141
	MLR_SGD	0.00030	0.01570	0.01307
	NN	0.00017	0.01268	0.01002

Najnižu grešku za kukuruz u 2019. godini ostvario je MLR nad realnim snimcima, sa RMSE od 0.01136 i MAE od 0.00876. Neuronska mreža bila je neposredno iza njega (RMSE=0.01152, MAE=0.00895). U GAN režimu najbolji je bio RF, sa RMSE od 0.01240 i MAE od 0.00968, dok je u kombinovanom režimu isti model ostvario RMSE od 0.01258 i MAE od 0.00987. Rezultati pokazuju da uključivanje svih realnih i svih GAN snimaka nije poboljšalo predikciju kukuruza u odnosu na realni ili GAN režim. Proseci su izračunati preko 12 LOPO test parcela.

Tabela 7.7: Rezultati svih modela za predikciju prinosa soje u 2019. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00003	0.00557	0.00442
	XGBoost	0.00004	0.00599	0.00473
	MLR	0.00310	0.03685	0.01939
	SVR	0.00004	0.00607	0.00498
	MLR_SGD	0.00013	0.00942	0.00661
	NN	0.00010	0.00904	0.00624
GAN rekonstruisani snimci	RF	0.00003	0.00544	0.00432
	XGBoost	0.00004	0.00608	0.00482
	MLR	0.00318	0.03777	0.02030
	SVR	0.00004	0.00646	0.00546
	MLR_SGD	0.00007	0.00801	0.00622
	NN	0.00006	0.00756	0.00580
Kombinovani snimci	RF	0.00003	0.00528	0.00415
	XGBoost	0.00004	0.00600	0.00474
	MLR	0.00266	0.03346	0.01714
	SVR	0.00004	0.00623	0.00522
	MLR_SGD	0.00902	0.05906	0.02910
	NN	0.00002	**0.00477**	**0.00376**

7.1.10.2 Predikcija prinosa soje

Za soju u 2019. godini najnižu grešku ostvarila je neuronska mreža u kombinovanom režimu, sa RMSE od 0.00477 i MAE od 0.00376. Ovaj rezultat je prema RMSE-u bio približno 12% bolji od najboljeg GAN rezultata, koji je ostvario RF (RMSE=0.00544, MAE=0.00432). RF je bio najbolji i nad realnim snimcima (RMSE=0.00557, MAE=0.00442). Kombinovanje svih realnih i svih GAN blokova zato je za ovu kulturu donelo jasno poboljšanje, ali procena ostaje ograničena na samo tri prostorno nezavisne parcele.

7.1.10.3 Predikcija prinosa pšenice

Najbolji rezultat za pšenicu u 2019. godini ostvario je SVR nad realnim snimcima, sa RMSE od 0.00897 i MAE od 0.00714. Isti model bio je najbolji i u kombinovanom režimu (RMSE=0.00928, MAE=0.00750), dok je u GAN režimu najnižu grešku ostvario RF (RMSE=0.00935, MAE=0.00759). Kombinovani režim je neznatno poboljšao rezultat u

Tabela 7.8: Rezultati svih modela za predikciju prinosa pšenice u 2019. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00010	0.00935	0.00753
	XGBoost	0.00011	0.00971	0.00793
	MLR	0.00015	0.01042	0.00826
	SVR	0.00009	**0.00897**	**0.00714**
	MLR_SGD	0.00028	0.01291	0.01104
	NN	0.00009	0.00913	0.00732
GAN rekonstruisani snimci	RF	0.00010	0.00935	0.00759
	XGBoost	0.00011	0.00986	0.00810
	MLR	0.00023	0.01144	0.00885
	SVR	0.00010	0.00938	0.00761
	MLR_SGD	0.00013	0.01009	0.00809
	NN	0.00011	0.00972	0.00784
Kombinovani snimci	RF	0.00010	0.00933	0.00757
	XGBoost	0.00011	0.00980	0.00805
	MLR	0.00013	0.01032	0.00826
	SVR	0.00010	0.00928	0.00749
	MLR_SGD	0.00015	0.01105	0.00899
	NN	0.00011	0.00963	0.00788

odnosu na GAN snimke, ali nije nadmašio realni režim. Rezultati su agregirani preko 10 LOPO iteracija.

7.1.10.4 Predikcija prinosa suncokreta

Za suncokret u 2019. godini SVR je ostvario najnižu grešku u sva tri režima. Najbolji rezultat dobijen je nad realnim snimcima, sa RMSE od 0.00556 i MAE od 0.00421. GAN režim dao je tek neznatno veće vrednosti (RMSE=0.00564, MAE=0.00428), dok je u kombinovanom režimu RMSE iznosio 0.00568, a MAE 0.00432. Razlike su male, ali kombinovanje svih blokova nije poboljšalo rezultat u odnosu na pojedinačne izvore. Procena je zasnovana na osam prostornih test parcela.

Tabela 7.9: Rezultati svih modela za predikciju prinosa suncokreta u 2019. godini.

Eksperiment	**Model**	**MSE**	**RMSE**	**MAE**
Realni snimci	RF	0.00004	0.00577	0.00447
	XGBoost	0.00005	0.00652	0.00517
	MLR	0.00004	0.00597	0.00455
	SVR	0.00003	**0.00556**	**0.00421**
	MLR_SGD	0.00005	0.00719	0.00593
	NN	0.00003	0.00564	0.00429
GAN rekonstruisani snimci	RF	0.00004	0.00577	0.00446
	XGBoost	0.00005	0.00652	0.00517
	MLR	0.00004	0.00644	0.00502
	SVR	0.00003	0.00564	0.00428
	MLR_SGD	0.00005	0.00686	0.00561
	NN	0.00004	0.00572	0.00436
Kombinovani snimci	RF	0.00003	0.00577	0.00445
	XGBoost	0.00005	0.00652	0.00518
	MLR	0.00004	0.00607	0.00468
	SVR	0.00003	0.00568	0.00432
	MLR_SGD	0.00004	0.00624	0.00494
	NN	0.00004	0.00581	0.00444

7.1.11 Poređenje modela i eksperimentalnih režima

Na nivou 24 kombinacije godine, kulture i režima ulaznih podataka, RF je najčešće ostvarivao najnižu grešku. Prema RMSE-u bio je najbolji u 13 kombinacija, SVR u sedam, neuronska mreža u tri, a MLR u jednoj kombinaciji. Prema MAE-u RF je bio najbolji u 13 kombinacija, SVR u sedam, dok su MLR i neuronska mreža bili najbolji u po dve kombinacije. XGBoost i MLR_SGD nisu ostvarili najnižu grešku ni u jednoj kombinaciji, iako je XGBoost u više eksperimenata dao rezultate bliske najbolje rangiranim modelima.

RF je pokazao najveću ukupnu stabilnost, posebno za kukuruz i u većini eksperimenata sa pojedinačnim izvorima snimaka. Bio je najbolji i u više kombinovanih eksperimenata, ali povećanje broja ulaznih obeležja nije uvek dovelo do smanjenja greške.

SVR je bio najuspešniji prema RMSE-u u sedam kombinacija. Pored realnih snimaka

pšenice u 2018. godini, najbolje rezultate ostvario je u kombinovanom eksperimentu za soju 2018. godine, realnom i kombinovanom eksperimentu za pšenicu 2019. godine, kao i u sva tri režima za suncokret 2019. godine. To pokazuje da su RBF jezgro i odabrani parametri regularizacije omogućili dobru generalizaciju u više različitih skupova podataka.

MLR je ostvario najbolji rezultat za kukuruz u 2019. godini nad realnim snimcima, a imao je i najniži MAE za pšenicu u 2018. godini. Međutim, u eksperimentima za soju i u pojedinim visokodimenzionalnim kombinacijama pokazao je znatno veće greške, što ukazuje na osetljivost linearnog modela na multikolinearnost i razlike između parcela.

Neuronska mreža nije pokazivala numeričku divergenciju i bila je najbolja prema RMSE-u u tri kombinacije: za suncokret u 2018. godini nad realnim snimcima, za pšenicu u GAN režimu iste godine i za soju u kombinovanom režimu 2019. godine. Poslednji rezultat predstavlja i najjasnije poboljšanje ostvareno objedinjavanjem svih realnih i svih GAN snimaka.

MLR_SGD je uspešno završio sve eksperimente, ali je najčešće ostvarivao veće greške od ostalih modela. Njegova nestabilnost bila je posebno izražena u pojedinim kombinovanim eksperimentima, što pokazuje da jedinstven skup fiksnih parametara nije jednako pogodan za sve kulture i godine.

Poređenje režima pokazuje da su najbolji ukupni rezultati u šest od osam kombinacija kulture i godine ostvareni korišćenjem realnih snimaka. Izuzetak je soja u obe godine, za koju je najmanja greška ostvarena u ispravljenom kombinovanom režimu: SVR modelom u 2018. i neuronskom mrežom u 2019. godini. Za ostale kulture uključivanje svih realnih i svih GAN snimaka nije nadmašilo najbolji rezultat pojedinačnog izvora. To ukazuje da dodatni blokovi mogu biti korisni kada donose komplementarne informacije, ali mogu i povećati redundantnost ulaznog prostora bez poboljšanja generalizacije.

7.1.12 Interpretacija rezultata u odnosu na vrednosti prinosa

Vrednosti prinosa u ulaznom skupu podataka izražene su u tonama po prostornoj jedinici površine $100\,\mathrm{m}^2$, koja odgovara rasterskoj ćeliji dimenzija $10 \times 10\,\mathrm{m}$. Kako jedan hektar obuhvata površinu od $10\,000\,\mathrm{m}^2$, odnosno 100 rasterskih ćelija dimenzija $10 \times 10\,\mathrm{m}$, vrednosti prinosa izražene u tonama po pikselu mogu se prevesti u $\mathrm{t\,ha}^{-1}$ množenjem sa faktorom 100:

$$Y_{\mathrm{t/ha}} = 100 \cdot Y_{\mathrm{t/piksel}}. \tag{7.1}$$

Na primer, vrednost prinosa od $0.128\,\mathrm{t/piksel}$ odgovara prinosu od $12.8\,\mathrm{t\,ha}^{-1}$. Na isti način, RMSE od $0.01469\,\mathrm{t/piksel}$ odgovara grešci od približno $1.47\,\mathrm{t\,ha}^{-1}$, dok MAE

od 0.01153 t/piksel odgovara prosečnoj apsolutnoj grešci od približno $1.15\,\mathrm{t\,ha^{-1}}$.

Za razliku od RMSE i MAE, koji se pri konverziji u $\mathrm{t\,ha^{-1}}$ množe sa faktorom 100, srednja kvadratna greška menja se kvadratom faktora konverzije. Prema tome, MSE izražen u $(\mathrm{t\,ha^{-1}})^2$ dobija se množenjem originalne vrednosti sa 100^2, odnosno sa 10 000:

$$\mathrm{MSE}_{(\mathrm{t/ha})^2} = 10\,000 \cdot \mathrm{MSE}_{(\mathrm{t/piksel})^2}. \tag{7.2}$$

Tabela 7.10 prikazuje prosečne vrednosti prinosa i najbolje ostvarene vrednosti RMSE i MAE za svaku kulturu i godinu, izražene u $\mathrm{t\,ha^{-1}}$. Najbolji RMSE i najbolji MAE određeni su nezavisno, pa kod pšenice u 2018. godini potiču iz dva različita modela. U poslednje dve kolone prikazan je odnos greške i prosečnog prinosa.

Tabela 7.10: Prosečne vrednosti prinosa i najbolje ostvarene greške predikcije izražene u $\mathrm{t\,ha^{-1}}$.

Godina	**Kultura**	**Prosečan prinos** ($\mathrm{t\,ha^{-1}}$)	**Najbolji RMSE** ($\mathrm{t\,ha^{-1}}$)	**Najbolji MAE** ($\mathrm{t\,ha^{-1}}$)	**RMSE / prinos** (%)	**MAE / prinos** (%)
2018	Kukuruz	12.84	1.47	1.15	11.4	9.0
2018	Soja	3.83	0.46	0.37	12.0	9.5
2018	Pšenica	7.10	1.37	1.13	19.3	16.0
2018	Suncokret	3.90	0.53	0.42	13.7	10.9
2019	Kukuruz	9.76	1.14	0.88	11.6	9.0
2019	Soja	3.23	0.48	0.38	14.8	11.6
2019	Pšenica	6.69	0.90	0.71	13.4	10.7
2019	Suncokret	3.43	0.56	0.42	16.2	12.3

Rezultati prikazani u Tabeli 7.10 pokazuju da su dobijene vrednosti grešaka u skladu sa rasponom ulaznih vrednosti prinosa. Najveće apsolutne greške dobijene su za kukuruz i pšenicu, što je očekivano jer te kulture imaju i najveće prosečne vrednosti prinosa. Soja i suncokret imaju niže apsolutne vrednosti RMSE i MAE, ali to ne znači nužno da su modeli za njih relativno precizniji. Zbog toga je potrebno posmatrati i odnos greške prema prosečnom prinosu.

7.1.12.1 Interpretacija rezultata za 2018. godinu

Prosečan prinos kukuruza u 2018. godini iznosio je približno $12.84\,\mathrm{t\,ha^{-1}}$. Najbolji model ostvario je RMSE od $1.47\,\mathrm{t\,ha^{-1}}$ i MAE od $1.15\,\mathrm{t\,ha^{-1}}$. Relativni RMSE iznosio je

približno 11.4%, a relativni MAE oko 9.0%. Ovakav rezultat može se smatrati razumnim za prostorno nezavisnu LOPO validaciju, u kojoj se model testira na celoj parceli koja nije bila uključena u obuku.

Soja je imala prosečan prinos od približno $3.83\,\mathrm{t\,ha^{-1}}$. Najbolji rezultat ostvaren je u kombinovanom režimu, sa RMSE od približno $0.46\,\mathrm{t\,ha^{-1}}$ i MAE od $0.37\,\mathrm{t\,ha^{-1}}$. Relativni RMSE iznosio je približno 12.0%, a relativni MAE 9.5%. Rezultat je povoljan, ali ga treba tumačiti sa oprezom zbog samo tri dostupne parcele.

Prosečna vrednost prinosa pšenice iznosila je približno $7.10\,\mathrm{t\,ha^{-1}}$. Najbolji RMSE iznosio je $1.37\,\mathrm{t\,ha^{-1}}$, dok je najbolji MAE iznosio $1.13\,\mathrm{t\,ha^{-1}}$. Relativni RMSE od približno 19.3% i relativni MAE od 16.0% predstavljaju najveće relativne greške u 2018. godini. To ukazuje da je prenos modela između parcela za pšenicu bio zahtevniji nego za ostale kulture.

Za suncokret je prosečan prinos iznosio približno $3.90\,\mathrm{t\,ha^{-1}}$. Najbolji RMSE iznosio je $0.53\,\mathrm{t\,ha^{-1}}$, a MAE $0.42\,\mathrm{t\,ha^{-1}}$. Relativni RMSE bio je približno 13.7%, a relativni MAE 10.9%, što ukazuje na umereno dobru preciznost.

Na nivou 2018. godine najbolji relativni rezultat ostvaren je za kukuruz, dok je najslabiji rezultat dobijen za pšenicu. Razlike između kultura potvrđuju da apsolutne vrednosti grešaka ne treba porediti bez uzimanja u obzir prosečnog nivoa prinosa.

7.1.12.2 Interpretacija rezultata za 2019. godinu

Tokom 2019. godine kukuruz je imao prosečan prinos od približno $9.76\,\mathrm{t\,ha^{-1}}$. Najbolji RMSE iznosio je $1.14\,\mathrm{t\,ha^{-1}}$, a MAE $0.88\,\mathrm{t\,ha^{-1}}$. Relativni RMSE iznosio je približno 11.6%, dok je relativni MAE iznosio oko 9.0%. Relativna preciznost bila je veoma slična rezultatu za kukuruz u 2018. godini.

Prosečan prinos soje iznosio je približno $3.23\,\mathrm{t\,ha^{-1}}$. Najbolji rezultat ostvaren je neuronskom mrežom u kombinovanom režimu, sa RMSE od približno $0.48\,\mathrm{t\,ha^{-1}}$ i MAE od $0.38\,\mathrm{t\,ha^{-1}}$. Relativni RMSE bio je približno 14.8%, a relativni MAE 11.6%. U poređenju sa 2018. godinom, relativna greška je bila nešto veća, ali je razlika manja nego što su pokazivali prethodni rezultati.

Pšenica je imala prosečan prinos od približno $6.69\,\mathrm{t\,ha^{-1}}$. Najbolji RMSE iznosio je $0.90\,\mathrm{t\,ha^{-1}}$, dok je MAE iznosio $0.71\,\mathrm{t\,ha^{-1}}$. Relativni RMSE od 13.4% bio je znatno povoljniji nego u 2018. godini, kada je iznosio približno 19.3%. To ukazuje da su odnosi između satelitskih obeležja i prinosa pšenice bili prenosiviji između parcela u 2019. godini.

Kod suncokreta je prosečan prinos iznosio približno $3.43\,\mathrm{t\,ha^{-1}}$. Najbolji RMSE iznosio je $0.56\,\mathrm{t\,ha^{-1}}$, a MAE $0.42\,\mathrm{t\,ha^{-1}}$. Relativni RMSE bio je približno 16.2%, dok

je relativni MAE iznosio 12.3%. U poređenju sa 2018. godinom, relativna greška bila je nešto veća.

Na nivou 2019. godine najbolji relativni rezultat ostvaren je za kukuruz. Pšenica je takođe pokazala umereno dobre rezultate, dok su veće relativne greške dobijene za soju i suncokret.

7.1.12.3 Poređenje kultura i godina

Najveće apsolutne greške ne predstavljaju nužno i najslabije relativne rezultate. Kukuruz je imao RMSE između približno 1.14 i 1.47 $\mathrm{t\,ha^{-1}}$, što je više nego kod ostalih kultura. Međutim, zbog visokog prosečnog prinosa, relativni RMSE za kukuruz iznosio je samo oko 11.4–11.6%.

Soja i suncokret imali su apsolutne RMSE vrednosti manje od 0.60 $\mathrm{t\,ha^{-1}}$, ali je zbog nižeg prosečnog prinosa njihova relativna greška bila veća. U 2019. godini relativni RMSE iznosio je približno 14.8% za soju i 16.2% za suncokret.

Najpovoljniji rezultat u odnosu na prosečan prinos ostvaren je za kukuruz u 2018. godini, sa relativnim RMSE od približno 11.4%, a neposredno zatim za kukuruz u 2019. godini sa 11.6%. Najveća relativna greška dobijena je za pšenicu u 2018. godini, gde je RMSE iznosio približno 19.3% prosečnog prinosa.

7.1.12.4 Ocena smislenosti dobijenih rezultata

Dobijene vrednosti MSE, RMSE i MAE imaju smisla u odnosu na ulazne vrednosti prinosa. Male decimalne vrednosti u originalnim rezultatima posledica su toga što je prinos izražen u tonama po prostornoj jedinici površine 100 $\mathrm{m^2}$, a ne direktno u $\mathrm{t\,ha^{-1}}$. Nakon konverzije, greške dobijaju agronomski razumljive vrednosti.

Najbolji modeli ostvarili su RMSE između približno 0.46 $\mathrm{t\,ha^{-1}}$ za soju u 2018. godini i 1.47 $\mathrm{t\,ha^{-1}}$ za kukuruz u 2018. godini. Najbolje vrednosti MAE kretale su se između približno 0.37 $\mathrm{t\,ha^{-1}}$ i 1.15 $\mathrm{t\,ha^{-1}}$. U relativnom smislu, RMSE je iznosio između približno 11.4% i 19.3% prosečnog prinosa, dok je MAE iznosio između približno 9.0% i 16.0%.

S obzirom na to da je evaluacija sprovedena primenom LOPO validacije, pri kojoj se model testira na potpuno novoj parceli, ovakve greške mogu se smatrati realističnim. LOPO postupak je znatno zahtevniji od slučajne podele piksela, jer model mora da prenese naučene odnose na prostornu jedinicu koja se može razlikovati prema osobinama zemljišta, agrotehničkim merama, fenološkom razvoju i raspodeli prinosa.

Ipak, preciznost nije bila jednaka za sve kulture i godine. Relativne greške od približno 15% ili više, dobijene za pšenicu u 2018. godini, soju u 2019. godini i

suncokret u 2019. godini, ukazuju na ograničenu prenosivost modela između parcela. U tim slučajevima dodatna obeležja koja opisuju osobine zemljišta, meteorološke uslove, topografiju i agrotehničke postupke mogla bi doprineti smanjenju greške.

Rezultati pokazuju da su modeli relativno najpreciznji za kukuruz u obe godine, dok su veće relativne greške zabeležene za pšenicu u 2018. godini, soju u 2019. godini i suncokret u 2019. godini. Najvažnije ograničenje predstavlja broj prostorno nezavisnih parcela, posebno kod soje i suncokreta u 2018. godini, gde su bile dostupne samo tri parcele.

Uspešnost predikcije zavisila je istovremeno od kulture, proizvodne godine, izbora modela i izvora ulaznih snimaka. RF je pokazao najveću ukupnu stabilnost, dok su SVR, MLR i neuronska mreža bili najbolji u pojedinim specifičnim kombinacijama. MLR_SGD je nakon optimizacije uspešno završio sva izvršavanja, ali je uglavnom ostvarivao veće greške. Kombinovanje realnih i GAN snimaka donelo je jasno poboljšanje za soju u obe godine, dok za ostale kulture povećanje broja ulaznih blokova nije konzistentno smanjilo grešku.

7.1.13 Evaluacija generalizacije modela na podacima iz različitih izvora

Jedan od važnih aspekata procene robusnosti razvijenih modela jeste analiza njihove sposobnosti generalizacije u uslovima kada se treniranje i testiranje vrše na podacima koji potiču iz različitih vremenskih ili produkcionih izvora. U ovom eksperimentu modeli su trenirani na podacima iz 2018. godine, dok je evaluacija izvršena na podacima iz 2019. godine. Na ovaj način simuliran je realističniji scenario primene, u kome model treba da prenese naučene obrasce na podatke koji nisu samo prostorno, već i vremenski i uslovno različiti u odnosu na skup za obuku.

U eksperimentu su upoređena dva modela, *Random Forest* (RF) i *Long Short-Term Memory* (LSTM), pri čemu su korišćene četiri različite grupe ulaznih podataka: satelitski podaci (*sat*), kombinacija satelitskih i zemljišnih podataka (*sat+soil*), kombinacija satelitskih i meteoroloških podataka (*sat+meteo*) i objedinjeni skup svih raspoloživih podataka (*sat+soil+meteo*).

Rezultati prikazani u Tabeli 7.11 ukazuju na nekoliko važnih zaključaka. Pre svega, model *Random Forest* ostvaruje konzistentno bolje rezultate u odnosu na LSTM u svim posmatranim kombinacijama ulaznih podataka. Najbolji rezultat RF modela ostvaren je pri korišćenju isključivo satelitskih podataka, gde je dobijena vrednost od 0.01269. Veoma slične vrednosti dobijene su i za ostale kombinacije ulaza, što ukazuje da dodatni zemljišni i meteorološki podaci u ovom eksperimentu nisu doprineli značajnom

Tabela 7.11: Rezultati testiranja modela pri scenariju treniranja na podacima iz 2018. godine i testiranja na podacima iz 2019. godine. Oznaka *sat* odnosi se na satelitske ulazne podatke, *soil* na zemljišne karakteristike, dok *meteo* označava meteorološke promenljive. Vrednosti grešaka izražene su u t/pikselu.

Ulazni podaci	**RF**	**LSTM**
sat	0.01269	0.01391
sat + soil	0.01337	0.01400
sat + meteo	0.01313	0.01386
sat + soil + meteo	0.01308	0.01640

unapređenju performansi modela.

Kod LSTM modela primećuje se nešto slabija i manje stabilna generalizacija. Najbolji rezultat ostvaren je za kombinaciju *sat+meteo* (0.01386), dok je najlošiji rezultat zabeležen pri korišćenju kompletnog skupa ulaznih podataka *all* (0.01640). Ovakav pad performansi može ukazivati na to da LSTM model, u uslovima ograničenog broja uzoraka i heterogenosti ulaznih promenljivih, teže usvaja reprezentacije koje se uspešno prenose na podatke iz druge godine. Drugim rečima, povećanje broja ulaznih karakteristika nije nužno dovelo do bolje prediktivne moći, već je moglo uneti dodatni šum i otežati proces učenja.

Posmatrano sa aspekta ulaznih promenljivih, rezultati sugerišu da satelitski podaci nose najveći deo korisne informacije za predikciju u ovom eksperimentalnom okviru. Dodavanje zemljišnih i meteoroloških podataka nije dovelo do jasnog i sistematskog poboljšanja, ni kod RF ni kod LSTM modela. Jedno moguće objašnjenje je da dodatni izvori podataka nisu dovoljno informativni u odnosu na ciljni problem ili da njihov prostorni i vremenski nivo agregacije nije u potpunosti usklađen sa skalom na kojoj se modeluje prinos.

Dodatno, ovakav rezultat može biti povezan i sa problemom visoke dimenzionalnosti ulaznog prostora. U konkretnom slučaju, broj izlaznih, odnosno referentnih podataka je fiksan i relativno ograničen, dok se dodavanjem satelitskih, zemljišnih i meteoroloških atributa broj ulaznih promenljivih značajno povećava. Takav nesrazmer između broja uzoraka i broja obeležja može dovesti do tzv. prokletstva dimenzionalnosti, pri čemu dodatna obeležja ne doprinose nužno boljoj generalizaciji, već mogu povećati šum, redundantnost i rizik od preprilagođavanja modela.

Zbog toga bi u budućim istraživanjima trebalo razmotriti pažljiviju selekciju i redukciju ulaznih obeležja pre integracije dodatnih izvora podataka. Potencijalni pravci uključuju primenu metoda redukcije dimenzionalnosti, kao što je analiza glavnih kompo-

nenti (PCA), zatim selekciju obeležja na osnovu važnosti promenljivih, kao i iterativne postupke dodavanja ili uklanjanja obeležja, poput forward feature addition i backward feature elimination pristupa. Na taj način bi se moglo ispitati koji skupovi promenljivih zaista doprinose predikciji prinosa, a koji uvode redundantne ili nedovoljno informativne informacije.

Na osnovu sprovedenog eksperimenta može se zaključiti da je *Random Forest* pokazao veću robusnost i bolju sposobnost generalizacije u scenariju prenosa modela između različitih godina, odnosno različitih izvora podataka. Sa druge strane, LSTM model pokazuje veću osetljivost na izbor ulaznih atributa i kompleksnost skupa podataka. Ovi rezultati ukazuju da, u uslovima ograničenih i heterogenih podataka, klasični ansambl pristupi mogu predstavljati pouzdanije rešenje od složenijih dubokih arhitektura.

7.2 Predikcija proteinskog sastava useva

Predikcija proteinskog sastava soje na osnovu satelitskih snimaka predstavlja izrazito kompleksan istraživački problem. Za razliku od zadataka kao što su klasifikacija useva ili procena biomase, sadržaj proteina nije direktno vidljiva osobina useva, već je rezultat složene interakcije genotipa, agroekoloških uslova, režima đubrenja, dostupnosti azota, vodnog stresa, fenološke faze i drugih faktora. Zbog toga se odnos između spektralnog odgovora useva i konačnog proteinskog sastava ne može smatrati jednostavnim ni stabilnim u različitim godinama, lokalitetima i proizvodnim sistemima. Ovakva složenost dodatno otežava generalizaciju modela kada se podaci prikupljaju u različitim agroekološkim uslovima, kao što je bio slučaj u ovom istraživanju, gde su korišćeni podaci iz Italije i Srbije.

U cilju ispitivanja mogućnosti predikcije proteinskog sastava soje, formiran je višestepeni metodološki okvir koji obuhvata: (i) pripremu i ekstrakciju satelitskih podataka iz Sentinel-2 vremenskih serija, (ii) transformaciju vremenske serije satelitskih snimaka u robusnije sezonske indikatore i (iii) razvoj i evaluaciju klasifikacionih modela za procenu nivoa proteinskog sastava. Posebna pažnja posvećena je testiranju modela u uslovima prenosa između različitih izvora podataka, odnosno između različitih lokaliteta i godina, kako bi se procenila stvarna mogućnost generalizacije modela.

7.2.1 Opis podataka, transformacija vremenskih serija i konstrukcija ulaznih obeležja

Za potrebe istraživanja korišćeni su skupovi podataka iz više ogleda uspostavljenih u Italiji i Srbiji. Svaki skup sadržao je tačkaste ili ogledne lokacije sa laboratorijski

određenim sadržajem proteina, kao i pridružene satelitske podatke izvedene iz Sentinel-2 snimaka. Ekstrakcija satelitskih podataka sprovedena je u okruženju Google Earth Engine korišćenjem kolekcije `COPERNICUS/S2_SR_HARMONIZED`. Za svaku tačku uzorkovanja učitan je odgovarajući prostorni objekat, pri čemu je geometrija definisana na osnovu prethodno pripremljenih tabela, a zatim je oko svake tačke formiran bafer poluprečnika 10 m kako bi se smanjio uticaj geolokacione greške i obezbedilo reprezentativnije uzorkovanje piksela. Analiza je sprovedena za vegetacioni period od 1. maja do 31. oktobra u zavinosti od godine kada su ekeprimenti sprovedneni (2017, 2018 i 2020)

Za uklanjanje nepouzdanih opservacija primenjena je maska zasnovana na `SCL` sloju (Scene Classification Layer), pri čemu su iz analize uklonjeni pikseli označeni kao saturisani ili defektni, senke oblaka, oblaci srednje i visoke verovatnoće, cirusi i sneg/led. Pored toga, na nivou cele scene primenjen je uslov da `CLOUDY_PIXEL_PERCENTAGE` bude manji ili jednak 90%, uz istovremeno uklanjanje najproblematičnijih piksela na lokalnom nivou. Ovakav pristup predstavlja kompromis između količine dostupnih podataka i kvaliteta spektralnih informacija.

Iz svakog validnog snimka izdvojeni su osnovni Sentinel-2 kanali $B2$, $B3$, $B4$, $B5$, $B6$, $B7$, $B8$, $B8A$, $B11$ i $B12$, kao i skup izvedenih vegetacionih i spektralnih indeksa: NDVI, GNDVI, NDRE, NDWI, EVI, SAVI, CIRE i MSI. Vrednosti su izračunate nad baferisanim tačkama pomoću srednje vrednosti piksela u prostornoj rezoluciji od 10 m. Rezultat ove faze bila je tabela u kojoj svaki red predstavlja kombinaciju lokacije i datuma snimanja, sa pridruženim spektralnim obeležjima i terenskim podacima o proteinskom sastavu.

Pošto su pojedinačni datumi snimanja između lokaliteta i godina bili neujednačeni, direktno korišćenje višedatumskih kolona oblika `YYYY_MM_DD_FEATURE` nije omogućavalo stabilno objedinjavanje skupova iz različitih izvora. Zbog toga je razvijena procedura agregacije vremenskih serija na mesečni nivo. Prepoznate su sve kolone koje odgovaraju satelitskim kanalima i indeksima, nakon čega su grupisane prema mesecu i tipu promenljive, na primer `06_B5`, `07_NDRE`, `08_MSI`. Za svaku takvu grupu izračunata je srednja vrednost kroz sve raspoložive datume u datom mesecu. Na taj način je višedatumska vremenska serija transformisana u skup robusnijih sezonskih deskriptora koji su uporedivi između godina i lokaliteta.

U modelovanju je korišćen redukovani skup spektralnih obeležja koji obuhvata kanale $B5$, $B6$, $B7$, $B8$, $B8A$, $B11$, $B12$ i indekse NDRE, CIRE, NDWI i MSI. Ovaj izbor zasnovan je na pretpostavci da su upravo red-edge i SWIR opsezi, kao i odgovarajući indeksi, najinformativniji za procenu biohemijskih i fizioloških karakteristika soje povezanih sa sadržajem proteina. Agregacija je sprovedena za mesece od maja do oktobra, čime su obuhvaćene ključne faze vegetacionog razvoja useva.

7.2.2 Klasifikacija proteinskog sastava

Strategija modelovanja zasnivala se na binarnoj klasifikaciji proteinskog sastava, pri čemu je ciljna promenljiva definisana prema praktično relevantnom pragu od 40%. Uzorci sa sadržajem proteina većim od 40% označeni su kao pozitivna klasa, dok su uzorci sa sadržajem manjim ili jednakim 40% označeni kao negativna klasa. Ovakva formulacija problema ima direktan praktični značaj, budući da se soja sa sadržajem proteina iznad 40% može koristiti za ljudsku ishranu, dok se niži nivoi uglavnom posmatraju kao manje pogodni za tu namenu. Time je istraživanje pomereno sa zadatka precizne regresione procene na zadatak donošenja odluke sa jasnim tehnološkim značajem.

Za klasifikaciju je korišćen model RF klasifikator. U modelu je primenjena opcija balansiranja skupa podataka kako bi se ublažio uticaj potencijalne neuravnoteženosti klasa. Evaluacija je sprovedena u režimu *leave-one-source-out*, gde je u svakoj iteraciji jedan izvor podataka korišćen kao nezavisni test skup, dok su preostali izvor korišćena za obuku. Na taj način procenjena je sposobnost modela da generalizuje između lokaliteta i godina. Kao metrike uspešnosti korišćene su tačnost i matrica konfuzije.

Za razliku od klasične nasumične podele na trening i test skup, koja može dovesti do optimistične procene performansi usled prisustva veoma sličnih uzoraka u oba podskupa, u ovom istraživanju naglasak je stavljen na evaluaciju između izvora podataka. Strategija *leave-one-source-out* omogućava strožu i realističniju procenu, jer model u svakoj iteraciji mora da predviđa proteinski sastav na potpuno novom lokalitetu i/ili u novoj godini, bez prethodnog uvida u raspodelu tog skupa. Na taj način moguće je razdvojiti dve situacije: (i) sposobnost modela da prepoznaje obrasce unutar homogenog skupa podataka i (ii) sposobnost modela da te obrasce prenese između heterogenih agroekoloških uslova.

Dobijeni rezultati ukazuju da je predikcija proteinskog sastava na osnovu satelitskih snimaka moguća u određenoj meri, ali da njena uspešnost snažno zavisi od homogenosti ulaznih podataka i od izbora načina formulacije problema. Pokazalo se da regresioni pristup, naročito kada se modeli treniraju na jednom skupu i testiraju na drugom, dovodi do izraženog pada performansi usled razlika između lokaliteta, godina i proizvodnih sistema. Zbog toga su klasifikacioni pristupi, posebno oni zasnovani na praktično relevantnim pragovima ili percentilnim klasama, predstavljali robusniju i interpretabilniju alternativu za procenu proteinskog kvaliteta soje.

7.2.3 Rezultati klasifikacije proteinskog sastava

U okviru binarne klasifikacije cilj je bio razlikovati uzorke sa sadržajem proteina manjim ili jednakim 40% od uzoraka sa sadržajem proteina većim od 40%. Ovakva formulacija

Tabela 7.12: Rezultati binarne klasifikacije proteinskog sastava ($\leq 40\%$ naspram $> 40\%$) u režimu *leave-one-source-out* evaluacije.

Test skup	**Tačnost (%)**
Test 1	0.75
Test 2	0.41
Test 3	0.99
Test 4	1.00

problema ima direktan praktični značaj, jer prag od 40% predstavlja važnu granicu za procenu pogodnosti soje za ljudsku ishranu. Model je evaluiran u režimu *leave-one-source-out*, pri čemu je u svakoj iteraciji jedan skup korišćen kao testni, dok su preostala tri korišćena za obuku.

Rezultati prikazani u Tabeli 7.12 pokazuju izrazito veliku varijabilnost tačnosti između različitih testnih scenarija. Ostvarene vrednosti tačnosti predikcije kretale su se od 0.41 do 1.0000, što ukazuje da uspešnost modela značajno zavisi od konkretnog skupa na kome se vrši testiranje. U jednom slučaju postignuta je tačnost od 0.75, u drugom svega 0.41, dok su u preostala dva scenarija ostvarene veoma visoke vrednosti od 0.99 i 1.0000.

Iako ovakvi rezultati na prvi pogled mogu ukazivati na dobru uspešnost modela, njihova interpretacija zahteva oprez. Visoke vrednosti tačnosti u pojedinim testnim scenarijima ne moraju nužno značiti da model pouzdano razlikuje uzorke sa visokim i niskim sadržajem proteina, već mogu biti posledica neravnomerne raspodele klasa u testnom skupu.

Poglavlje 8

SAPI i njegove primene

Sadržaj

8.1 Mape produktivnosti useva – SAPI (Satellite Agriculture Productivity Index) . . . 127
8.1.1 Ulazni podaci i izračunavanje spektralnih pokazatelja . . . 128
8.1.2 Rekonstrukcija vremenske serije i integral sezonske aktivnosti . . 129
8.1.3 Klasifikacija relativne produktivnosti . . . 129
8.1.4 Višegodišnja agregacija i prostorna interpretacija . . . 130
8.1.5 Poređenje SAPI klasa i prinosa useva u 2018. godini . . . 130
8.1.6 Poređenje SAPI klasa i prinosa useva u 2019. godini . . . 134
8.1.7 Značaj i primena SAPI indeksa . . . 138
8.2 Analiza anomalija produktivnosti povezanih sa štetnim događajima . . 138
8.2.1 Formiranje baze olujnih događaja, geolokacijska standardizacija i povezivanje olujnih događaja sa katastarskim opštinama . . . 140
8.2.2 Izdvajanje statistika produktivnosti po katastarskim opštinama iz SAPI mapa . . . 142
8.2.3 Rezultati analize anomalija u katastarskim opštinama pogođenim olujama . . . 143
8.3 Analiza povezanosti SAPI sa poznatim navodnjavanim površinama . . . 143
8.3.1 Analiza usklađenosti između navodnjavanja i produktivnosti useva zasnovane na SAPI indeksu . . . 145

8.1 Mape produktivnosti useva – SAPI (Satellite Agriculture Productivity Index)

Za prostornu procenu relativne produktivnosti useva razvijen je indeks *Satellite Agriculture Productivity Index* (SAPI), zasnovan na vremenskim serijama Sentinel-2 snimaka i izvedenih spektralnih odziva u prostornoj rezoluciji od 10 m. Osnovni cilj ovog indeksa

jeste da na nivou piksela opiše prostorne razlike u intenzitetu i dinamici vegetacionog razvoja tokom proizvodne sezone, odnosno da izdvoji stabilno manje i više produktivne delove poljoprivrednih parcela. SAPI je generisan za sve kulture za koje su bile dostupne mape klasifikacije useva: kukuruz, pšenicu, soju, šećernu repu, uljanu repicu i suncokret.

8.1.1 Ulazni podaci i izračunavanje spektralnih pokazatelja

Prvi korak u izračunavanju SAPI indeksa podrazumevao je prostorno izdvajanje pikselâ koji pripadaju ciljnoj kulturi pojedinačno. U tu svrhu korišćene su binarne maske dobijene na osnovu prethodno generisanih mapa klasifikacije useva, odnosno za svaku kulturu generisana je posebna binarsna maska. Na taj način analiza je ograničena isključivo na piksele posmatrane kulture, čime je izbegnuto mešanje spektralnog signala između različitih useva i obezbeđena konzistentnost između klasifikacionog i produktivnosnog sloja.

Ovakav pristup je od posebnog značaja u heterogenim agroekološkim uslovima, kao i na parcelama složene prostorne strukture, jer omogućava da se produktivnost procenjuje unutar homogenih proizvodnih površina.

Za svaki validni piksel korišćeni su Sentinel-2 spektralni kanali B01, B02, B03, B04, B05, B06, B07, B08, B8A, B09, B10, B11 i B12, kao i sloj klasifikacije scene (*Scene Classification Layer* – SCL). Na spektralnih kanala izračunat je skup vegetacionih indeksa koji opisuju različite biofizičke aspekte stanja useva: NDVI, EVI, NDVIRE, VARI, SAVI i NDWI. Na osnovu SCL sloja detektovani su pikseli pod oblacima koje je trebalo ukloniti iz dalje analize.

Radi objedinjavanja informacija iz više komplementarnih spektralnih odziva, za svaku opservaciju formiran je kompozitni indeks kao aritmetička sredina navedenih vegetacionih indeksa:

$$SAPI_{raw}(t) = \frac{1}{6}\left(NDVI_t + EVI_t + NDVIRE_t + VARI_t + SAVI_t + NDWI_t\right), \quad (8.1)$$

$SAPI_{raw}(t)$ predstavlja trenutnu vrednost kompozitnog spektralnog signala za posmatrani piksel u trenutku t.

Na ovaj način dobijen je jedinstven pokazatelj koji ne zavisi od samo jednog vegetacionog indeksa, već integriše više aspekata vegetacionog razvoja, uključujući opštu biomasu, fotosintetsku aktivnost, odziv u red-edge domenu, uticaj zemljišta i prisustvo vode u vegetaciji.

Kontrola optičkih snimaka sprovedena je korišćenjem SCL sloja. Pikseli označeni kao oblaci, senke oblaka, sneg i druge nepoželjne klase isključeni su iz daljeg proračuna.

Dodatno, primenjena je morfološka dilatacija maske kako bi se uklonili i rubni pikseli u zoni oblaka i senki, odnosno pikseli kod kojih klasifikacija scene često ne uspeva u potpunosti da obuhvati kontaminirani deo signala.

Ovakav postupak smanjuje verovatnoću da preostali artefakti utiču na izračunate vrednosti vegetacionih indeksa i doprinosi većoj pouzdanosti konačne procene produktivnosti.

8.1.2 Rekonstrukcija vremenske serije i integral sezonske aktivnosti

Nakon uklanjanja piksela pod oblacima, vremenska serija za svaki piksel transformisana je u vremensku seriju vegetacionog signala. Posmatrani datumi snimanja raspoređeni su unutar vegetacionog perioda, definisanog od aprila do oktobra. Nedostajuće vrednosti između raspoloživih satelitskih opservacija popunjene su linearnom interpolacijom, čime je dobijena vremenska serija pogodna za integralnu analizu.

Na rekonstruisanoj sezonskoj krivoj izračunavana je površina ispod krive (*Area Under the Curve* – AUC), koja predstavlja integrisanu meru vegetacione aktivnosti tokom cele sezone:

$$AUC_SAPI_i = \int_{t_{start}}^{t_{end}} SAPI_{raw,i}(t)\, dt, \tag{8.2}$$

gde AUC_SAPI_i predstavlja integralnu vrednost za piksel i, dok su t_{start} i t_{end} početak i kraj posmatranog vegetacionog perioda.

U diskretnoj implementaciji, kada su vrednosti poznate za niz vremenskih trenutaka, integral je moguće aproksimirati trapeznim pravilom:

$$AUC_SAPI_i \approx \sum_{k=1}^{n-1} \frac{SAPI_{raw,i}(t_k) + SAPI_{raw,i}(t_{k+1})}{2} (t_{k+1} - t_k). \tag{8.3}$$

Veće vrednosti integrala ukazuju na duže trajanje i/ili veći intenzitet aktivne vegetacije, što se u okviru ovog pristupa tumači kao indikator veće relativne produktivnosti posmatranog piksela.

8.1.3 Klasifikacija relativne produktivnosti

Dobijene integralne vrednosti prevedene su u diskretnu skalu relativne produktivnosti sa 10 klasa. Klasifikacija je izvršena na osnovu percentilnih pragova raspodele AUC_SAPI vrednosti, pri čemu su korišćeni 10., 20., 30., 40., 50., 60., 70., 80. i 90. percentil.

Formalno, klasa produktivnosti za piksel i može se definisati na sledeći način:

$$C_i = \begin{cases} 1, & \mathrm{A}UC_SAPI_i \leq P_{10}, \\ 2, & \mathrm{P}_{10} < AUC_SAPI_i \leq P_{20}, \\ 3, & \mathrm{P}_{20} < AUC_SAPI_i \leq P_{30}, \\ \vdots & \vdots \\ 10, & \mathrm{A}UC_SAPI_i > P_{90}. \end{cases} \tag{8.4}$$

gde je C_i klasa produktivnosti piksela i, a $P_{10}, P_{20}, \ldots, P_{90}$ percentilni pragovi raspodele integralnih vrednosti.

Na ovaj način klasa 1 označava najniži, a klasa 10 najviši nivo relativne produktivnosti unutar analiziranog skupa podataka. Prednost ovakvog pristupa ogleda se u robusnoj relativnoj interpretaciji prostorne varijabilnosti, nezavisno od apsolutnog raspona vrednosti i međugodišnjih promena u intenzitetu vegetacije.

8.1.4 Višegodišnja agregacija i prostorna interpretacija

Pored godišnjih mapa produktivnosti, formirane su i višegodišnje prosečne mape produktivnosti parcela. Ove mape dobijene su objedinjavanjem godišnjih SAPI slojeva za istu kulturu kroz više sezona, nakon čega je za svaki piksel izračunata prosečna vrednost indeksa:

$$\overline{SAPI}_i = \frac{1}{N} \sum_{y=1}^{N} C_{i,y}, \tag{8.5}$$

gde $\overline{SAPI}_i$ predstavlja višegodišnju prosečnu produktivnost piksela i, $C_{i,y}$ klasu produktivnosti tog piksela u godini y, a N broj raspoloživih sezona.

Na ovaj način ublažen je uticaj međugodišnje klimatske varijabilnosti, pomerenih rokova setve i drugih kratkoročnih faktora, dok su istovremeno izdvojeni stabilni prostorni obrasci proizvodnog potencijala. Višegodišnje prosečne mape su zbog toga pogodnije za identifikaciju trajnih upravljačkih zona od pojedinačnih sezonskih mapa, koje mogu biti pod snažnijim uticajem specifičnosti konkretne godine.

8.1.5 Poređenje SAPI klasa i prinosa useva u 2018. godini

Radi provere odnosa između SAPI klasifikacije i ostvarenog prinosa, izvršeno je poređenje za četiri analizirane kulture u 2018. godini: kukuruz, soju, suncokret i pšenicu. SAPI podaci predstavljeni su rasterom sa klasama od 1 do 10, dok su podaci o prinosu dati u obliku vektorskih poligona koji odgovaraju Sentinel-2 pikselima prostorne rezolucije 10 m. Za svaki piksel određena je pripadajuća SAPI klasa, nakon čega su po klasama

izračunati broj piksela, prosečan prinos i medijana prinosa. Prinos je izražen u tonama po pikselu, pri čemu jedan piksel obuhvata površinu od 100 m^2.

Tabela 8.1: Broj piksela i prinos kukuruza po SAPI klasama u 2018. godini

SAPI	**Broj piksela**	**Prosečan prinos [t/piksel]**	**Medijana [t/piksel]**
1	4 002	0,1094	0,1091
2	5 394	0,1139	0,1141
3	3 275	0,1178	0,1190
4	2 967	0,1193	0,1208
5	3 771	0,1194	0,1207
6	4 930	0,1228	0,1236
7	7 083	0,1262	0,1261
8	14 669	0,1337	0,1339
9	22 169	0,1383	0,1396
10	2 952	0,1259	0,1264
Ukupno	**71 212**	–	–

Rezultati za kukuruz pokazuju postepeno povećanje prosečnog prinosa od nižih ka višim SAPI klasama. U prvoj klasi prosečan prinos iznosio je 0,1094 t/pikselu, dok je u devetoj klasi dostignuta najveća vrednost od 0,1383 t/pikselu. Sličan obrazac uočava se i kod medijane, koja raste sa 0,1091 t/pikselu u prvoj na 0,1396 t/pikselu u devetoj klasi. Ovakva saglasnost prosečne vrednosti i medijane ukazuje da rast prinosa nije posledica samo pojedinačnih ekstremnih vrednosti, već predstavlja stabilan obrazac unutar klasa.

U desetoj SAPI klasi zabeležen je pad prosečnog prinosa na 0,1259 t/pikselu. Ovaj rezultat pokazuje da najviša SAPI klasa nije nužno povezana sa najvećim ostvarenim prinosom. Odstupanje može biti povezano sa manjim brojem piksela u ovoj klasi, lokalnim razlikama u uslovima proizvodnje ili činjenicom da SAPI opisuje relativni položaj vegetacionog signala, a ne apsolutnu vrednost prinosa. Najveći broj piksela kukuruza nalazi se u klasama 8 i 9, što dodatno potvrđuje da je najveći deo analiziranih površina pripadao višim SAPI kategorijama.

Kod soje se ne uočava isti rastući obrazac kao kod kukuruza. Prosečan prinos u nižim i srednjim klasama bio je relativno stabilan i kretao se od 0,0394 do 0,0405 t/pikselu. Najveća prosečna vrednost zabeležena je u četvrtoj SAPI klasi, dok je u klasama 6, 7 i 8 došlo do smanjenja prinosa. Najniža prosečna vrednost iznosila je 0,0363 t/pikselu u sedmoj klasi.

Medijane prinosa potvrđuju sličan obrazac. U klasama 2–5 medijana se kretala oko 0,041 t/pikselu, dok je u sedmoj i osmoj klasi pala ispod 0,037 t/pikselu. U najvišim

Tabela 8.2: Broj piksela i prinos soje po SAPI klasama u 2018. godini

SAPI	Broj piksela	Prosečan prinos [t/piksel]	Medijana [t/piksel]
1	1 283	0,0394	0,0396
2	1 846	0,0397	0,0407
3	2 428	0,0400	0,0407
4	2 098	0,0405	0,0411
5	2 349	0,0401	0,0411
6	2 642	0,0380	0,0388
7	4 429	0,0363	0,0361
8	4 177	0,0369	0,0366
9	1 977	0,0380	0,0378
10	700	0,0385	0,0389
Ukupno	**23 929**	–	–

klasama dolazi do blagog oporavka, ali vrednosti ostaju niže u odnosu na srednje SAPI klase. Najveći broj piksela soje nalazi se u klasama 7 i 8, što znači da značajan deo analiziranih površina pripada klasama sa nešto nižim prosečnim prinosom. Dobijeni rezultati ukazuju da za soju SAPI klase u 2018. godini nisu pratile jednostavan rast prinosa od nižih ka višim vrednostima.

Tabela 8.3: Broj piksela i prinos suncokreta po SAPI klasama u 2018. godini

SAPI	Broj piksela	Prosečan prinos [t/piksel]	Medijana [t/piksel]
1	18 989	0,0394	0,0400
2	9 262	0,0391	0,0394
3	4 455	0,0386	0,0383
4	3 411	0,0387	0,0382
5	3 011	0,0385	0,0382
6	2 456	0,0384	0,0382
7	2 061	0,0381	0,0380
8	1 990	0,0379	0,0376
9	4 715	0,0395	0,0398
10	1 675	0,0402	0,0407
Ukupno	**52 025**	–	–

Za suncokret su razlike u prosečnom prinosu između SAPI klasa bile veoma male. Vrednosti su se kretale u uskom opsegu od 0,0379 do 0,0402 t/pikselu. Od prve do osme klase uočava se blago smanjenje prosečnog prinosa, nakon čega vrednosti ponovo rastu

u devetoj i desetoj klasi. Najveća prosečna vrednost zabeležena je u desetoj klasi, dok je najniža vrednost zabeležena u osmoj klasi.

Medijane su bile veoma bliske prosečnim vrednostima i pratile su isti obrazac. To pokazuje da unutar pojedinačnih klasa nije bilo izraženog uticaja ekstremnih vrednosti. Posebno je značajno da je najveći broj piksela suncokreta, čak 18 989, pripadao prvoj SAPI klasi, dok su više klase bile znatno slabije zastupljene. Zbog malih razlika u prinosu među klasama, SAPI je u ovom slučaju bolje opisivao raspodelu vegetacionog signala nego jasno razdvajanje zona različitog prinosa.

Tabela 8.4: Broj piksela i prinos pšenice po SAPI klasama u 2018. godini

SAPI	**Broj piksela**	**Prosečan prinos [t/piksel]**	**Medijana [t/piksel]**
1	6 576	0,0693	0,0664
2	4 553	0,0658	0,0645
3	5 063	0,0682	0,0668
4	5 033	0,0720	0,0707
5	4 089	0,0754	0,0736
6	3 344	0,0790	0,0764
7	3 099	0,0859	0,0827
8	2 769	0,0847	0,0823
9	527	0,0769	0,0743
10	1 410	0,0741	0,0721
Ukupno	**36 463**	–	–

Kod pšenice se nakon početnog pada između prve i druge klase uočava postepeno povećanje prinosa do sedme SAPI klase. Prosečan prinos raste sa 0,0658 t/pikselu u drugoj klasi na 0,0859 t/pikselu u sedmoj klasi, gde je zabeležena najveća vrednost. U osmoj klasi prinos ostaje veoma sličan, dok u devetoj i desetoj klasi dolazi do smanjenja.

Medijane prinosa pokazuju isti trend. Najveća medijana zabeležena je u sedmoj klasi i iznosila je 0,0827 t/pikselu, dok je u osmoj klasi bila gotovo jednaka. Pad u devetoj i desetoj klasi ukazuje da odnos između SAPI klase i prinosa pšenice nije potpuno linearan. Takođe, deveta klasa sadrži samo 527 piksela, pa njenu prosečnu vrednost treba tumačiti sa većim oprezom u odnosu na klase sa znatno većim brojem opažanja.

Poređenjem sve četiri kulture može se uočiti da se odnos između SAPI klasa i prinosa razlikuje u zavisnosti od vrste useva. Najjasniji rast prinosa kroz SAPI klase zabeležen je kod kukuruza, gde prosečne vrednosti i medijane rastu gotovo neprekidno do devete klase. Kod pšenice je takođe prisutan rast u većem delu raspona, ali se maksimalne vrednosti javljaju u sedmoj i osmoj klasi, nakon čega sledi pad.

Kod soje i suncokreta razlike među SAPI klasama bile su manje izražene. Soja pokazuje više prosečne prinose u nižim i srednjim klasama, a niže vrednosti u klasama 6–8. Kod suncokreta se prosečan prinos menja u veoma uskom rasponu, zbog čega pojedinačne SAPI klase ne razdvajaju jasno zone različitog prinosa.

Rezultati potvrđuju da SAPI treba posmatrati kao relativni pokazatelj sezonske vegetacione dinamike, a ne kao direktnu procenu prinosa. Njegova povezanost sa ostvarenim prinosom zavisi od kulture, fenološkog razvoja, vremenskih uslova, osobina zemljišta i primenjene agrotehnike. Zbog toga je za pravilno tumačenje SAPI klasa neophodno posmatrati svaku kulturu zasebno i rezultate potvrditi kroz više proizvodnih sezona.

8.1.6 Poređenje SAPI klasa i prinosa useva u 2019. godini

Radi provere odnosa između SAPI klasifikacije i ostvarenog prinosa, izvršeno je poređenje za četiri analizirane kulture u 2019. godini: kukuruz, soju, suncokret i pšenicu. SAPI podaci predstavljeni su rasterom sa klasama od 1 do 10, dok su podaci o prinosu dati u obliku vektorskih poligona koji odgovaraju Sentinel-2 pikselima prostorne rezolucije 10 m. Za svaki piksel određena je pripadajuća SAPI klasa, nakon čega su po klasama izračunati broj piksela, prosečan prinos i medijana prinosa. Prinos je izražen u tonama po pikselu, pri čemu jedan piksel obuhvata površinu od 100 m^2.

Tabela 8.5: Broj piksela i prinos kukuruza po SAPI klasama u 2019. godini

SAPI	**Broj piksela**	**Prosečan prinos [t/piksel]**	**Medijana [t/piksel]**
1	11 751	0,0712	0,0719
2	11 518	0,0856	0,0876
3	11 680	0,0975	0,1000
4	10 598	0,1044	0,1071
5	10 117	0,1064	0,1082
6	6 483	0,1103	0,1117
7	4 465	0,1110	0,1116
8	4 168	0,1123	0,1136
9	3 625	0,1166	0,1173
10	899	0,1192	0,1218
Ukupno	**75 304**	–	–

Rezultati za kukuruz pokazuju jasno i gotovo neprekidno povećanje prinosa od nižih ka višim SAPI klasama. Prosečan prinos raste sa 0,0712 t/pikselu u prvoj klasi na

0,1192 t/pikselu u desetoj klasi. Najizraženiji porast javlja se između prve i četvrte klase, dok se u višim klasama povećanje nastavlja sporijim tempom. Ovakav obrazac ukazuje na postepeno zasićenje, odnosno na manju razliku u prinosu između susednih viših SAPI kategorija.

Medijana prinosa prati isti trend i povećava se sa 0,0719 na 0,1218 t/pikselu. Bliskost prosečnih vrednosti i medijana pokazuje da dobijeni trend nije posledica samo malog broja ekstremnih vrednosti. Najveći broj piksela nalazi se u klasama 1–5, dok je deseta klasa zastupljena sa svega 899 piksela. Zbog manjeg broja opažanja rezultat za najvišu klasu treba tumačiti sa nešto većim oprezom, iako je obrazac rasta prinosa dosledan kroz čitav raspon.

Tabela 8.6: Broj piksela i prinos soje po SAPI klasama u 2019. godini

SAPI	**Broj piksela**	**Prosečan prinos [t/piksel]**	**Medijana [t/piksel]**
1	107	0,0137	0,0129
2	467	0,0187	0,0194
3	826	0,0220	0,0222
4	1 581	0,0248	0,0254
5	2 302	0,0266	0,0273
6	3 162	0,0282	0,0287
7	3 712	0,0310	0,0312
8	5 110	0,0352	0,0358
9	6 012	0,0387	0,0390
10	836	0,0391	0,0396
Ukupno	**24 115**	–	–

Kod soje je zabeležen veoma izražen rast prosečnog prinosa sa povećanjem SAPI klase. Prosečan prinos u prvoj klasi iznosio je 0,0137 t/pikselu, dok je u desetoj klasi dostigao 0,0391 t/pikselu. Vrednost se na taj način povećala gotovo tri puta između krajnjih kategorija. Rast je postepen kroz sve klase, bez pada između susednih kategorija, što ukazuje na veoma pravilno razdvajanje zona različitog prinosa.

Medijana prinosa potvrđuje isti obrazac i povećava se sa 0,0129 na 0,0396 t/pikselu. Najveći broj piksela soje pripada osmoj i devetoj klasi, dok su prve dve klase slabo zastupljene. Posebno prva klasa sadrži samo 107 piksela, zbog čega njena prosečna vrednost može biti manje stabilna. Ipak, doslednost rasta kroz preostale klase pokazuje da je u 2019. godini viša SAPI kategorija uglavnom odgovarala većem ostvarenom prinosu soje.

Za suncokret se uočava postepeno povećanje prosečnog prinosa od prve do devete

Tabela 8.7: Broj piksela i prinos suncokreta po SAPI klasama u 2019. godini

SAPI	Broj piksela	Prosečan prinos [t/piksel]	Medijana [t/piksel]
1	1 470	0,0261	0,0245
2	2 419	0,0293	0,0289
3	2 401	0,0313	0,0317
4	2 932	0,0321	0,0327
5	4 919	0,0333	0,0339
6	5 050	0,0354	0,0361
7	5 590	0,0365	0,0370
8	5 091	0,0368	0,0372
9	4 068	0,0372	0,0377
10	1 164	0,0359	0,0366
Ukupno	**35 104**	–	–

SAPI klase. Prosečna vrednost raste sa 0,0261 t/pikselu u prvoj klasi na 0,0372 t/pikselu u devetoj klasi. Najizraženije povećanje prisutno je u nižim i srednjim klasama, dok se od sedme do devete klase vrednosti menjaju veoma malo.

U desetoj klasi zabeležen je blag pad prosečnog prinosa na 0,0359 t/pikselu. Medijana pokazuje gotovo isti obrazac, sa maksimumom od 0,0377 t/pikselu u devetoj klasi i padom na 0,0366 t/pikselu u desetoj. Ovakav rezultat može ukazivati na zasićenje vegetacionog signala ili na uticaj drugih faktora koji ograničavaju prinos na površinama sa najvišim SAPI vrednostima. Najveći broj piksela nalazi se u klasama 6–8, dok su krajnje klase slabije zastupljene.

Kod pšenice je između prve i druge SAPI klase zabeležen početni pad prosečnog prinosa sa 0,0609 na 0,0572 t/pikselu. Nakon druge klase vrednosti kontinuirano rastu i u desetoj klasi dostižu maksimum od 0,0738 t/pikselu. Rast je izraženiji između druge i šeste klase, dok se u klasama 6–9 javlja postepeno zasićenje i veoma male razlike između susednih kategorija.

Medijana prinosa prati isti obrazac. Najniža vrednost zabeležena je u drugoj klasi i iznosila je 0,0532 t/pikselu, dok je najveća medijana od 0,0727 t/pikselu ostvarena u desetoj klasi. Najveći broj piksela nalazi se u klasama 7 i 8, a deseta klasa je znatno slabije zastupljena. Uprkos početnom odstupanju u drugoj klasi, opšti obrazac pokazuje da više SAPI klase uglavnom odgovaraju višim vrednostima prinosa pšenice.

Poređenje sve četiri kulture pokazuje da je u 2019. godini odnos između SAPI klasa i prinosa bio ujednačeniji nego u prethodnoj sezoni. Kod kukuruza i soje zabeležen je gotovo neprekidan rast prosečnog prinosa i medijane od najniže ka najvišoj SAPI klasi.

Tabela 8.8: Broj piksela i prinos pšenice po SAPI klasama u 2019. godini

SAPI	Broj piksela	Prosečan prinos [t/piksel]	Medijana [t/piksel]
1	5 854	0,0609	0,0604
2	7 020	0,0572	0,0532
3	7 443	0,0625	0,0594
4	6 031	0,0668	0,0641
5	6 342	0,0687	0,0670
6	7 469	0,0704	0,0697
7	7 996	0,0711	0,0707
8	8 157	0,0714	0,0707
9	6 512	0,0718	0,0711
10	1 306	0,0738	0,0727
Ukupno	**64 130**	–	–

Posebno izražen obrazac prisutan je kod soje, kod koje se prosečan prinos između prve i desete klase povećao gotovo tri puta.

Suncokret je pokazao postepen rast do devete klase, nakon čega je u desetoj klasi došlo do blagog smanjenja. Kod pšenice je prisutno početno odstupanje u drugoj klasi, ali se od druge do desete klase uočava kontinuirano povećanje prinosa. U višim klasama kukuruza, suncokreta i pšenice razlike između susednih kategorija postaju manje, što može ukazivati na zasićenje vegetacionog signala ili na sve veći uticaj drugih proizvodnih faktora.

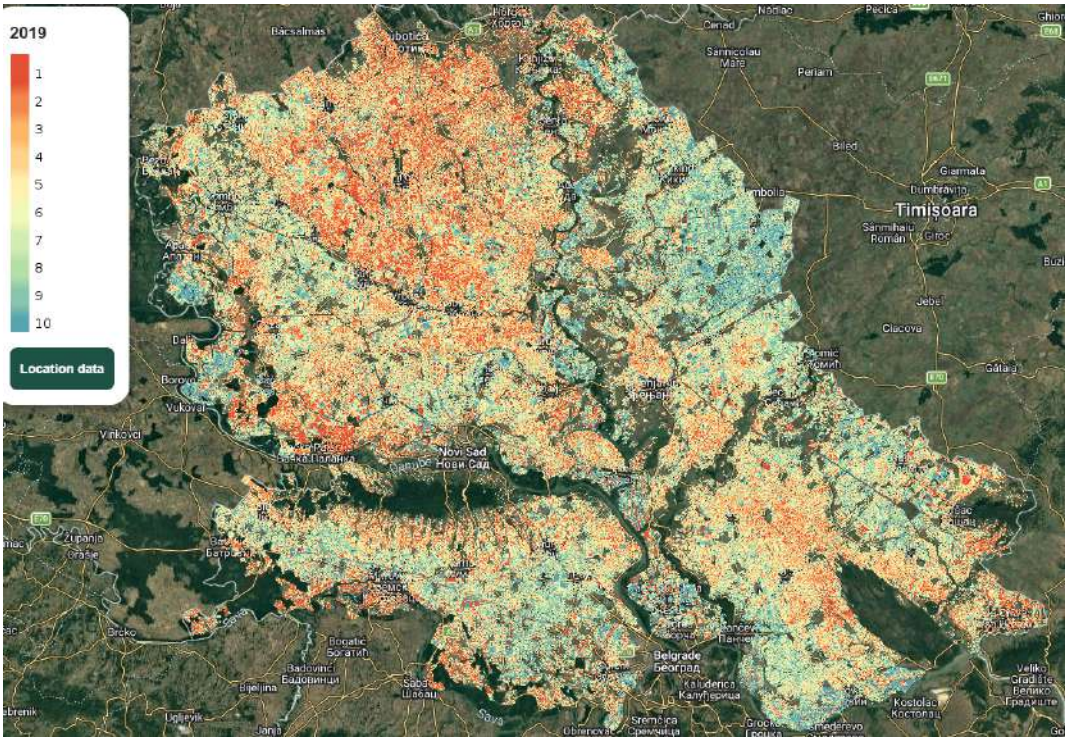
Raspodela broja piksela među klasama razlikovala se u zavisnosti od kulture. Kukuruz je bio najzastupljeniji u nižim i srednjim klasama, soja u višim klasama 8 i 9, suncokret u srednjim i višim klasama 5–8, dok je pšenica bila relativno ravnomerno raspoređena kroz veći deo raspona. Slabija zastupljenost krajnjih klasa kod pojedinih kultura zahteva oprez pri tumačenju njihovih prosečnih vrednosti.

Dobijeni rezultati pokazuju da su SAPI klase u 2019. godini uspešno opisivale relativne razlike u ostvarenom prinosu kod većine analiziranih kultura. Ipak, SAPI ne predstavlja direktnu procenu prinosa, već relativni pokazatelj zasnovan na sezonskoj dinamici vegetacionog signala. Na konačni prinos istovremeno utiču meteorološki uslovi, osobine zemljišta, primenjena agrotehnika, dostupnost vode i lokalne stresne pojave. Zbog toga rezultate treba tumačiti kao potvrdu prostorne saglasnosti SAPI klasifikacije sa prinosom, a ne kao dokaz da SAPI samostalno određuje ili u potpunosti objašnjava ostvareni prinos.

8.1.7 Značaj i primena SAPI indeksa

Posebna vrednost SAPI pristupa ogleda se u tome što omogućava standardizovanu procenu relativne produktivnosti unutar iste kulture na velikom prostornom obuhvatu. Pošto je indeks zasnovan na fenološki integrisanom satelitskom signalu, on ne opisuje samo trenutno stanje useva u jednom trenutku, već sumarni odgovor vegetacije tokom cele proizvodne sezone.

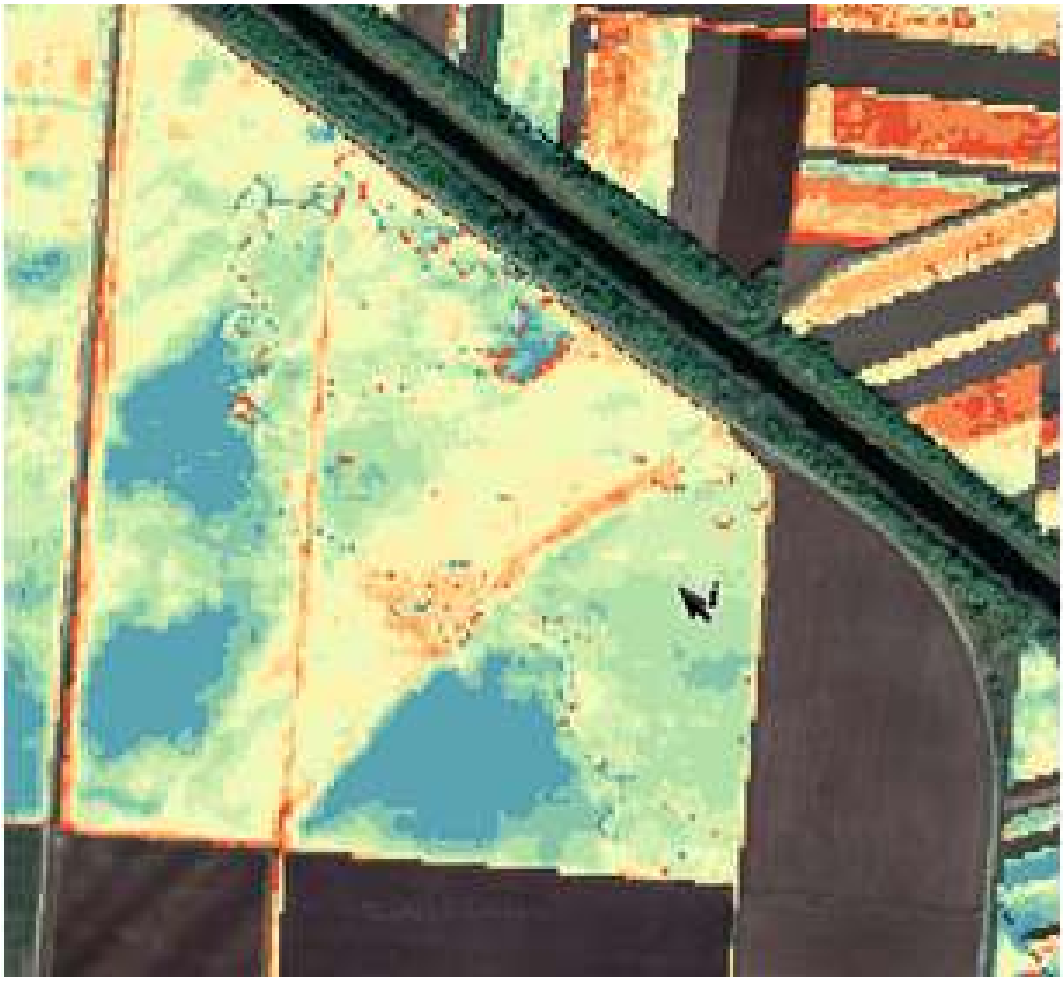
Na taj način SAPI predstavlja prostorni indikator relativnog proizvodnog potencijala, koji može biti korišćen za identifikaciju stabilnih zona visokog i niskog potencijala, analizu unutrašnje heterogenosti parcela, podršku zoniranju u preciznoj poljoprivredi, kao i za dalje povezivanje sa prinosom, zemljišnim osobinama i upravljačkim praksama.



Slika 8.1: Grafički prikaz mape produktivnosti useva za 2019. godinu (SAPI).

8.2 Analiza anomalija produktivnosti povezanih sa štetnim događajima

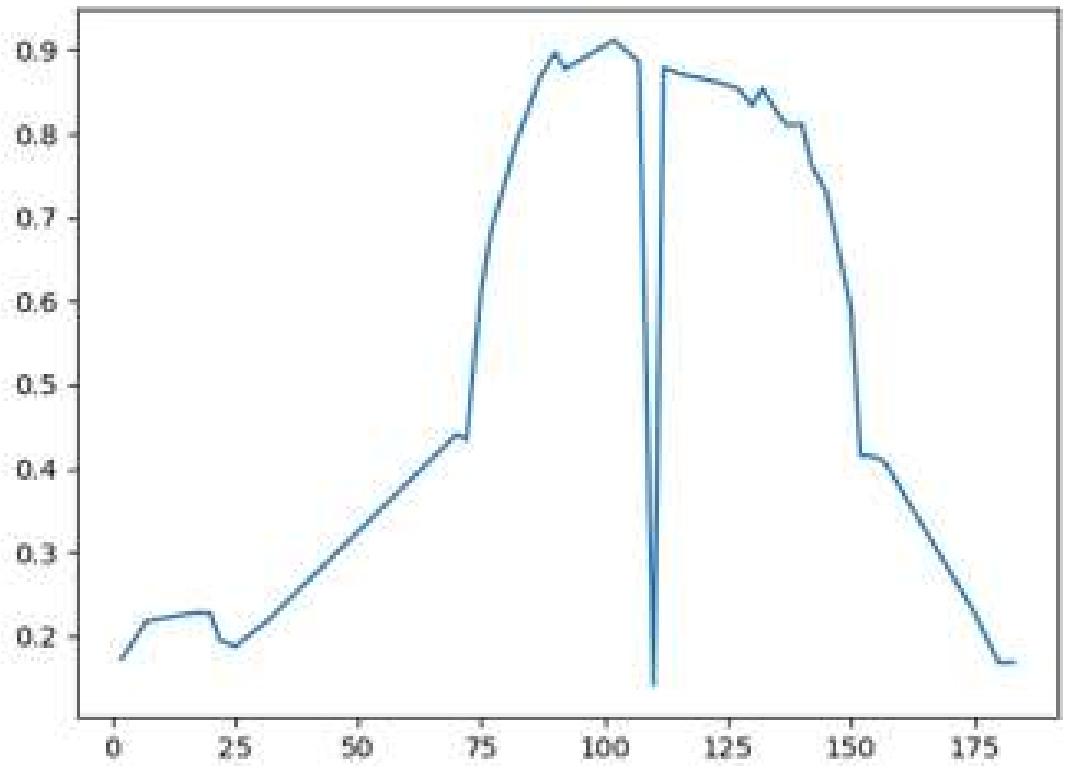
Pored primene u proceni prostorne varijabilnosti produktivnosti useva, SAPI je iskorišćen i za identifikaciju područja na kojima je potencijalno došlo do oštećenja useva usled ekstremnih vremenskih nepogoda. Za tu svrhu korišćen je spisak evidentiranih događaja



Slika 8.2: Grafički prikaz parcele sa SAPI indeksima produktivnosti po pikselu.

iz baze podataka o olujama, gradu, jakim pljuskovima, poplavama i drugim nepogodama, organizovan po godinama, datumima nastanka i pogođenim oblastima. Ovi podaci su povezani sa višegodišnjim i godišnjim SAPI vrednostima kako bi se detektovala odstupanja produktivnosti u katastarskim opštinama u kojima je registrovana pojava olujnih događaja.

Osnovna pretpostavka pristupa jeste da nepogode kao što su grad, oluja, jak vetar ili intenzivne padavine izazivaju narušavanje stanja useva, što se u satelitskom signalu manifestuje kao pad ili anomalija vrednosti vegetacionih indeksa, a samim tim i kompozitnog SAPI pokazatelja. Za razliku od višegodišnjih ili sezonski integrisanih mapa produktivnosti, koje prikazuju opšti proizvodni potencijal prostora, procena štete zasniva se na analizi odstupanja godišnjih SAPI vrednosti u odnosu na višegodišnje referentno stanje iste prostorne jedinice. Na taj način moguće je razlikovati stabilno manje produktivne zone od površina na kojima je došlo do iznenadnog poremećaja



Slika 8.3: Dijagram NDVI vegetacionog indeksa kroz vreme. U trenutku t(110) može se videti nagli pad vrednosti indeksa koji je prouzrokovan pojavom oblaka.

izazvanog spoljnim faktorom.

Metodološki, postupak je započet formiranjem i standardizacijom baze olujnih događaja, nakon čega je sprovedena njihova geolokacijska interpretacija u odnosu na prostorne jedinice analize. Pošto su se tekstualni opisi pogođenih oblasti najčešće odnosili na nazive mesta, naselja ili širih lokaliteta, analiza je u završnoj fazi usmerena na nivo katastarskih opština, kao finiji i pogodniji nivo za detekciju lokalnih promena produktivnosti nego što je to slučaj sa agregacijom na nivou opština.

8.2.1 Formiranje baze olujnih događaja, geolokacijska standardizacija i povezivanje olujnih događaja sa katastarskim opštinama

Baza olujnih događaja formirana je objedinjavanjem evidencija o registrovanim nepogodama za period 2016–2025 u jedinstvenu tabelarnu strukturu. Polazni podaci bili su organizovani po godinama, sa osnovnim atributima koji su obuhvatali datum događaja, oblast, tip nepogode i dodatni komentar. U procesu pripreme baze izvršena je standardizacija naziva kolona, ujednačavanje formata datuma, ispravljanje očiglednih tipografskih

grešaka i normalizacija tekstualnih zapisa za oblast i tip događaja. Pored originalnih atributa, uvedene su i pomoćne kolone koje omogućavaju dalju analizu, uključujući standardizovan naziv oblasti, izdvojenu listu mesta, nivo prostorne preciznosti zapisa, grupu hazarda i procenjenu jačinu događaja. Na taj način formirana je jedinstvena i analitički pripremljena baza pogodna za dalje prostorno povezivanje sa katastarskim opštinama i mapama produktivnosti.

Nakon formiranja jedinstvene baze olujnih događaja, sproveden je postupak geolokacijske standardizacije sa ciljem da se tekstualni opisi pogođenih oblasti prevedu u oblik pogodan za dalju prostornu analizu. Pošto su podaci o nepogodama izvorno sadržali nazive mesta, širih oblasti ili kombinovanih prostornih opisa, bilo je neophodno izvršiti njihovu harmonizaciju i povezivanje sa prostornim jedinicama na teritoriji Vojvodine.

U okviru ovog koraka izvršeno je izdvajanje pojedinačnih mesta iz tekstualnih zapisa, njihova standardizacija i upoređivanje sa nazivima katastarskih opština i pripadajućih opština u vektorskom sloju katastarskih opština Vojvodine. Na taj način formirana je tabela kandidata za prostorno povezivanje, pri čemu su zadržani samo oni zapisi za koje je bilo moguće uspostaviti dovoljno pouzdanu vezu sa katastarskom opštinom na teritoriji Vojvodine. Zapisi koji su se odnosili na suviše široke oblasti, regionalne oznake ili prostorno neodređene opise izdvojeni su i nisu korišćeni u finijoj analizi.

Rezultat ovog postupka je baza olujnih događaja na nivou katastarske opštine, u kojoj je svaki događaj, kada je to bilo moguće, jednoznačno povezan sa odgovarajućom katastarskom opštinom i godinom pojave. Time je omogućeno formiranje analitičkog skupa podataka pogodnog za direktno povezivanje sa statistikama produktivnosti izvedenim iz SAPI rastera.

Nakon geolokacijske standardizacije baze olujnih događaja, sproveden je postupak agregacije podataka na nivo katastarske opštine i godine, sa ciljem da se formira analitička tabela pogodna za dalje statističko povezivanje sa mapama produktivnosti. U slučajevima kada je jedan zapis obuhvatao više mesta, takav događaj je razložen na više pojedinačnih redova, tako da svaka kombinacija događaja i katastarske opštine predstavlja zasebnu jedinicu evidencije.

Na osnovu tako pripremljenih zapisa formirana je zbirna tabela na nivou godina i katastarska opština. Za svaku katastarsku opštinu i svaku godinu zadržana je informacija o prisustvu olujnog događaja, broju evidentiranih događaja, broju jedinstvenih dana sa nepogodama, maksimalnoj procenjenoj jačini događaja i dominantnim tipovima oluja. Ovako formirana tabela predstavlja osnovu za identifikaciju pogođenih katastarskih opština i njihovo dalje povezivanje sa godišnjim i višegodišnjim SAPI statistikama.

8.2.2 Izdvajanje statistika produktivnosti po katastarskim opštinama iz SAPI mapa

Nakon formiranja skupa pogođenih katastarskih opština, sprovedeno je izdvajanje statistika produktivnosti iz SAPI rastera na nivou katastarskih opština. Ovaj korak imao je za cilj da se za svaku katastarsku opštinu i za svaku godinu posmatranog perioda kvantifikuju osnovne karakteristike prostorne raspodele produktivnosti, kako bi se omogućilo poređenje sa višegodišnjim referentnim stanjem iste jedinice.

Postupak je sproveden korišćenjem godišnjih SAPI rastera za period 2016–2025, kao i višegodišnjeg median rastera koji predstavlja referentno stanje produktivnosti. Za svaku katastarsku opštinu izvršeno je prostorno preklapanje sa odgovarajućim rasterom, pri čemu su izdvojeni svi pikseli koji se nalaze unutar granica date katastarske opštine. Na tako izdvojenim vrednostima izračunate su osnovne deskriptivne statistike, uključujući srednju vrednost, median, standardnu devijaciju, minimalnu i maksimalnu vrednost, kao i ukupan broj validnih piksela.

Pored osnovnih statističkih pokazatelja, analizirana je i klasna struktura SAPI vrednosti. S obzirom na to da je produktivnost izražena kroz klasifikovane vrednosti u rasponu od 1 do 10, za svaku katastarsku opštinu izračunat je procenat piksela koji pripadaju klasama niske produktivnosti (1–3), srednje produktivnosti (4–7) i visoke produktivnosti (8–10). Na taj način, pored centralnih tendencija, obuhvaćena je i unutrašnja struktura raspodele produktivnosti unutar svake katastarske opštine.

Za višegodišnji median raster izdvojene su iste statistike, čime je formirana referentna baza produktivnosti na nivou katastarskih opština. Ovaj referentni skup korišćen je za izračunavanje anomalija godišnje produktivnosti u odnosu na višegodišnje stanje, odnosno za procenu odstupanja pojedinačnih godina od dugoročnog prostornog obrasca iste katastarske opštine.

Nakon izdvajanja statistika produktivnosti po katastarskim opštinama iz godišnjih i višegodišnjeg median SAPI rastera, sprovedeno je njihovo povezivanje sa tabelom olujnih događaja agregiranom na nivou godina i katastarskih opština. Na taj način formirana je jedinstvena analitička baza u kojoj su za svaku katastarsku opštinu i svaku godinu objedinjene informacije o godišnjoj produktivnosti, referentnom višegodišnjem stanju i registrovanim olujnim događajima.

Iz ovako formirane baze izdvojene su samo katastarske opštine u kojima je registrovana pojava oluje. Za te jedinice analizirane su anomalije produktivnosti u odnosu na sopstveno višegodišnje referentno stanje, čime je svaka pogođena katastarska opština posmatrana u odnosu na svoju dugoročnu normalu, a ne u odnosu na druge prostorne jedinice. Ovakav pristup omogućava pouzdaniju detekciju lokalnih posledica vremenskih

nepogoda.

8.2.3 Rezultati analize anomalija u katastarskim opštinama pogođenim olujama

Rezultati analize pokazuju da u 58.82% slučajeva srednja vrenost SAPI indeksa za katastarske opštine pogođene olujama je bila manja od višegodišnjeg proseka. Korišćenjem median vrednosti SAPI indeksa pokazalo se da je za opštine pogođenje olujama produktivnost u tim godinama opala za 50%.

Pored centralnih pokazatelja produktivnosti, analizirana je i promena klasne strukture SAPI vrednosti. Posebno je razmatran indikator koji predstavlja promenu udela piksela niske produktivnosti u odnosu na višegodišnje referentno stanje. Dobijeni rezultati ukazuju na izražen pozitivan pomeraj ovog pokazatelja, sa prosečnom vrednošću od +10.48, dok je u 82.35% slučajeva zabeležen porast udela klasa niske produktivnosti u opštinama gde su za datu godinu registrovane nepogode. Ovakav nalaz potvrđuje da su olujni događaji povezani ne samo sa padom prosečne produktivnosti, već i sa prostornim širenjem površina slabijeg produktivnog statusa unutar pogođenih katastarskih opština.

Na osnovu sprovedene analize može se zaključiti da katastarske opštine pogođene olujama pokazuju sistematsko negativno odstupanje produktivnosti u odnosu na sopstveno višegodišnje referentno stanje. Najstabilniji i interpretativno najjasniji pokazatelji ovog efekta bili su srednja anomalija produktivnosti, medijanska anomalija i porast udela klasa niske produktivnosti. Ovi rezultati potvrđuju da analiza na nivou katastarskih opština predstavlja pogodniji okvir za detekciju lokalnih posledica olujnih događaja nego agregacija na višim administrativnim nivoima.

8.3 Analiza povezanosti SAPI sa poznatim navodnjavanim površinama

Pored primene u proceni prostorne varijabilnosti produktivnosti i detekciji potencijalnih oštećenja useva, SAPI je korišćen i za procenu verovatnoće navodnjavanja na parcelama. Osnovna logika ovog pristupa zasniva se na pretpostavci da parcele sa stabilno višim nivoom vegetacione aktivnosti i izraženijim razvojem useva tokom sezone imaju veću verovatnoću da su bile pod uticajem dodatnog navodnjavanja. U agroekološkim uslovima Vojvodine, gde je vodni deficit tokom letnjih meseci jedan od glavnih ograničavajućih faktora proizvodnje, navodnjavanje se često manifestuje kroz očuvaniji i intenzivniji vegetacioni signal u odnosu na okolne nenavodnjavane površine.

Metodološki pristup zasnovan je na povezivanju godišnjih i višegodišnjih SAPI mapa sa referentnim podacima o navodnjavanim površinama razvijenim u prethodnim istraživanjima za područje Vojvodine. Kao osnova za interpretaciju korišćeni su rezultati studije koja se bavi detekcijom navodnjavanja primenom metoda mašinskog učenja nad Sentinel-2 podacima, kao i prvi geoprostorni skup podataka o navodnjavanim parcelama u Vojvodini za period 2020–2024 Radulović, Brdar, et al. 2023; Radulović, Marković, et al. 2025. Ovi radovi pružili su referentni okvir za razumevanje prostornog rasporeda navodnjavanih površina i potvrdu da satelitski signal vegetacije može biti pouzdan indikator prisustva navodnjavanja.

Za razliku od direktnih klasifikacionih pristupa, u disertaciji procena navodnjavanih parcela nije zasnovana na eksplicitnom modelu nadgledanog učenja, već na interpretaciji prostornih obrazaca produktivnosti izvedenih iz SAPI indeksa. Polazi se od činjenice da navodnjavanje, ukoliko je adekvatno primenjeno, smanjuje stres kod biljaka i doprinosi održavanju više produktivnosti, veće biomase i dužeg trajanja aktivne vegetacije. Takav efekat se u satelitskom signalu odražava kroz više vrednosti vegetacionih indeksa tokom kritičnih faza razvoja useva, a posledično i kroz više integralne vrednosti sezonske SAPI krive. Stoga su zone sa visokom i veoma visokom relativnom produktivnošću posmatrane kao područja sa povećanom verovatnoćom prisustva navodnjavanja, posebno ukoliko se takav obrazac ponavlja kroz više godina.

Analiza je sprovedena unutar maski ciljnih kultura, dobijenih na osnovu prethodno generisanih mapa klasifikacije useva, čime je obezbeđeno da se procena verovatnoće navodnjavanja vrši samo unutar proizvodnih površina sa istim usevom. Ovaj korak je od posebnog značaja, jer efekat navodnjavanja nije jednak kod svih useva, niti se može pouzdano interpretirati bez razdvajanja kultura sa različitim fenološkim i fiziološkim karakteristikama. U tom smislu, SAPI je analiziran za kukuruz, soju, i šećernu repu pri čemu je procena navodnjavanja naročito relevantna za letnje useve i kulture osetljivije na nedostatak vode tokom vegetacije.

Za identifikaciju potencijalno navodnjavanih parcela korišćene su dve komplementarne logike. Prva se odnosi na analizu godišnjih SAPI mapa, gde su pikseli i zone sa natprosečno visokim vrednostima indeksa izdvojeni kao kandidati za površine sa mogućim prisustvom navodnjavanja u datoj sezoni. Druga logika zasniva se na višegodišnjim prosečnim SAPI mapama, gde su kao posebno indikativne prepoznate parcele koje kroz više sezona zadržavaju visok nivo relativne produktivnosti. Takva stabilnost ukazuje da visoka vegetaciona aktivnost nije posledica samo povoljnih meteoroloških uslova u jednoj godini, već može biti povezana sa trajnim prisustvom infrastrukture ili prakse navodnjavanja.

Na mapama procene navodnjavanih parcela izdvojene su zone sa visokom verovat-

noćom navodnjavanja na osnovu prostornih obrazaca SAPI vrednosti. U praksi, to podrazumeva da su pikseli iz viših klasa produktivnosti, posebno oni koji formiraju prostorno povezane i konzistentne celine unutar parcela, interpretirani kao indikatori mogućeg navodnjavanja. Dodatna pouzdanost ovakve procene postiže se poređenjem sa referentnim slojevima navodnjavanih površina iz literature Radulović, Brdar, et al. 2023; Radulović, Marković, et al. 2025, čime se omogućava kvalitativna validacija dobijenih obrazaca.

Važno je naglasiti da visoka produktivnost ne mora uvek nužno biti posledica navodnjavanja. Na vegetacioni signal i vrednosti SAPI indeksa utiču i brojni drugi faktori, uključujući tip zemljišta, sadržaj organske materije, reljef, agrotehničke mere, izbor hibrida ili sorte, rokovi setve, primena đubriva i lokalne mikroklimatski uslovi. Zbog toga ovako dobijene mape ne predstavljaju direktnu detekciju navodnjavanja, već prostornu procenu verovatnoće da je parcela navodnjavana. Njihova najveća vrednost ogleda se u mogućnosti da izdvoje potencijalne navodnjavane površine koje zatim mogu biti proverene pomoću dodatnih izvora podataka, kao što su postojeći geoprostorni slojevi navodnjavanja, terenske informacije ili specijalizovani modeli za detekciju navodnjavanja.

8.3.1 Analiza usklađenosti između navodnjavanja i produktivnosti useva zasnovane na SAPI indeksu

U cilju ispitivanja povezanosti između navodnjavanja i produktivnosti useva, sprovedena je analiza zasnovana na poređenju vrednosti indeksa SAPI (*Satellite Agriculture Productivity Index*) između navodnjavanih i nenavodnjavanih piksela. SAPI predstavlja prostorno eksplicitnu meru relativne produktivnosti poljoprivrednih površina, pri čemu su vrednosti indeksa skalirane u opsegu od 1 do 10, gde veće vrednosti ukazuju na veću produktivnost.

Za potrebe analize korišćene su godišnje SAPI mape za 2021. i 2022. godinu, kao i mape navodnjavanja za iste godine, posebno za tri posmatrane kulture: kukuruz, soju i šećernu repu. U rasterima navodnjavanja vrednost 0 označavala je piksele koji nisu bili identifikovani kao navodnjavani, dok je vrednost 1 označavala navodnjavane piksele. Analiza je sprovedena odvojeno za svaku godinu i svaku kulturu, kako bi se izbeglo mešanje efekata različitih useva i različitih agroekoloških uslova.

Metodološki postupak realizovan je tako što su za svaku kombinaciju godine i kulture učitani odgovarajući raster SAPI i raster navodnjavanja. Nakon toga izvršena je provera njihove prostorne usklađenosti, uključujući koordinatni referentni sistem, prostornu rezoluciju, dimenzije rastera i geometrijsku transformaciju. U slučajevima kada rasteri nisu bili potpuno poravnati, raster navodnjavanja je reprojektovan i usklađen sa

prostornim gridom SAPI rastera primenom metode najbližih suseda, kako bi se očuvala kategorička priroda binarne maske navodnjavanja.

U narednom koraku izdvojeni su samo validni pikseli. To je podrazumevalo uklanjanje vrednosti piksela bez podataka iz oba rastera, kao i zadržavanje samo onih SAPI vrednosti koje pripadaju očekivanom intervalu od 1 do 10. Nakon toga, svi validni pikseli su podeljeni u dve grupe: nenavodnjavani pikseli (vrednost 0 u rasteru navodnjavanja) i navodnjavani pikseli (vrednost 1 u rasteru navodnjavanja).

Za obe grupe izračunate su osnovne deskriptivne statistike, uključujući broj piksela, srednju vrednost, medijanu, standardnu devijaciju i interkvartilni opseg. Pored toga, razlika između navodnjavanih i nenavodnjavanih površina kvantifikovana je kroz razliku srednjih vrednosti i razliku medijana:

$$\Delta\mu = \mu_{\text{nav}} - \mu_{\text{nenav}},$$

gde je μ_{nav} prosečna SAPI vrednost navodnjavanih piksela, a μ_{nenav} prosečna SAPI vrednost nenavodnjavanih piksela.

Radi statističke procene značajnosti razlika primenjen je neparametarski Mann–Whitney U test, koji je pogodan za poređenje dve nezavisne grupe bez pretpostavke normalnosti raspodele. Dodatno je korišćen i Welch-ov t-test, ali je interpretacija rezultata primarno zasnovana na Mann–Whitney testu i veličini efekta, s obzirom na veoma veliki broj piksela u analiziranim rasterima. Kao mera jačine efekta korišćen je Cliff-ov delta koeficijent, pri čemu pozitivne vrednosti ukazuju da navodnjavani pikseli teže imaju veće SAPI vrednosti od nenavodnjavanih.

Rezultati analize prikazani su u Tabeli 8.9. U svim analiziranim slučajevima za 2021. i 2022. godinu prosečne i medijalne vrednosti SAPI indeksa bile su više kod navodnjavanih piksela u odnosu na nenavodnjavane. Ovakav obrazac ukazuje na postojanje pozitivne povezanosti između prisustva navodnjavanja i povećane produktivnosti useva.

U 2021. godini najizraženiji efekat navodnjavanja uočen je kod soje. Prosečna SAPI vrednost nenavodnjavanih piksela iznosila je 5.50, dok je kod navodnjavanih piksela iznosila 7.95, što predstavlja razliku od 2.44 jedinice. Medijana je porasla sa 5 na 8, a Cliff-ov delta koeficijent od 0.484 ukazuje na izražen pozitivan efekat. Ovakav rezultat sugeriše da je kod soje navodnjavanje bilo snažno povezano sa višim nivoom produktivnosti.

Kod kukuruza je u 2021. godini takođe utvrđena pozitivna razlika u korist navodnjavanih površina. Prosečna SAPI vrednost porasla je sa 5.37 na 6.05, dok je medijana porasla sa 5 na 7. Iako je razlika bila manja nego kod soje, rezultat ipak potvrđuje dosledan pozitivan odnos između navodnjavanja i produktivnosti. Veličina efekta je bila manja ($\delta = 0.135$), što ukazuje da je uticaj navodnjavanja prisutan, ali umereniji.

Tabela 8.9: Poređenje SAPI vrednosti između navodnjavanih i nenavodnjavanih piksela za 2021. i 2022. godinu. Kolona "Broj uzoraka" označava broj nenavodnjavanih piksela, dok N_1 označava broj navodnjavanih piksela. Vrednost μ_{nenav} predstavlja prosečnu SAPI vrednost nenavodnjavanih piksela, μ_{nav} prosečnu SAPI vrednost navodnjavanih piksela, dok $\Delta\mu$ označava razliku između prosečnih vrednosti navodnjavanih i nenavodnjavanih piksela, odnosno $\Delta\mu = \mu_{nav} - \mu_{nenav}$. Cliff's δ predstavlja meru veličine efekta između dve grupe.

Godina	**Usev**	N_0	N_1	μ_{nenav}	μ_{nav}	$\Delta\mu$	**Cliff's** δ
2021	Kukuruz	54 127 756	2 025 330	5.37	6.05	0.68	0.135
2021	Soja	18 940 473	551 042	5.50	7.95	2.44	0.484
2021	Šećerna repa	52 140	286	5.71	6.29	0.58	0.107
2022	Kukuruz	46 741 902	2 892 062	5.34	7.60	2.27	0.456
2022	Soja	15 745 239	1 735 087	5.43	7.32	1.89	0.372
2022	Šećerna repa	251 314	5 038	5.14	7.20	2.06	0.388

Kod šećerne repe u 2021. godini registrovano je povećanje prosečne SAPI vrednosti sa 5.71 na 6.29, uz porast medijane sa 6 na 8. Iako su i ovde navodnjavani pikseli imali više vrednosti indeksa produktivnosti, potrebno je napomenuti da je broj navodnjavanih piksela bio veoma mali ($N = 286$), zbog čega rezultat treba tumačiti sa određenim oprezom. Ipak, i u ovom slučaju smer razlike ostaje pozitivan.

Rezultati za 2022. godinu dodatno potvrđuju uočeni obrazac. Kod kukuruza je prosečna SAPI vrednost porasla sa 5.34 na 7.60, što predstavlja razliku od 2.27 jedinica, dok je medijana porasla sa 5 na 8. Cliff-ov delta koeficijent od 0.456 ukazuje na umeren pozitivan efekat navodnjavanja. Ovakav rezultat pokazuje da su navodnjavani pikseli kod kukuruza u 2022. godini bili jasno povezani sa višim nivoom produktivnosti.

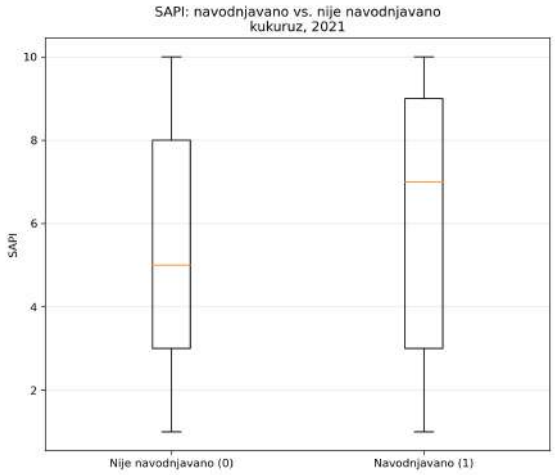
Kod soje je i u 2022. godini potvrđen stabilan pozitivan efekat navodnjavanja. Prosečna vrednost SAPI indeksa povećana je sa 5.43 na 7.32, dok je medijana porasla sa 5 na 8. Veličina efekta ($\delta = 0.372$) ukazuje na srednje izraženu do izraženu razliku između dve grupe, čime soja ostaje kultura kod koje je pozitivan uticaj navodnjavanja najkonzistentnije izražen kroz obe posmatrane godine.

Kod šećerne repe u 2022. godini zabeležen je takođe jasan pozitivan obrazac. Prosečna SAPI vrednost povećana je sa 5.14 na 7.20, uz porast medijane sa 5 na 8. Za razliku od prethodne godine, broj navodnjavanih piksela bio je znatno veći, što ovaj rezultat čini pouzdanijim. Cliff-ov delta koeficijent od 0.388 potvrđuje srednje izražen pozitivan efekat.

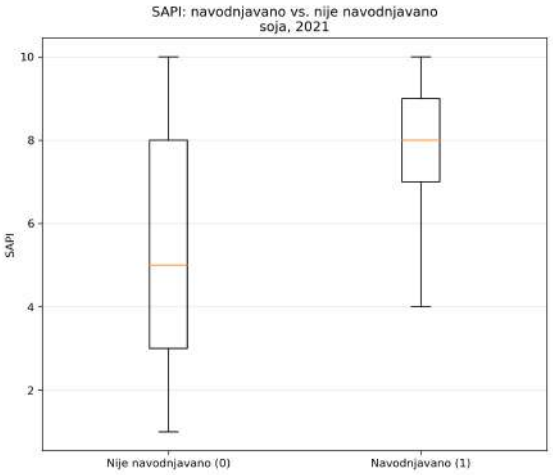
Mann–Whitney U test pokazao je statistički značajne razlike između navodnjavanih

i nenavodnjavanih piksela u svim analiziranim slučajevima u 2021. i 2022. godini ($p < 0.01$). Ipak, zbog izuzetno velikog broja posmatranih piksela, statistička značajnost sama po sebi nije dovoljna za interpretaciju, pa je poseban akcenat stavljen na razlike u srednjim i medijalnim vrednostima, kao i na veličinu efekta. Upravo ove mere potvrđuju da je navodnjavanje u posmatranim godinama uglavnom bilo povezano sa povećanom produktivnošću useva.

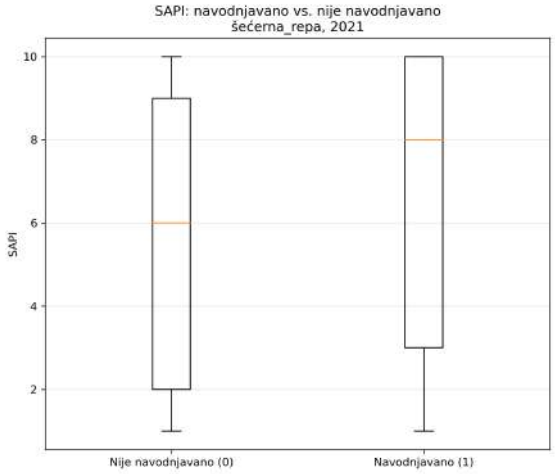
Na osnovu dobijenih rezultata može se zaključiti da postoji jasna i dosledna pozitivna povezanost između prisustva navodnjavanja i viših vrednosti SAPI indeksa, posebno u 2022. godini, kao i kod soje u obe analizirane godine. Time je potvrđeno da sloj navodnjavanja predstavlja važan faktor u tumačenju prostorne raspodele produktivnosti useva, te da kombinovanje SAPI mapa i informacija o navodnjavanju može biti korisno za detaljniju analizu proizvodnog potencijala poljoprivrednih površina.



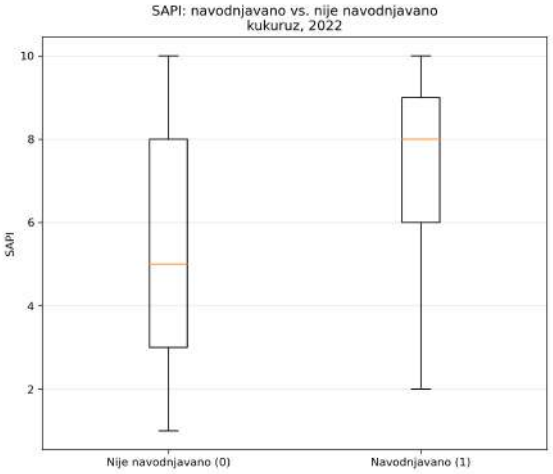
(a) Kukuruz, 2021



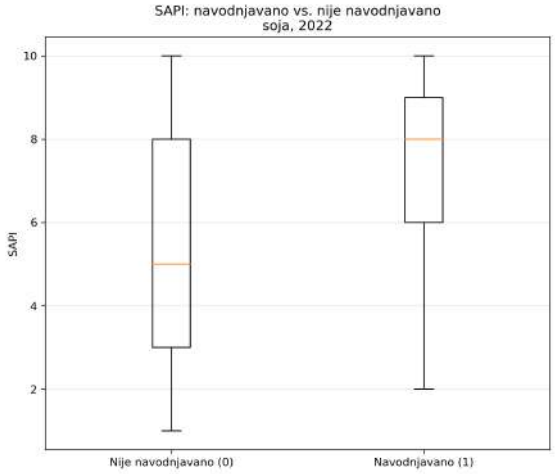
(b) Soja, 2021



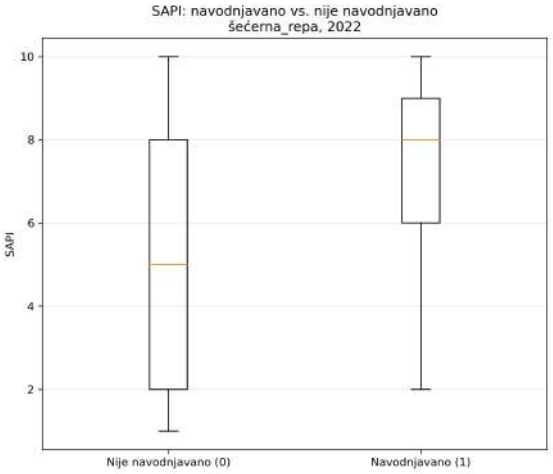
(c) Šećerna repa, 2021



(d) Kukuruz, 2022



(e) Soja, 2022



(f) Šećerna repa, 2022

Slika 8.4: Poređenje distribucija SAPI vrednosti između navodnjavanih i nenavodnjavanih piksela za tri posmatrane kulture u 2021. i 2022. godini.

Poglavlje 9

Delineacija upravljačkih zona i izbor mesta uzorkovanja

Sadržaj

9.1 Delineacija zona i određivanje mesta uzorkovanja unutar poljoprivrednih parcela 150
9.1.1 Ulazni podaci i formiranje spektralnih obeležja za segmentaciju 151
9.1.2 Klasterizacija piksela unutar parcele, prostorna homogenizacija i uklanjanje sitnih fragmenata 152
9.1.3 Određivanje reprezentativnih mesta uzorkovanja unutar zona . . 153
9.1.4 Izlazni proizvodi modela za određivanje homogenih zona 154
9.1.5 Validacija modela delineacije zona i strategije uzorkovanja . . . 155

9.1 Delineacija zona i određivanje mesta uzorkovanja unutar poljoprivrednih parcela

U cilju definisanja homogenih upravljačkih zona i određivanja reprezentativnih mesta za uzorkovanje unutar poljoprivrednih parcela, razvijen je postupak zasnovan na analizi multispektralnih satelitskih podataka Sentinel-2 i naknadnoj prostornoj segmentaciji unutar granica svake pojedinačne parcele. Metodologija je osmišljena tako da objedini dva međusobno povezana koraka: (i) delineaciju zona unutar parcele na osnovu spektralnih i izvedenih indeksa vegetacije, i (ii) izbor reprezentativnih tačaka uzorkovanja unutar svake izdvojene zone. Celokupan postupak implementiran je u programskom jeziku Python korišćenjem biblioteka GDAL/OGR, NumPy, OpenCV i scikit-learn. Glavni tok obrade podataka obuhvata učitavanje granica parcela i Sentinel-2 snimaka, formiranje maske parcele, računanje ulaznih obeležja, klasterizaciju piksela unutar parcele, prostornu homogenizaciju dobijenih klasa, konverziju rastera zona u vektorski oblik i generisanje tačaka uzorkovanja unutar svake zone.

9.1.1 Ulazni podaci i formiranje spektralnih obeležja za segmentaciju

Kao ulazni podaci korišćeni su Sentinel-2 spektralni kanali B02, B03, B04, B08 i B11, zajedno sa vektorskim granicama parcela u shapefile formatu. Izbor ovih kanala omogućava obuhvatanje informacija iz plavog, zelenog, crvenog, bliskog infracrvenog i kratkotalasnog infracrvenog dela spektra, što je posebno značajno za opis vegetacije, sadržaja vode i opšteg biofizičkog stanja useva. U glavnoj skripti definisane su putanje ka ulaznim rasterima i odgovarajućem sloju parcela, kao i broj zona koje treba izdvojiti unutar svake parcele. U prikazanoj implementaciji broj zona je zadat ručno, pri čemu je ostavljena mogućnost izbora između više vrednosti, na primer dve, tri ili četiri zone.

Za svaku parcelu prvo se formira binarna maska u raster prostoru. Ovaj korak je realizovan tako što se geometrija parcele transformiše iz vektorskog oblika u matricu dimenzija rastera. Funkcija za konverziju koordinata iz prostornog sistema u piksel-koordinate koristi geotransformaciju ulaznog rastera, dok funkcija za rasterizaciju poligona prolazi kroz sve temene tačke geometrije i popunjava pripadajuće piksele unutar granica poligona. Implementacija podržava poligon i multipoligon geometrije, što omogućava obradu i složenijih oblika parcela. Na ovaj način dalja analiza se ograničava isključivo na piksele koji pripadaju konkretnoj parceli, čime se izbegava uticaj okolnih površina.

Nakon izdvajanja maske parcele, iz Sentinel-2 kanala formiraju se ulazna obeležja za segmentaciju. U prikazanoj implementaciji korišćena su dva izvedena pokazatelja: NDWI (*Normalized Difference Water Index*) i BI (*Bare Soil Index* u konkretnoj implementaciji definisan kroz odnos plavog, crvenog i zelenog kanala). NDWI je računat kao normalizovana razlika između bliskog infracrvenog i kratkotalasnog infracrvenog kanala, dok je BI računat kao odnos kombinacije plavog, crvenog i zelenog kanala. Ovakav izbor obeležja ima za cilj da objedini informaciju o vlažnosti vegetacije i opštim spektralnim razlikama unutar parcele.

Pošto je kanal B11 Sentinel-2 dostupan u prostornoj rezoluciji od 20 m, dok su ostali korišćeni kanali u rezoluciji od 10 m, pre izračunavanja NDWI vrši se njegovo prostorno usklađivanje sa ostalim kanalima. To je realizovano bilinearnom interpolacijom, odnosno uvećanjem matrice B11 za faktor 2. Time se obezbeđuje da svi ulazni slojevi imaju istu dimenziju i mogu biti korišćeni na nivou piksela u zajedničkom postupku klasterizacije.

9.1.2 Klasterizacija piksela unutar parcele, prostorna homogenizacija i uklanjanje sitnih fragmenata

Nakon računanja ulaznih obeležja, svi pikseli unutar maske parcele prevode se u skup uzoraka pogodnih za nenadgledanu klasifikaciju. Za svaki piksel formira se vektor obeležja sastavljen od izračunatih spektralnih pokazatelja, pri čemu se analiziraju samo pikseli koji pripadaju parceli. Pre same klasterizacije primenjuje se standardizacija obeležja, kako bi se obezbedilo da različite promenljive imaju uporediv raspon i da nijedno obeležje ne dominira usled razlike u numeričkoj skali.

Za delineaciju zona primenjen je algoritam *K-means*, kojim se pikseli grupišu u unapred definisan broj klastera. Broj klastera predstavlja broj zona koje se izdvajaju unutar parcele i definiše se kao ulazni parametar modela. Rezultat klasterizacije je oznaka klastera za svaki piksel unutar parcele, kao i koordinate centroida u prostoru obeležja. Ove centroidne vrednosti predstavljaju prototipske spektralne profile izdvojenih zona. Pored numeričkih oznaka klastera, skripta generiše i vizuelni prikaz klastera u boji, kao i jednokanalni raster u kojem svaka zona dobija jedinstvenu vrednost pogodnu za dalju obradu i vektorizaciju.

Nakon klasterizacije formira se georeferencirani raster sa zonama za svaku parcelu. Time se dobija prostorna mapa homogenih zona unutar parcele.

Jedan od čestih problema kod metoda nenadgledane klasifikacije na bazi piksela jeste pojava sitnih, prostorno nepovezanih fragmenata i šuma u rezultatima. Da bi se povećala prostorna kompaktnost zona, nakon inicijalne klasterizacije primenjen je dodatni postupak homogenizacije. U okviru ove procedure za svaki piksel razmatra se njegova 3×3 okolina, a nova oznaka piksela određuje se na osnovu najčešće zastupljene nenulte vrednosti u toj lokalnoj okolini. Postupak se iterativno ponavlja sve dok ne dođe do stabilizacije rešenja ili dok se ne dostigne maksimalan broj zadatih iteracija. Na kraju se homogenizovana mapa ponovo maskira granicom parcele čime se obezbeđuje da vrednosti van parcele ostanu jednake nuli. Ovaj korak efikasno smanjuje broj izolovanih piksela i daje prostorno konzistentnije zone, pogodnije za dalju agronomsku interpretaciju i definisanje mesta uzorkovanja.

Po formiranju homogenizovanog rastera zona, rezultat se prevodi u vektorski oblik kako bi se omogućila lakša vizuelizacija, analiza površina i dalja upotreba u GIS okruženju. Za tu svrhu koristi se postupak poligonizacije rastera, pri čemu se raster sa oznakama zona pretvara u skup poligona. Nakon inicijalne poligonizacije, sve geometrije koje pripadaju istoj zoni se objedinjavanju u jedinstvene višedelne poligone. Za svaku zonu izračunava se površina koja je izražena u hektarima i dodeljuje se kao atribut poligona, odnosno zone. Dobijeni vektorski sloj sadrži identifikator zone i njenu površinu,

čime se obezbeđuje direktna primena rezultata u preciznoj poljoprivredi i planiranju terenskog uzorkovanja.

9.1.3 Određivanje reprezentativnih mesta uzorkovanja unutar zona

Nakon delineacije zona, sledeći korak predstavlja određivanje reprezentativnih mesta uzorkovanja unutar svake zone. Ovaj deo metodologije zasniva se na pretpostavci da tačka uzorkovanja treba da bude dovoljno centralno pozicionirana unutar prostorno kompaktne zone, odnosno udaljena od granica i ivica gde je verovatniji uticaj mešanja susednih zona ili rubnih efekata.

Konačne, prostorno homogenizovane zone predstavljale su neposrednu osnovu za određivanje mesta terenskog uzorkovanja. Izbor tačaka nije vršen nezavisno na nivou cele parcele, već zasebno unutar svake prethodno delineirane zone. Na taj način obezbeđeno je da svaka izdvojena prostorna celina bude zastupljena odgovarajućim reprezentativnim mestom uzorkovanja. Ove tačke su u validacionom eksperimentu označene kao OPT tačke.

U implementaciji se najpre za svaku jedinstvenu zonu formira binarna maska. Zatim se u okviru te maske izdvajaju povezane komponente korišćenjem osam susednih piksela, što omogućava razdvajanje prostorno odvojenih delova iste klase. Za svaku povezanu komponentu najpre se računa prostorni centroid na osnovu srednje vrednosti koordinata piksela komponente. Nakon toga sledi iterativna morfološka erozija komponente pomoću kernela dimenzija 3×3. Nakon svake iteracije ponovo se određuju povezane komponente i njihovi centroidi, pri čemu se zadržavaju samo oni centroidi koji i dalje pripadaju originalnoj početnoj komponenti. Poslednji centroid dobijen tokom ovog procesa tumači se kao najunutrašnjija, odnosno najreprezentativnija lokacija unutar zone. Na taj način izbor tačke nije zasnovan samo na geometrijskom centru, već na centru jezgra zone, što doprinosi stabilnijem izboru lokacije za terensko uzorkovanje. Prema tome, prostorni položaj OPT tačke zavisi istovremeno od rezultata delineacije i geometrije konkretne zone. Algoritam najpre identifikuje prostorni obuhvat zone, a zatim unutar njega određuje lokaciju koja je udaljena od granica i prostorno nestabilnih delova. Time je obezbeđena direktna veza između satelitski izvedene prostorne varijabilnosti i lokacije na kojoj se sprovodi terensko uzorkovanje.

Posebna prednost ovog pristupa jeste što prirodno eliminiše uske strukture, uske prelaze i rubne delove zone, jer se oni tokom iterativne erozije prvi uklanjaju. U kodu je dodat i prag kojim se vrlo male komponente mogu isključiti iz izbora tačaka, čime se izbegava generisanje uzoraka u prostorno nestabilnim ili zanemarljivo malim fragmentima

ukoliko dođe do njihovog formiranja. Konačno, odabrane piksele i njihove koordinate prevode se u geografske koordinate korišćenjem geotransformacije izvornog rastera i upisuju se u vektorsku strukturu tačaka sa atributima identifikatora i koordinatama.

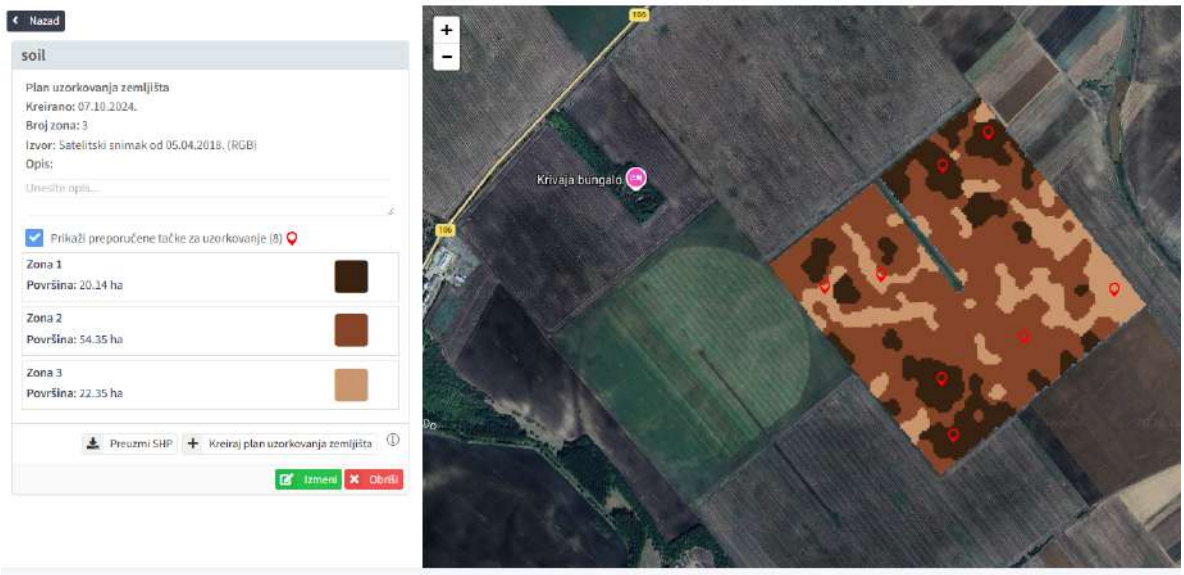
9.1.4 Izlazni proizvodi modela za određivanje homogenih zona

Kao rezultat opisanog postupka formiraju se tri međusobno povezana proizvoda: (i) georeferencirani raster zona unutar parcele, (ii) vektorski sloj poligona zona sa pripadajućim identifikatorima i površinama, i (iii) vektorski sloj reprezentativnih tačaka uzorkovanja. Ovakva organizacija izlaza omogućava integraciju sa GIS sistemima, vizuelnu proveru izdvojenih zona i direktnu primenu na terenu. Zone se mogu koristiti kao osnova za upravljačke intervencije ili detaljnije analize prostorne varijabilnosti, dok tačke uzorkovanja omogućavaju efikasno planiranje terenskih kampanja, na primer za uzorkovanje zemljišta, biljaka ili proveru stanja useva.

Prednost razvijenog modela ogleda se u tome što omogućava potpuno automatizovan prelaz od multispektralnog satelitskog snimka do operativnog GIS proizvoda pogodnog za preciznu poljoprivredu. Korišćenjem maske parcele analiza se ograničava na relevantnu površinu, nenadgledana klasterizacija omogućava izdvajanje unutrašnje heterogenosti bez potrebe za unapred definisanim klasama, a postupak morfološke erozije daje intuitivan i robustan način određivanja reprezentativnih tačaka unutar zona. Posebno je značajno što metod kombinuje spektralnu segmentaciju i prostornu regularizaciju, čime se postiže bolja primenljivost rezultata u praktičnoj primeni na parcelama.

Istovremeno, određena ograničenja treba jasno naglasiti. Prvo, broj zona se zadaje unapred i nije određen adaptivno na osnovu same strukture podataka, što može uticati na kvalitet segmentacije između različitih parcela. Drugo, skup ulaznih obeležja je relativno ograničen i zasnovan pre svega na dva izvedena indeksa, pa bi uključivanje dodatnih spektralnih, teksturnih ili vremenskih pokazatelja moglo unaprediti razdvajanje zona. Treće, homogenizacija većinskim pravilom i izbor tačke putem iterativne erozije predstavljaju heurističke postupke koji su praktični i efikasni, ali ne garantuju optimalnost u svim geometrijskim slučajevima.

Iako je validacija sprovedena korišćenjem terenskih i laboratorijskih podataka sa tri lokacije, za potpuniju procenu prenosivosti razvijenog postupka potrebno je njegovo testiranje na većem broju parcela, različitim tipovima zemljišta, kulturama i proizvodnim sezonama. Primenjeni model određuje zone na osnovu pojedinačne satelitske opservacije, pa rezultati mogu zavisiti od fenološke faze useva, prisustva korova i kvaliteta izabranog snimka. Dalji razvoj metode zato bi trebalo usmeriti ka uključivanju vremenskih serija podataka, automatskom izboru broja klastera i dodatnoj eksternoj validaciji



Slika 9.1: Primer zoniranja parcele za potrebe uzorkovanja zemljišta. Parcela je podeljena na tri zone sličnih karakteristika, dok crveni markeri predstavljaju preporučene tačke za uzorkovanje unutar izdvojenih zona.

zona korišćenjem nezavisnih podataka o zemljištu, prinosu i drugim agronomskim pokazateljima.

U celini posmatrano, razvijeni model predstavlja metodološki i praktični modul za delineaciju homogenih zona i određivanje mesta uzorkovanja unutar poljoprivrednih parcela. Validacioni eksperiment pokazao je da izdvojene zone u značajnom broju slučajeva odgovaraju razlikama u nezavisno izmerenim svojstvima zemljišta, kao i da algoritamski odabrane lokacije mogu dati rezultate uporedive sa konvencionalnim GRID uzorkovanjem uz manji broj terenskih tačaka. Njegova osnovna vrednost stoga se ogleda u povezivanju satelitski izvedene prostorne varijabilnosti sa konkretnim operativnim potrebama terenskog rada, čime se omogućava racionalnije planiranje uzorkovanja i preciznije upravljanje unutar parcele.

9.1.5 Validacija modela delineacije zona i strategije uzorkovanja

Validacija razvijenog postupka sprovedena je terenskim uzorkovanjem i laboratorijskom analizom zemljišta na tri eksperimentalne lokacije: Kikinda, Krivaja i Miletićevo. Cilj validacije bio je da se ispita da li zone izdvojene na osnovu satelitskih podataka odgovaraju prostornoj varijabilnosti fizičkih i hemijskih svojstava zemljišta, kao i da se proceni opravdanost algoritamskog izbora mesta uzorkovanja u odnosu na konvencionalni pristup.

U validacionom eksperimentu razmatrana su tri koncepta uzorkovanja. Prvi koncept,

označen kao GRID, predstavlja konvencionalno uzorkovanje prema unapred definisanoj prostornoj mreži i korišćen je kao referentni pristup. Drugi koncept, OPT, obuhvata uzorkovanje na lokacijama odabranim razvijenim algoritmom unutar prethodno delineiranih zona. Treći koncept, ONE, predstavlja minimalni scenario u kojem se karakteristike zone procenjuju na osnovu samo jednog reprezentativnog uzorka po zoni. Na ovaj način bilo je moguće analizirati ne samo prostornu opravdanost izdvojenih zona, već i uticaj intenziteta i strategije uzorkovanja na dobijene rezultate.

Ukupna baza korišćena za validaciju sadržala je 186 laboratorijski analiziranih zapisa. Na lokaciji Kikinda analizirano je 48 uzoraka, na lokaciji Krivaja 104 uzorka, dok su za Miletićevo bila dostupna 34 uzorka. Na Kikindi i Krivaji izdvojene su po tri zone, dok su na Miletićevu analizirane dve zone. U okviru GRID koncepta bilo je raspoloživo ukupno 100 tačaka, dok je OPT pristup obuhvatio 39 tačaka. Time je broj tačaka u OPT strategiji bio približno 61% manji u odnosu na konvencionalno GRID uzorkovanje.

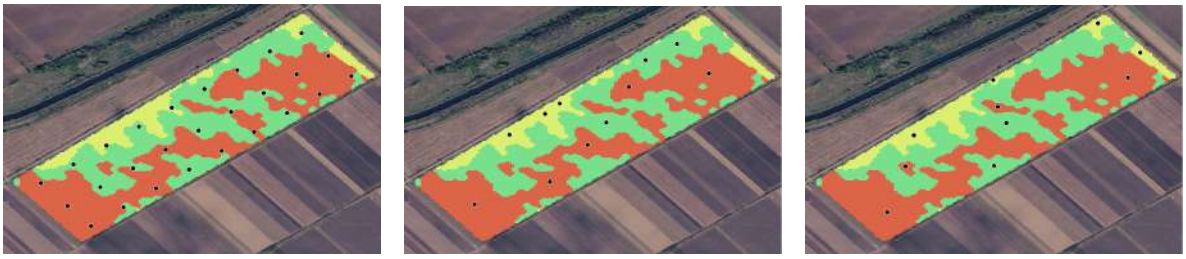
Laboratorijska analiza obuhvatila je osnovna fizička i hemijska svojstva zemljišta. Analizirani su sadržaj krupnog i sitnog peska, praha i gline, retencija vode pri pritiscima od -33, -625 i -1500 kPa, pH u vodi i KCl-u, sadržaj karbonata i humusa, kao i raspoloživi sadržaj fosfora i kalijuma. Na taj način validacija nije bila ograničena na jedan pokazatelj, već je obuhvatila više svojstava koja opisuju teksturu, vodni režim i hemijski status zemljišta.

Statistička analiza sprovedena je na nivou pojedinačnih lokacija i zona. Razlike između grupa interpretirane su korišćenjem LSD testa homogenih grupa sa nivoom značajnosti $\alpha = 0{,}05$. Vrednosti označene istim slovom pripadaju istoj homogenoj grupi, dok različite slovne oznake ukazuju na statistički značajnu razliku između posmatranih grupa.

9.1.5.1 Validacija izdvojenih zona

Prvi deo validacije odnosio se na proveru da li zone dobijene delineacijom odgovaraju razlikama utvrđenim terenskim merenjima. U ovom slučaju nije očekivano da se svaka zona statistički razlikuje za svako pojedinačno svojstvo zemljišta, budući da su zone formirane na osnovu spektralnih karakteristika površine, a ne direktnim merenjem osobina zemljišta. Ipak, prisustvo sistematskih razlika u većem broju fizičkih i hemijskih svojstava predstavlja pokazatelj da izdvojene zone odražavaju stvarnu prostornu heterogenost parcele.

Na lokaciji Kikinda statističke razlike između najmanje dve zone utvrđene su za 11 od 13 analiziranih svojstava. Razlike su zabeležene za sadržaj krupnog i sitnog peska i gline, parametre retencije vode, pH u KCl-u, sadržaj karbonata i humusa, kao i sadržaj



(a) GRID (b) OPT (c) ONE

Slika 9.2: Poređenje tri strategije uzorkovanja unutar prethodno definisanih homogenih zona. Različite boje predstavljaju izdvojene prostorne zone, dok crne tačke označavaju mesta terenskog uzorkovanja. (a) GRID – uzorkovanje prema pravilnoj prostornoj mreži, koje predstavlja konvencionalni pristup; (b) OPT – uzorkovanje na lokacijama određenim razvijenim algoritmom, pri čemu je broj i položaj uzoraka optimizovan u odnosu na geometriju i prostornu strukturu zona; (v) ONE – redukovani pristup sa jednom reprezentativnom tačkom uzorkovanja po homogenoj zoni.

raspoloživog fosfora i kalijuma. Najizraženije razdvajanje zona uočeno je kod sitnog peska, gline, retencije pri -625 i -1500 kPa i sadržaja fosfora i kalijuma. Prah i pH meren u vodi nisu pokazali jasno statističko razdvajanje svih izdvojenih zona.

Na lokaciji Krivaja razlike između zona utvrđene su za 8 od 13 analiziranih svojstava. Razdvajanje je bilo posebno izraženo kod sadržaja praha, parametara retencije vode, sadržaja karbonata, humusa, fosfora i kalijuma. Nasuprot tome, sadržaj krupnog i sitnog peska i pH vrednosti bili su relativno ujednačeni između izdvojenih zona.

Na lokaciji Miletićevo statistički uočljive razlike dobijene su za 6 od 12 svojstava uključenih u poređenje. Zone su se razlikovale prema sadržaju sitnog peska, parametrima retencije vode, sadržaju humusa i kalijuma. Za sadržaj karbonata nije bilo moguće izdvojiti razlike zbog odsustva odgovarajuće varijabilnosti u analiziranom skupu podataka.

Posmatrano zajedno za tri lokacije, razlike između delineiranih zona utvrđene su za 25 od ukupno 38 analiziranih kombinacija lokacije i svojstva, odnosno u približno 65,8% slučajeva. Ovaj rezultat pokazuje da zone formirane na osnovu satelitski izvedenih obeležja u značajnoj meri odgovaraju nezavisno izmerenoj prostornoj varijabilnosti karakteristika zemljišta.

9.1.5.2 Poređenje GRID i OPT strategija uzorkovanja

Drugi deo validacije odnosio se na poređenje rezultata dobijenih konvencionalnim GRID uzorkovanjem i uzorkovanjem na lokacijama odabranim algoritmom. GRID pristup je korišćen kao operativna referenca, budući da predstavlja uobičajenu praksu sistematskog

Tabela 9.1: Sažetak rezultata validacije delineiranih zona i poređenja GRID i OPT strategija uzorkovanja

Lokacija	**Broj zona**	**Svojstva sa razlikama između zona**	**GRID–OPT bez statistički značajne razlike**
Kikinda	3	11/13 (84,6%)	33/39 (84,6%)
Krivaja	3	8/13 (61,5%)	23/39 (59,0%)
Miletićevo	2	6/12 (50,0%)	23/24 (95,8%)
Ukupno	–	25/38 (65,8%)	79/102 (77,5%)

prostornog uzorkovanja. Pri tome GRID rezultat ne treba tumačiti kao apsolutnu stvarnu vrednost zone, već kao referentnu procenu dobijenu gušćim rasporedom terenskih tačaka.

Za Kikindu, od ukupno 39 poređenja svojstva, zone i strategije uzorkovanja, u 33 slučaja, odnosno 84,6%, nije utvrđena statistički značajna razlika između GRID i OPT pristupa. Razlike su se pojavile samo kod pojedinih teksturnih i retencionih pokazatelja i sadržaja fosfora. Ovakav rezultat pokazuje da su lokacije odabrane algoritmom u većini slučajeva davale procenu zone uporedivu sa procenom dobijenom gušćim GRID uzorkovanjem.

Na Krivaji je saglasnost bila niža. Od 39 poređenja, u 23 slučaja, odnosno 59,0%, GRID i OPT pripadali su istoj statističkoj grupi. Razlike su prvenstveno bile izražene kod mehaničkog sastava i parametara retencije vode, dok je saglasnost za većinu hemijskih pokazatelja bila veća. Ovaj rezultat ukazuje da efikasnost optimizovanog uzorkovanja zavisi i od unutrašnje heterogenosti konkretne parcele i geometrije izdvojenih zona.

Najveća saglasnost ostvarena je na lokaciji Miletićevo. Od 24 raspoloživa poređenja, 23, odnosno 95,8%, nisu pokazala statistički značajnu razliku između GRID i OPT strategije. Jedina jasna razlika zabeležena je za sadržaj humusa u jednoj od zona.

Na nivou celog eksperimenta, GRID i OPT nisu se statistički razlikovali u 79 od 102 raspoloživa poređenja, odnosno u približno 77,5% slučajeva. Istovremeno, OPT strategija je u eksperimentalnom skupu koristila 39 tačaka u odnosu na 100 tačaka GRID uzorkovanja, što predstavlja smanjenje broja terenskih lokacija za približno 61%. Dobijeni rezultat ukazuje da algoritamski izbor tačaka može zadržati značajan deo informacije dobijene konvencionalnim prostornim uzorkovanjem uz znatno manji broj terenskih lokacija.

9.1.5.3 Procena strategije sa jednim uzorkom po zoni

Treći deo eksperimenta odnosio se na scenario ONE, kojim je procenjeno koliko se reprezentativnost zone menja ukoliko se uzorkovanje svede na samo jedan uzorak. Za razliku od GRID i OPT pristupa, kod strategije sa jednim uzorkom nije moguće proceniti unutarzonsku varijabilnost na osnovu jedne terenske opservacije. Zbog toga je ovaj scenario interpretiran deskriptivno, poređenjem pojedinačnih ONE merenja sa prosečnom vrednošću dobijenom GRID pristupom u istoj zoni.

Za svako svojstvo izračunato je relativno apsolutno odstupanje:

$$\Delta = \frac{|x - \overline{x}_{\mathrm{GRID}}|}{|\overline{x}_{\mathrm{GRID}}|} \cdot 100, \tag{9.1}$$

gde je x vrednost posmatrane strategije, dok $\overline{x}_{\mathrm{GRID}}$ predstavlja prosečnu vrednost GRID uzoraka unutar iste zone. Za OPT pristup korišćena je prosečna vrednost algoritamski izabranih tačaka, dok su pojedinačna ONE merenja posmatrana kao zasebne realizacije scenarija sa jednim uzorkom. Kombinacije kod kojih je referentna GRID vrednost bila jednaka nuli nisu uključene u relativno poređenje.

Medijana apsolutnog relativnog odstupanja OPT procene u odnosu na GRID iznosila je približno 5,7%, dok je za pojedinačne ONE realizacije iznosila približno 7,2%. Isti obrazac dobijen je i kada su lokacije analizirane odvojeno. Na Kikindi su medijalna odstupanja iznosila približno 4,5% za OPT i 5,7% za ONE, na Krivaji 8,1% i 9,2%, a na Miletićevu približno 5,0% i 5,5%.

Razlike između ove dve strategije nisu velike u svim slučajevima, ali dosledno veće medijalno odstupanje ONE pristupa pokazuje povećanu osetljivost procene na tačnu lokaciju pojedinačnog uzorka. Jedan uzorak može dobro reprezentovati homogenu zonu, ali ne omogućava procenu unutrašnje varijabilnosti i povećava rizik da lokalna mikrovarijabilnost utiče na zaključak o celoj zoni. Zbog toga se ONE pristup može koristiti kao minimalna ili orijentaciona strategija u slučajevima kada su resursi za terensko uzorkovanje izrazito ograničeni, dok OPT predstavlja pouzdaniji kompromis između gustog GRID uzorkovanja i minimalnog broja terenskih tačaka.

9.1.5.4 Zaključak validacionog eksperimenta

Rezultati validacije pružaju dva komplementarna pokazatelja uspešnosti razvijenog postupka. Prvo, izdvojene zone pokazale su statistički uočljive razlike u većem broju nezavisno izmerenih fizičkih i hemijskih svojstava zemljišta, što ukazuje da satelitski izvedena segmentacija nije proizvela samo geometrijsku podelu parcele, već zone koje u značajnoj meri prate stvarnu prostornu heterogenost terena.

Drugo, algoritamski odabrane lokacije uzorkovanja dale su rezultate koji su u većini analiziranih slučajeva bili statistički uporedivi sa konvencionalnim GRID pristupom, iako je za njih bilo potrebno znatno manje tačaka. Istovremeno, rezultati ONE scenarija pokazuju da dalje smanjenje broja uzoraka na samo jednu tačku po zoni može biti prihvatljivo u pojedinim homogenim zonama, ali povećava zavisnost rezultata od lokalne varijabilnosti.

Na osnovu ovih rezultata OPT strategija može se posmatrati kao najpovoljniji kompromis između količine terenskog rada i reprezentativnosti uzorkovanja u analiziranom skupu podataka. Ipak, rezultati sa Krivaje pokazuju da kvalitet takvog pristupa zavisi od prostorne heterogenosti konkretne parcele, zbog čega algoritamski odabrane lokacije ne treba posmatrati kao univerzalnu zamenu za terensko uzorkovanje, već kao metod za njegovo prostorno usmeravanje i racionalizaciju.

Poglavlje 10

Integracija modula i grafički korisnički interfejs

Sadržaj

10.1 Povezivanje izlaza pojedinačnih modula 161
10.2 Grafički korisnički interfejs 162
10.3 Primer operativnog toka 165

10.1 Povezivanje izlaza pojedinačnih modula

Predloženi sistem nije realizovan kao jedan monolitni model koji neposredno izvršava sve faze analize, već kao skup međusobno povezanih analitičkih i vizuelizacionih modula. Svaki modul obrađuje određeni zadatak, kao što su rekonstrukcija nedostajućih satelitskih opservacija, klasifikacija useva, predikcija prinosa, procena proteinskog sadržaja, izračunavanje SAPI pokazatelja, analiza potencijalnih šteta, procena povezanosti sa navodnjavanjem i delineacija upravljačkih zona. Rezultati ovih postupaka pripremaju se u zasebnim programskim tokovima, nakon čega se čuvaju u standardizovanom geoprostornom formatu i prosleđuju modulu za prikaz.

Osnovu povezivanja modula predstavljaju zajednička prostorna referenca, jedinstvena organizacija piksela i parcela, usklađeni koordinatni sistemi i standardizovani izlazni rasteri. Na taj način rezultati različitih analiza mogu se prikazivati pojedinačno ili međusobno preklapati. Rekonstrukcija nedostajućih optičkih opservacija dopunjuje vremenske serije koje se koriste u kasnijim fazama obrade, dok klasifikaciona mapa obezbeđuje tematsku masku za izdvajanje kultura od interesa. Modeli za predikciju prinosa i proteinskog sadržaja proizvode kontinuirane prostorne pokazatelje, dok SAPI pokazatelj i modul za zoniranje opisuju relativno stanje vegetacije i unutarparcelnu heterogenost.

Ovakva organizacija omogućava da rezultati različitih analitičkih postupaka budu povezani preko njihovog zajedničkog prostornog položaja. Na taj način izlaz jednog

modula može predstavljati ulaz ili pomoćnu informaciju za narednu fazu analize, ali i nezavisan tematski sloj koji se može zajednički interpretirati sa ostalim rezultatima. Time se različiti postupci razvijeni u okviru disertacije povezuju u jedinstven modularni sistem za obradu i analizu poljoprivrednih podataka.

U grafičkom interfejsu izlazi pojedinačnih modula organizovani su u unapred definisane kategorije slojeva. Implementirane kategorije obuhvataju mape procene štete, mape navodnjavanja, SAPI mape, mape klasifikacije useva, mape predikcije prinosa, mape predikcije proteinskog sadržaja i mape upravljačkih zona. Za svaku kategoriju moguće je unapred definisati putanju do odgovarajućeg rastera, dok korisnik može i naknadno izabrati proizvoljan GeoTIFF fajl sa lokalnog računara.

Svaki raster u aplikaciji predstavljen je zasebnim objektom koji sadrži naziv sloja, putanju do datoteke, matricu vrednosti, prostornu transformaciju, koordinatni referentni sistem, oznaku nedostajućih podataka, prostorni obuhvat, izabranu paletu boja, nivo transparentnosti i status vidljivosti. Ovakva organizacija omogućava da se funkcije učitavanja, prikaza i analize primenjuju na različite vrste izlaznih mapa bez potrebe za izmenom osnovne programske logike.

Prilikom učitavanja GeoTIFF datoteke iz rastera se preuzima prvi kanal, zajedno sa pripadajućim georeferenciranim informacijama. Vrednosti označene kao *NoData* transformišu se u nedostajuće numeričke vrednosti, čime se sprečava njihovo uključivanje u izračunavanje statistika i vizuelizaciju. Sačuvani prostorni obuhvat rastera koristi se za pravilno pozicioniranje sloja unutar koordinatnog sistema prikaza.

Predložena modularna organizacija omogućava da se postojeći analitički postupci naknadno zamene ili unaprede bez izmene grafičkog interfejsa. Na primer, promena modela za predikciju prinosa ili rekonstrukciju satelitskih snimaka ne zahteva promenu vizuelizacionog modula, pod uslovom da izlaz novog modela ostane georeferencirani raster odgovarajućeg formata. Time se postiže funkcionalna odvojenost između faze modelovanja i faze interpretacije rezultata i omogućava dalje proširivanje sistema novim analitičkim modulima.

10.2 Grafički korisnički interfejs

Grafički korisnički interfejs (engl. *Graphical User Interface* – GUI) razvijen je u programskom jeziku Python primenom biblioteke *Tkinter*, dok je za rad sa georeferenciranim rasterima korišćena biblioteka *Rasterio*. Vizuelizacija podataka realizovana je pomoću biblioteke *Matplotlib*, čiji je prikaz integrisan u glavni prozor aplikacije. Interfejs je namenjen objedinjavanju, pregledu, vizuelizaciji i osnovnoj analizi prethodno pripremljenih

izlaznih slojeva.

Grafički interfejs predstavlja završni korisnički nivo razvijenog modularnog sistema. Njegova uloga je da rezultate različitih analitičkih postupaka objedini u zajedničkom vizuelnom okruženju i time olakša njihovu interpretaciju. Umesto da se svaki rezultat posmatra zasebno, korisnik može da prikaže i međusobno poredi mape procene štete, informacije o navodnjavanju, SAPI mape, rezultate klasifikacije useva, predikcije prinosa i proteinskog sadržaja i upravljačke zone.

Glavni prozor aplikacije organizovan je u nekoliko funkcionalnih celina. U gornjem delu nalazi se kontrolna traka sa poljima za izbor godine, parcele i poljoprivredne kulture, kao i komandama za učitavanje slojeva, osvežavanje prikaza, otvaranje proizvoljnog rastera i čuvanje trenutnog kartografskog prikaza. Ova polja predstavljaju osnovu za kasnije povezivanje interfejsa sa bazom podataka ili automatizovanim sistemom za izbor odgovarajućih izlaza prema sezoni, parceli i kulturi.

Levi panel sadrži pregled dostupnih tematskih slojeva. Korisnik može uključiti ili isključiti pojedinačne slojeve i za svaku kategoriju ručno definisati putanju do odgovarajuće GeoTIFF datoteke. Na ovaj način moguće je istovremeno učitati više mapa i prikazati ih jednu preko druge. Redosled prikaza i podešena transparentnost omogućavaju poređenje različitih rezultata, na primer SAPI mape, predikcije prinosa i upravljačkih zona za isto područje.

Mogućnost zajedničkog prikaza više slojeva posebno je značajna za interpretaciju njihovih prostornih odnosa. Na primer, SAPI mapa može se porediti sa mapom predikcije prinosa i raspoloživim informacijama o navodnjavanju, čime se može vizuelno analizirati da li se prostorni obrasci relativnog stanja vegetacije podudaraju sa razlikama u predikovanom prinosu i informacijama o upravljanju parcelom. Na sličan način moguće je porediti upravljačke zone sa prostornom raspodelom drugih analiziranih pokazatelja.

Centralni deo interfejsa namenjen je kartografskom prikazu. Raster se prikazuje u izvornom prostornom obuhvatu koristeći koordinate preuzete iz georeferenciranih podataka. Ugrađena *Matplotlib* navigaciona traka omogućava uvećavanje, umanjivanje, pomeranje prikaza i vraćanje na početni obuhvat. Time korisnik može pregledati rezultate na regionalnom nivou, ali i detaljnije analizirati pojedinačne parcele ili njihove delove.

Desni panel sadrži funkcije za podešavanje prikaza i pregled informacija o aktivnom sloju. Transparentnost se može kontinuirano menjati u opsegu od 0,1 do 1,0, što je posebno značajno pri preklapanju više tematskih rastera. Za svaki sloj moguće je izabrati jednu od više kontinuiranih ili kategorijskih paleta boja. Kontinuirane palete pogodne su za prikaz prinosa, SAPI vrednosti ili proteinskog sadržaja, dok se kategorijske palete mogu koristiti za klasifikacione mape i upravljačke zone.

Za svaki učitani raster automatski se izračunavaju osnovne deskriptivne statistike:

- minimalna vrednost;
- maksimalna vrednost;
- aritmetička sredina;
- medijana;
- standardna devijacija;
- ukupan broj validnih piksela.

Iz izračunavanja se isključuju pikseli koji sadrže nedostajuće ili nevažeće vrednosti. Prikazane statistike omogućavaju brzu početnu procenu raspodele vrednosti u aktivnom sloju bez potrebe za otvaranjem dodatnih GIS ili statističkih programa. Na ovaj način kartografski prikaz dopunjen je osnovnim kvantitativnim informacijama o prikazanim podacima.

Interfejs omogućava i interaktivno očitavanje vrednosti piksela. Klikom na određenu lokaciju na mapi prikazuju se prostorne koordinate izabrane tačke, red i kolona odgovarajućeg piksela i njegova numerička vrednost. Ova funkcija olakšava pregled rezultata na konkretnim lokacijama i omogućava direktno povezivanje kartografskog prikaza sa pojedinačnim numeričkim vrednostima.

Korisnik može sačuvati trenutni prikaz u formatima PNG, JPEG ili PDF. Izvoz se vrši u rezoluciji od 300 tačaka po inču (engl. *dots per inch* – dpi), što omogućava korišćenje sačuvanih prikaza u izveštajima, prezentacijama, naučnim publikacijama i drugoj dokumentaciji. Donji statusni panel pruža informacije o izvršenim radnjama, kao što su uspešno učitavanje rastera, promena putanje ili čuvanje slike.

Za unapred definisane kategorije slojeva postavljene su podrazumevane palete boja. Mape štete prikazuju se crvenim tonovima, mape navodnjavanja plavim, SAPI mape zelenim, dok su za klasifikaciju useva i upravljačke zone predviđene kategorijske palete. Korisnik ove postavke može naknadno promeniti u skladu sa karakteristikama i opsegom vrednosti prikazanog rastera.

Grafički interfejs treba posmatrati kao demonstracioni i interpretacioni sloj razvijenog sistema. Njegova osnovna namena nije obuka modela, već objedinjavanje, pregled i interpretacija njihovih izlaza. Obuka, validacija i primena pojedinačnih modela izvršavaju se u zasebnim programskim tokovima opisanim u prethodnim poglavljima, dok interfejs omogućava da se dobijeni rezultati prikažu u jedinstvenom radnom okruženju.

Sa metodološkog stanovišta, grafički interfejs predstavlja završni nivo integracije razvijenih analitičkih modula. Dok su prethodne faze usmerene na prikupljanje i pripremu podataka, rekonstrukciju satelitskih opservacija, modelovanje i prostornu analizu, GUI omogućava da se njihovi rezultati objedine i interpretiraju na nivou krajnjeg korisnika. Time se uspostavlja veza između algoritamske obrade podataka i njihove praktične primene u analizi poljoprivrednih parcela.

Razvijeni interfejs predstavlja i osnovu za dalje unapređenje sistema. Njegova modularna struktura omogućava naknadno povezivanje sa bazama podataka, automatizovano učitavanje odgovarajućih rezultata i dodavanje novih analitičkih ili vizuelizacionih funkcionalnosti. Na taj način postojeća implementacija može predstavljati polaznu osnovu za razvoj operativnih digitalnih alata i sistema podrške odlučivanju u preciznoj poljoprivredi.



Slika 10.1: Prikaz grafičkog korisničkog interfejsa za učitavanje i vizuelizaciju tematskih rastera. Interfejs omogućava podešavanje transparentnosti i palete boja (*colormap*), interaktivno očitavanje vrednosti piksela i prikaz osnovnih statističkih pokazatelja aktivnog sloja.

10.3 Primer operativnog toka

Tipičan operativni tok započinje preuzimanjem Sentinel-1 i Sentinel-2 satelitskih podataka i njihovom prostornom i vremenskom predobradom. U ovoj fazi sprovode

se geometrijsko usklađivanje, resamplovanje na zajedničku prostornu mrežu, kontrola kvaliteta i izdvajanje validnih opservacija. Ukoliko optička vremenska serija sadrži nedostajuće ili oblačnošću kontaminirane opservacije, odgovarajući delovi mogu se dopuniti primenom razvijenog generativnog modela.

Nakon pripreme satelitskih podataka primenjuje se model za klasifikaciju useva. Rezultat klasifikacije predstavlja tematsku mapu koja omogućava izdvajanje piksela određene kulture. Time se sprečava da se modeli namenjeni određenoj kulturi primenjuju na prostorno neodgovarajuće površine. Za izdvojene piksele zatim se mogu primeniti modeli za predikciju prinosa i proteinskog sadržaja, izračunavanje SAPI pokazatelja ili postupak delineacije upravljačkih zona.

Dobijeni rezultati izvoze se kao georeferencirani rasteri, pri čemu svaki piksel zadržava prostorni položaj u zajedničkom koordinatnom sistemu. Vektorski izlazi, kao što su granice parcela, izdvojene zone ili preporučena mesta uzorkovanja, mogu se čuvati u odgovarajućim vektorskim formatima i koristiti zajedno sa rasterima u eksternim GIS okruženjima.

Za prikaz rezultata korisnik najpre bira odgovarajuću kategoriju sloja i definiše putanju do pripremljene datoteke. Nakon uključivanja sloja aplikacija učitava njegovu prvu rastersku komponentu, prostorni obuhvat, koordinatni sistem i oznaku nedostajućih vrednosti. Aktivni raster se zatim prikazuje u centralnom delu interfejsa, dok se u panelu sa desne strane automatski prikazuju njegove osnovne statistike.

Korisnik može istovremeno uključiti više rasterskih slojeva. Na primer, mapa klasifikacije useva može se koristiti kao tematska osnova, preko nje se može prikazati raster predikcije prinosa, a zatim i raster upravljačkih zona. Podešavanjem transparentnosti omogućava se vizuelno poređenje prostornih obrazaca različitih pokazatelja.

Klikom na određeni deo mape korisnik dobija koordinate i vrednost izabranog piksela aktivnog sloja. Na taj način moguće je, na primer, uporediti područja niže SAPI vrednosti sa predikovanom vrednošću prinosa ili proveriti kojoj klasi ili upravljačkoj zoni pripada određena lokacija.

Završni prikaz može se izvesti kao slika visoke rezolucije. Pored vizuelizacije u razvijenom interfejsu, svi rezultati ostaju dostupni kao standardni geoprostorni proizvodi, pa se mogu dalje analizirati u programima kao što su QGIS ili ArcGIS, uključiti u baze podataka ili proslediti drugim sistemima za podršku odlučivanju.

Na ovaj način operativni tok povezuje faze prikupljanja i predobrade podataka, primenu razvijenih analitičkih metoda, generisanje prostornih izlaza i njihovu završnu vizuelizaciju. Iako pojedinačni moduli ostaju funkcionalno odvojeni, korišćenje zajedničke prostorne reference i standardizovanih izlaznih formata omogućava njihovo objedinjavanje u jedinstvenom sistemu.

Poglavlje 11

Diskusija

Sadržaj

11.0.1 Rekonstrukcija satelitskih podataka i predikcija prinosa 167
11.0.2 Klasifikacija useva, SAPI i prostorna analiza poljoprivrednih površina 170
11.0.3 Ograničenja i praktični značaj razvijenog sistema 172
11.0.4 Pravci budućih istraživanja 173

11.0.1 Rekonstrukcija satelitskih podataka i predikcija prinosa

Rezultati pokazuju da satelitske vremenske serije mogu da posluže kao zajednička osnova za više zadataka precizne poljoprivrede, ali i da pouzdanost zaključaka snažno zavisi od jedinice validacije i kvaliteta referentnih podataka. Najčvršći zaključci proizlaze iz eksperimenata u kojima su test uzorci prostorno odvojeni na nivou cele parcele. U takvim uslovima dobijene greške predstavljaju realističniju procenu primene modela na novim parcelama od slučajne podele piksela.

Generativni model je pokazao sposobnost da rekonstruiše prostorne strukture i spektralne obrasce u maskiranim oblastima Sentinel-2 snimaka. Rekonstrukcija je sprovedena sa ciljem dopune optičkih vremenskih serija u periodima kada su pojedine opservacije bile degradirane prisustvom oblaka. GAN rekonstruisani snimci zato nisu posmatrani kao direktna zamena za originalna satelitska posmatranja, već kao dopunski izvor informacija. U slučajevima kada za određeni datum nije bila dostupna odgovarajuća čista optička opservacija, rekonstrukcija omogućava dopunu vremenske serije.

Rezultati evaluacije rekonstrukcije pokazuju da je kvalitet GAN generisanih snimaka bio sezonski zavisan. Za sezonu 2019. dobijene su niže vrednosti greške i veća strukturna sličnost u odnosu na sezonu 2018. Prosečna vrednost $RMSE$ greške za ceo snimak u 2019. godini iznosila je 0.0165, dok je prosečni $SSIM$ iznosio 0.9151. Nasuprot tome, za 2018. godinu prosečni $RMSE$ za ceo snimak iznosio je 0.0246, a prosečni $SSIM$ 0.8788. Ovi rezultati ukazuju da je model u 2019. godini uspešnije očuvao prostorne i

teksturalne karakteristike scene, dok su rekonstrukcije iz 2018. godine bile opterećene većim odstupanjima u odnosu na referentne čiste snimke.

Ovakva razlika između godina može se objasniti većom složenošću scena, različitim fenološkim uslovima, stepenom oblačnosti, vremenskim razmakom između oblačnih i referentnih snimaka, kao i razlikama u odnosu između radarskog Sentinel-1 i optičkog Sentinel-2 signala. Posebno treba imati u vidu da rekonstrukcija satelitskih snimaka ne predstavlja jednostavno popunjavanje praznih vrednosti, već pokušaj generisanja spektralno i prostorno verodostojnih informacija u uslovima kada direktno optičko posmatranje nije dostupno. Zbog toga i relativno dobre vrednosti metrika, kao što su $RMSE$ i $SSIM$, ne znače nužno da rekonstruisani snimci u potpunosti zadržavaju sve fizičke i biofizičke karakteristike originalnih posmatranja.

Pored direktne evaluacije rekonstruisanih snimaka, važan kriterijum njihove upotrebljivosti predstavljao je njihov uticaj na modele za predikciju prinosa. U eksperimentima su razmatrana tri režima ulaznih podataka: korišćenje samo realnih snimaka bez značajne oblačnosti, korišćenje GAN rekonstrukcija oblačnih datuma i kombinovani režim.

U kombinovanom režimu za svaki datum korišćena je samo jedna opservacija: originalni Sentinel-2 snimak kada je scena bila dovoljno čista, odnosno GAN rekonstrukcija kada je originalna optička opservacija bila degradirana oblačnošću. U 2018. godini kombinovana vremenska serija sadržala je 12 realnih i 5 GAN rekonstruisanih datuma, dok je u 2019. godini sadržala 7 realnih i 10 GAN rekonstruisanih datuma. U obe godine kombinovani režim sadržao je ukupno 17 datuma.

Pri tome veći broj početnih obeležja ne znači da su svi atributi neposredno korišćeni u konačnom regresionom modelu. Nakon formiranja početnog prostora obeležja sprovedena je selekcija obeležja unutar odgovarajuće validacione procedure. Kod RF, XGBoost, MLR, SVR i MLR_SGD modela korišćeno je 200 odabranih obeležja, dok je kod neuronske mreže korišćeno 120 obeležja. Zbog toga stvarni doprinos pojedinačnih realnih i GAN rekonstruisanih snimaka zavisi i od toga koja su obeležja zadržana tokom selekcije.

Rezultati predikcije prinosa pokazali su da kombinovanje realnih i GAN rekonstruisanih snimaka ne dovodi do sistematskog poboljšanja performansi za sve kulture i sve algoritme. Najjasnija korist kombinovanog pristupa uočena je kod soje. U 2018. godini najniži $RMSE$ za soju ostvaren je SVR modelom u kombinovanom režimu i iznosio je 0.004605 t/piksel, uz MAE od 0.003657 t/piksel. U 2019. godini najniži $RMSE$ za soju ostvaren je neuronskom mrežom u kombinovanom režimu i iznosio je 0.004772 t/piksel, dok je MAE iznosio 0.003755 t/piksel. U oba slučaja kombinovanje dva izvora omogućilo je nižu grešku od korišćenja samo realnih ili samo GAN rekonstruisanih snimaka.

Kod kukuruza kombinovani režim nije doveo do poboljšanja najboljeg ukupnog

rezultata. U 2018. godini RF model u kombinovanom režimu ostvario je $RMSE$ od 0.015038 t/piksel, dok je nešto niža greška ostvarena korišćenjem samo realnih snimaka. Sličan obrazac zabeležen je i u 2019. godini, gde je RF u kombinovanom režimu ostvario $RMSE$ od 0.012582 t/piksel, dok je najbolji rezultat za ovu kulturu ostvaren MLR modelom nad realnim snimcima. Ovo pokazuje da povećanje broja ulaznih informacija ne mora nužno da donese dodatnu prediktivnu vrednost ukoliko su realne vremenske opservacije već dovoljno informativne za dati zadatak.

Kod pšenice su razlike između najboljih režima bile relativno male. U 2018. godini RF u kombinovanom režimu ostvario je $RMSE$ od 0.013741 t/piksel, što je veoma blisko najboljem rezultatu dobijenom nad realnim snimcima. U 2019. godini SVR u kombinovanom režimu ostvario je $RMSE$ od 0.009280 t/piksel, dok je isti tip modela nad realnim snimcima ostvario nešto nižu grešku. Ovakav rezultat ukazuje da GAN rekonstrukcije mogu zadržati značajan deo prediktivnog signala, ali da zamena oblačnih opservacija rekonstruisanim snimcima ne mora u svim slučajevima poboljšati predikciju u odnosu na korišćenje raspoloživih čistih realnih snimaka.

Sličan obrazac uočen je kod suncokreta. U 2018. godini RF u kombinovanom režimu ostvario je $RMSE$ od 0.005372 t/piksel, što je veoma blisko rezultatima dobijenim korišćenjem realnih i GAN rekonstruisanih snimaka odvojeno. U 2019. godini najbolji kombinovani rezultat ostvario je SVR sa $RMSE$ od 0.005683 t/piksel, dok je nešto niža greška ostvarena korišćenjem samo realnih snimaka. Razlike su male, pa se kod ove kulture ne može izdvojiti jasna i dosledna prednost kombinovanog režima.

Uticaj kombinovanja izvora zavisio je i od korišćenog algoritma. RF i SVR uglavnom su pokazivali stabilnije ponašanje u različitim režimima ulaznih podataka, dok su linearni modeli u pojedinim eksperimentima bili znatno osetljiviji na proširenje visokodimenzionalnog ulaznog prostora. Posebno izražene greške zabeležene su kod MLR i MLR_SGD modela za suncokret u 2018. godini u kombinovanom režimu. MLR je tada ostvario $RMSE$ od 0.045947 t/piksel, a MLR_SGD 0.047942 t/piksel. Ovakvi rezultati pokazuju da veći broj ulaznih blokova može povećati redundantnost i međusobnu korelaciju obeležja i otežati generalizaciju pojedinih modela, čak i kada se primenjuje selekcija obeležja.

Zbog toga se ne može zaključiti da dodatni GAN rekonstruisani snimci univerzalno poboljšavaju ili pogoršavaju predikciju prinosa. Njihova korisnost zavisi od kulture, godine, kvaliteta rekonstrukcije, strukture vremenske serije, korišćenog algoritma i izbora obeležja. Kombinovani pristup bio je naročito koristan kod soje, dok kod ostalih kultura nije sistematski nadmašio najbolje rezultate dobijene korišćenjem jednog izvora podataka.

Sa metodološkog stanovišta, doprinos GAN rekonstrukcije zato ne treba posmatrati

isključivo kroz zahtev da u svakoj kombinaciji modela i useva smanji grešku predikcije. Rekonstrukcija omogućava formiranje potpunijih optičkih vremenskih serija u uslovima kada oblaci onemogućavaju direktno posmatranje površine. To je posebno značajno tokom fenološki važnih perioda, kada nedostatak jedne ili više opservacija može prekinuti praćenje sezonskog razvoja useva. Istovremeno, rezultati kombinovanih eksperimenata pokazuju da dodatne rekonstruisane informacije treba koristiti selektivno i uz odgovarajuću procenu njihove informativne vrednosti.

Vizuelna sličnost i zbirne metrike rekonstrukcije nisu dovoljne za tvrdnju o potpunoj fizičkoj verodostojnosti generisanih podataka. Slične performanse modela zasnovanih na realnim i rekonstruisanim atributima mogu se tumačiti kao pokazatelj da generisani podaci zadržavaju deo relevantnog signala, ali ne dokazuju da svaki rekonstruisani snimak donosi novu i nezavisnu informaciju. Za snažniji zaključak bilo bi potrebno sprovesti dodatna poređenja sa jednostavnijim metodama dopune vremenskih serija, kao što su linearna ili spline interpolacija, kao i detaljniju evaluaciju po pojedinačnim spektralnim kanalima, vegetacionim indeksima i fenološkim fazama.

11.0.2 Klasifikacija useva, SAPI i prostorna analiza poljoprivrednih površina

Klasifikacija useva predstavlja važan tematski sloj za ostale module. Pri interpretaciji ukupne tačnosti potrebno je dati prednost metrikama koje su manje osetljive na neravnotežu klasa, kao što su F1 mera po klasi i makroprosek. Kod predikcije prinosa razlike između modela treba tumačiti zajedno sa varijabilnošću između izostavljenih parcela. Mali broj nezavisnih parcela za pojedine kulture ograničava širinu generalizacije, bez obzira na veliki broj piksela.

Pri interpretaciji rezultata klasifikacije potrebno je imati u vidu obim i strukturu ulaznih podataka. Iako satelitski snimci obezbeđuju veliki broj piksela, ti pikseli nisu potpuno nezavisni uzorci. Pikseli koji pripadaju istoj parceli često imaju slične spektralne karakteristike i dele iste uslove zemljišta, obrade i razvoja vegetacije. Zbog toga veliki broj piksela ne mora nužno da znači i veliki broj nezavisnih primera za obuku modela.

Dodatni izazov predstavlja neravnomerna zastupljenost klasa. Kulture koje zauzimaju veće površine mogu imati znatno veći broj uzoraka, dok su manje zastupljene kulture predstavljene manjim brojem parcela i piksela. U takvim uslovima ukupna tačnost može dati previše optimističnu sliku performansi, jer može biti dominantno određena uspehom klasifikacije najzastupljenijih klasa. Zato je, pored ukupne tačnosti, potrebno analizirati i metrike po klasama, kao što su preciznost, odziv i F1 mera, kao i makroprosečne vrednosti koje ravnopravnije prikazuju uspešnost klasifikacije svih useva.

SAPI indeks u ovom radu treba posmatrati kao relativni pokazatelj stanja i sezonske dinamike vegetacije, a ne kao direktnu meru prinosa ili apsolutne produktivnosti. Njegove klase izražavaju položaj pojedinačnog piksela u odnosu na referentnu raspodelu vrednosti vegetacionog signala u analiziranom prostoru i periodu. Zbog toga viša SAPI klasa ukazuje na povoljniji relativni vegetacioni odgovor u odnosu na druge piksele, ali se ne može automatski tumačiti kao veći prinos bez nezavisne terenske validacije.

Ovakav pristup je koristan jer omogućava prostorno poređenje parcela i identifikaciju zona koje tokom sezone odstupaju od očekivanog vegetacionog razvoja. SAPI može da ukaže na delove prostora sa stabilno visokim vegetacionim signalom, kao i na zone u kojima dolazi do pada ili stagnacije razvoja useva. Na taj način indeks ima praktičnu vrednost za monitoring, rangiranje proizvodnih površina i izdvajanje potencijalno problematičnih zona, posebno kada nisu dostupni detaljni podaci o prinosu, zemljištu ili agrotehničkim merama.

Međutim, interpretacija SAPI klasa mora biti oprezna. Budući da je indeks izveden iz sezonske dinamike vegetacionog signala, na njegove vrednosti mogu uticati brojni faktori: fenološka faza useva, datum setve, tip kulture, gustina sklopa, stanje zemljišta, dostupnost vode, prisustvo bolesti, korova i štetočina, kao i atmosferski i senzorski uslovi u trenutku snimanja. Zbog toga isti nivo SAPI indeksa ne mora imati isto značenje za različite kulture, godine ili agroekološke uslove.

Analiza štetnih događaja pokazuje da SAPI i izvedene anomalije mogu registrovati prostorno-vremenske promene koje su kompatibilne sa evidentiranim olujnim ili stresnim događajima. Nagli pad vegetacionog signala nakon poznatog događaja može ukazivati na potencijalno oštećenje useva, naročito ako se javlja prostorno konzistentno i u očekivanom vremenskom intervalu. Ipak, ovakav rezultat ne predstavlja pouzdanu atribuciju uzroka. Pad vegetacionog signala može biti posledica oluje, ali i suše, žetve, promene fenološke faze, bolesti, lokalnih agrotehničkih mera ili grešaka u satelitskim podacima. Zbog toga se rezultati mogu tumačiti kao detekcija anomalija kompatibilnih sa štetnim događajem, a ne kao jednoznačna potvrda uzroka oštećenja.

Slično važi i za analizu navodnjavanja. Veći broj piksela omogućava detaljnije prostorno sagledavanje razlika između navodnjavanih i nenavodnjavanih površina, ali ne povećava proporcionalno broj nezavisnih jedinica posmatranja. Susedni pikseli unutar iste parcele često su prostorno autokorelisani i dele iste uslove upravljanja, zemljišta i mikroklime. Zbog toga statistički značajne razlike na nivou piksela mogu preceniti stvarnu snagu dokaza ukoliko se ne uzme u obzir prostorna zavisnost podataka.

Iz tog razloga rezultate vezane za navodnjavanje treba opisivati kao povezanost između SAPI vrednosti i poznatih navodnjavanih površina, a ne kao potvrđenu automatsku detekciju navodnjavanja. Ukoliko navodnjavane parcele pokazuju više ili

stabilnije SAPI vrednosti tokom kritičnih fenoloških faza, to ukazuje da indeks može sadržati signal povezan sa boljom dostupnošću vode. Međutim, za pouzdanu detekciju navodnjavanja potrebna je nezavisna validacija, uključujući podatke o režimu navodnjavanja, količini vode, tipu sistema, vremenu primene i meteorološkim uslovima.

Velike razlike u tačnosti klasifikacije proteinskog sadržaja između izvora ukazuju na promenu domena i ograničenu pouzdanost modela. Ovaj rezultat je važan jer pokazuje da visoka tačnost na jednom izvoru ne garantuje generalizaciju na drugi. Zoniranje parcela daje praktično čitljive prostorne celine, a sprovedena validacija na nezavisnim laboratorijskim merenjima zemljišta pokazala je da izdvojene zone u značajnom broju slučajeva odgovaraju stvarnoj prostornoj varijabilnosti fizičkih i hemijskih svojstava zemljišta. Poređenje GRID i OPT strategija dodatno je pokazalo da algoritamski odabrane lokacije mogu obezbediti rezultate uporedive sa gušćim konvencionalnim uzorkovanjem uz značajno manji broj terenskih tačaka. Ipak, razlike između analiziranih lokacija pokazuju da efikasnost ovakvog pristupa zavisi od unutrašnje heterogenosti konkretne parcele.

11.0.3 Ograničenja i praktični značaj razvijenog sistema

Glavna ograničenja obuhvataju neujednačen broj parcela i godina između eksperimenata, prostornu autokorelaciju piksela, različitu pouzdanost referentnih izvora, ograničenu eksternu validaciju i nepostojanje jednakog nivoa kvantitativne potvrde za sve module. Sistem zato treba posmatrati kao metodološki okvir u kome su pojedine komponente validirane snažnije od drugih. Tvrdnje o uzročnosti, potpunoj operativnosti i univerzalnoj generalizaciji nisu opravdane bez dodatnih eksperimenata.

Praktični značaj rada ogleda se u jedinstvenoj prostornoj organizaciji heterogenih podataka i mogućnosti generisanja interoperabilnih GIS proizvoda koji se mogu koristiti u različitim fazama poljoprivrednog monitoringa. U okviru rada objedinjeni su satelitski podaci, prostorni slojevi, vegetacioni indeksi, rezultati klasifikacije, procene prinosa, GAN rekonstrukcije i izvedeni pokazatelji kao što je SAPI. Time je formiran okvir koji omogućava da se podaci različite prirode analiziraju u zajedničkom prostornom sistemu i da se rezultati prikažu u obliku tematskih karata pogodnih za dalju upotrebu u GIS okruženju.

Posebna vrednost ovakvog pristupa je u tome što pojedinačni moduli nisu izolovani, već mogu da se koriste kao povezani elementi šireg sistema za praćenje poljoprivredne proizvodnje. Klasifikacija useva obezbeđuje osnovni tematski sloj, GAN rekonstrukcija omogućava dopunu i proširenje optičkih vremenskih serija u periodima oblačnosti, modeli za predikciju prinosa daju procenu proizvodnog potencijala, dok SAPI omogućava

relativno poređenje sezonske dinamike vegetacije. Na taj način sistem ne pruža samo pojedinačne rezultate, već omogućava višeslojnu interpretaciju stanja useva kroz prostor i vreme.

11.0.4 Pravci budućih istraživanja

Dalji razvoj ovog pristupa treba usmeriti ka sistematskom poređenju različitih metoda i izvora podataka. Kod rekonstrukcije satelitskih snimaka potrebno je uporediti GAN pristup sa jednostavnijim metodama dopune vremenskih serija, kao što su linearna, spline ili vremenski ponderisana interpolacija. Takvo poređenje je važno kako bi se jasno utvrdilo u kojim situacijama generativni modeli zaista donose dodatnu vrednost, a u kojim slučajevima jednostavniji modeli mogu dati dovoljno dobre rezultate uz manju računarsku složenost i jednostavniju interpretaciju. Posebno bi bilo korisno analizirati da li je za predikciju prinosa korisnije uključiti sve rekonstruisane snimke ili samo rekonstruisane datume za koje nije dostupna odgovarajuća čista opservacija.

Za SAPI indeks budući rad može biti usmeren na poređenje različitih sezona i agroklimatskih uslova. Posebno bi bilo korisno analizirati ponašanje SAPI vrednosti u sušnim i kišnim godinama, kao i u sezonama sa izraženim stresnim događajima. Na taj način bi se moglo ispitati da li SAPI stabilno prepoznaje relativno bolje i lošije vegetacione odgovore u različitim uslovima ili je njegova interpretacija snažno zavisna od konkretne godine, kulture i raspodele vrednosti u analiziranom prostoru.

U oblasti predikcije prinosa dalji razvoj treba da obuhvati direktniju validaciju modela na nezavisnim sezonama, regionima i parcelama. Pored poređenja različitih algoritama, važno je uključiti i jednostavne referentne modele, kao što su prosečan prinos po kulturi, linearni modeli ili modeli zasnovani na osnovnim vegetacionim indeksima. Ovakvi referentni modeli omogućavaju realniju procenu doprinosa složenijih metoda mašinskog učenja. Takođe, korisno je uvoditi intervale pouzdanosti na nivou parcela, kako bi se rezultati predikcije prikazali ne samo kao jedna vrednost, već i kroz stepen nesigurnosti procene.

Jedan od važnih pravaca budućeg istraživanja jeste transfer obučenih modela na nove godine, nove regione i nove agroekološke uslove. Time bi se ispitalo u kojoj meri modeli naučeni na jednom prostoru i jednoj sezoni mogu da se primene u drugim uslovima bez ponovne obuke ili uz ograničeno prilagođavanje. Takva analiza je posebno važna za operativnu primenu, jer model koji dobro funkcioniše samo u okviru jedne godine ili jednog područja ima ograničenu praktičnu vrednost. U tom kontekstu potrebno je ispitati i pristupe kao što su transfer učenje, domensko prilagođavanje i kombinovanje podataka iz više sezona.

Metodološki razvoj može ići i u pravcu proširenja izvora podataka koji se koriste za validaciju i dopunu satelitskih analiza. Pored satelitskih snimaka, korisno je uključiti podatke prikupljene dronovima, terenskim senzorima, skeniranjem zemljišta sondama, meteorološkim stanicama i mašinama za preciznu poljoprivredu. Dronovi mogu obezbediti detaljnije prostorne informacije o stanju useva, dok senzori i sonde mogu dati direktne podatke o zemljišnoj vlazi, temperaturi, zbijenosti zemljišta i drugim faktorima koji utiču na razvoj biljaka. Integracija ovih izvora omogućila bi detaljniju validaciju satelitskih pokazatelja i bolje razumevanje uzroka prostornih razlika unutar parcela.

Pored toga, budući rad treba da obuhvati kvantitativnu evaluaciju izdvojenih zona i direktnu validaciju SAPI indeksa prema merenom prinosu, podacima o navodnjavanju i evidentiranim štetnim događajima. Time bi se omogućilo jasnije razdvajanje interpretacije SAPI indeksa kao relativnog pokazatelja vegetacionog stanja od njegove moguće primene kao indikatora proizvodnog potencijala, stresa ili efekata agrotehničkih mera. Posebno je važno da se evaluacija ne zasniva isključivo na pikselima, već da parcela ili prostorni blok budu osnovne jedinice analize, čime bi se smanjio uticaj prostorne autokorelacije.

Ukupno posmatrano, predloženi sistem predstavlja osnovu za razvoj integrisanog okvira za monitoring poljoprivrednih površina, procenu stanja useva i podršku odlučivanju. Njegova najveća vrednost ogleda se u povezivanju različitih izvora podataka, modela i prostornih proizvoda u jedinstvenu analitičku celinu. Rezultati predikcije prinosa dodatno pokazuju da uključivanje GAN rekonstruisanih podataka može doneti merljivo poboljšanje u pojedinim slučajevima, ali da veći broj ulaznih podataka sam po sebi ne garantuje bolju generalizaciju. Za prelazak ka operativnoj primeni potrebno je dalje raditi na eksternoj validaciji, proveri prenosivosti modela, uvođenju referentnih metoda, kvantifikaciji nesigurnosti i objavljivanju reproduktivnog programskog toka. Time bi se povećala pouzdanost rezultata, olakšalo poređenje sa drugim pristupima i omogućila šira primena razvijene metodologije u sistemima precizne poljoprivrede.

Poglavlje 12

Zaključak

Sadržaj

12.1 Odgovori na istraživačka pitanja i završni zaključci 175
12.2 Status postavljenih hipoteza 178
12.3 Završna napomena 179

12.1 Odgovori na istraživačka pitanja i završni zaključci

Ova doktorska disertacija bila je usmerena na razvoj integrisanog metodološkog okvira za analizu poljoprivrednih površina primenom podataka daljinske detekcije, terenskih merenja i metoda mašinskog učenja. Polazeći od potrebe za preciznijim, prostorno eksplicitnim i primenljivim informacijama o stanju useva i proizvodnom potencijalu parcela, razvijen je sistem koji objedinjuje više međusobno povezanih analitičkih tokova: klasifikaciju useva, procenu prostorne produktivnosti, detekciju potencijalnih šteta, procenu verovatnoće navodnjavanja, delineaciju zona unutar parcela, kao i predikciju prinosa i sadržaja proteina. Na taj način formiran je jedinstven okvir koji omogućava sagledavanje različitih aspekata poljoprivredne proizvodnje kroz zajedničku prostorno-vremensku osnovu.

Jedan od centralnih doprinosa disertacije ogleda se u razvoju i primeni pristupa SAPI (*Satellite Agriculture Productivity Index*), kojim je omogućena procena relativne produktivnosti useva na osnovu vremenskih serija Sentinel-2 podataka i izvedenih vegetacionih indeksa. Pokazano je da se satelitski signal može transformisati u prostorni indikator relativnog stanja i sezonske dinamike vegetacije na nivou piksela, pri čemu su izdvojeni prostorni obrasci višeg i nižeg relativnog vegetacionog odziva. Posebna vrednost ovog pristupa jeste u tome što ne opisuje samo trenutno stanje vegetacije, već integriše njen sezonski razvoj, čime omogućava robusniju interpretaciju prostorne varijabilnosti u odnosu na analize zasnovane na pojedinačnim satelitskim snimcima.

Na osnovu SAPI vremenskih serija razvijen je pristup za identifikaciju prostorno-vremenskih anomalija kompatibilnih sa evidentiranim štetnim događajima na usevima, pri čemu su evidentirani datumi i lokacije nepogoda povezani sa satelitski registrovanim odstupanjima vegetacionog signala. Time je omogućen prelaz sa opšte evidencije štetnih događaja ka prostorno eksplicitnom izdvajanju zona u kojima je verovatno došlo do oštećenja useva. Pored toga, SAPI je korišćen za analizu povezanosti vegetacionog odziva sa navodnjavanim površinama, polazeći od pretpostavke da površine sa stabilno višim nivoom vegetacione aktivnosti i višim klasama produktivnosti imaju veću verovatnoću da su pod uticajem dodatnog vodosnabdevanja. Iako ovakav pristup ne predstavlja direktnu detekciju navodnjavanja, rezultati pokazuju da SAPI može predstavljati koristan posredni indikator intenziteta upravljanja parcelama.

Značajan deo istraživanja bio je posvećen klasifikaciji useva na osnovu višedatumskih Sentinel-2 podataka i parcelnih referentnih informacija. Razvijeni pristup omogućio je formiranje detaljnih klasifikacionih mapa useva, pri čemu je posebna pažnja posvećena pravilnoj pripremi trening skupa, ekstrakciji spektralno-vremenskih atributa i validaciji na nivou parcela. Na taj način obezbeđen je metodološki pouzdan okvir za mapiranje useva, a dobijene klasifikacione mape predstavljaju ne samo samostalan rezultat, već i značajan pomoćni sloj za ostale analize sprovedene u okviru disertacije, naročito za maskiranje pojedinačnih kultura i interpretaciju produktivnosti.

U disertaciji je razvijen i model za delineaciju homogenih zona unutar poljoprivrednih parcela i određivanje reprezentativnih mesta uzorkovanja. Pokazano je da se kombinacijom satelitskih spektralnih deskriptora, klasterizacije i prostorne regularizacije mogu izdvojiti unutrašnje zone sa sličnim karakteristikama vegetacionog odgovora. Time je omogućeno ne samo bolje razumevanje heterogenosti unutar parcela, već i praktična podrška terenskom uzorkovanju i diferenciranom upravljanju u okviru precizne poljoprivrede. Posebna vrednost ovog dela rada ogleda se u povezivanju metoda daljinske detekcije sa konkretnim operativnim potrebama terenskog rada i donošenja odluka na nivou parcele.

Poseban segment istraživanja odnosio se na predikciju prinosa i proteinskog sastava useva. Korišćenjem višedatumskih satelitskih atributa, kao i dodatnih zemljišnih i meteoroloških podataka, ispitane su mogućnosti različitih modela mašinskog učenja za regresione i klasifikacione zadatke. Rezultati potvrđuju da integracija heterogenih izvora podataka može doprineti uspešnijem modelovanju agronomskih izlaza, ali i da performanse modela u velikoj meri zavise od kvaliteta referentnih podataka, prostorne reprezentativnosti uzorka i stabilnosti odnosa između spektralnog signala i ciljne promenljive. Posebno je značajno to što su ovi modeli evaluirani u uslovima prostorne generalizacije, čime je dodatno potvrđena potreba za pažljivim metodološkim pristupom

pri razdvajanju trening i test skupova. Rezultati su takođe pokazali da povećanje broja ulaznih obeležja samo po sebi ne garantuje poboljšanje predikcije, jer su kombinovani realni i GAN podaci kod pojedinih kultura poboljšali rezultat, dok su kod drugih performanse ostale slične ili su se pogoršale. Kod modelovanja proteinskog sadržaja uočene su i značajne razlike u performansama između različitih izvora podataka, što ukazuje na prisustvo promene domena i ograničava mogućnost neposrednog prenosa modela između različitih skupova podataka.

Važan doprinos rada predstavlja i uključivanje generativnih algoritama za rekonstrukciju optičkih satelitskih podataka u uslovima oblačnosti. GAN rekonstruisani snimci korišćeni su kako za dopunu nedostajućih opservacija, tako i za ispitivanje mogućnosti zajedničkog korišćenja realnih i rekonstruisanih podataka u modelima za predikciju prinosa.U eksperimentima predikcije prinosa GAN rekonstruisani snimci korišćeni su za datume kod kojih su originalne optičke opservacije bile degradirane oblačnošću. U kombinovanom režimu čisti datumi predstavljeni su originalnim Sentinel-2 snimcima, dok su za oblačne datume korišćene odgovarajuće GAN rekonstrukcije. Posebno je kod soje kombinovani režim ostvario najnižegreške u obe analizirane godine. Ovi rezultati ukazuju da rekonstruisani podaci mogu sadržati dodatnu prediktivnu informaciju, ali da njihov doprinos zavisi od kulture, kvaliteta rekonstrukcije, korišćenog modela i načina selekcije ulaznih obeležja.

Rezultati istraživanja potvrđuju da savremeni pristupi zasnovani na podacima daljinske detekcije i metodama mašinskog učenja imaju značajan potencijal za unapređenje monitoringa poljoprivredne proizvodnje. Korišćenjem višedatumskih Sentinel podataka moguće je pouzdano pratiti fenološku dinamiku useva, detektovati prostornu heterogenost unutar parcela i modelovati agronomski relevantne pokazatelje. Istovremeno, uključivanje terenskih i pomoćnih podataka pokazalo se kao ključni korak za prelazak sa deskriptivne analize na prediktivno modelovanje i operativnu primenu.

Dobijene rezultate, međutim, treba tumačiti i u svetlu određenih ograničenja. Kvalitet izlaza direktno zavisi od dostupnosti i pouzdanosti referentnih podataka, gustine vremenske serije satelitskih opservacija, tačnosti klasifikacionih slojeva i prostorne usklađenosti svih ulaznih podataka. Kod pojedinih zadataka, kao što su procena štete i procena navodnjavanja, satelitski indikatori prvenstveno ukazuju na verovatnoću i prostornu raspodelu pojave, ali ne mogu uvek jednoznačno razlikovati sve uzroke promena bez dodatne terenske, meteorološke ili upravljačke potvrde. Slično tome, regresioni modeli za prinos i kvalitet prinosa mogu biti dodatno unapređeni uključivanjem informacija o agrotehničkim praksama, sortimentu, režimu navodnjavanja i zemljišnim intervencijama.

U tom smislu, buduća istraživanja mogu biti usmerena u nekoliko pravaca. Prvi pravac odnosi se na dalje unapređenje vremenskog modelovanja korišćenjem arhitektura

kao što su GRU i transformeri za vremenske serije. Bidirekcioni LSTM modeli mogu biti razmatrani u retrospektivnim analizama kompletnih sezonskih vremenskih serija, dok su za operativno predviđanje tokom sezone prikladniji modeli koji koriste samo prethodno dostupne opservacije. Drugi pravac podrazumeva dublju integraciju Sentinel-1 i Sentinel-2 podataka, kao i razvoj robusnijih pristupa za rad u uslovima oblačnosti i nepotpunih vremenskih serija. Treći pravac odnosi se na proširenje analitičkog okvira uključivanjem dodatnih izvora podataka, kao što su UAV snimci, detaljniji zemljišni podaci, podaci o agrotehničkim merama i podaci sa IoT senzora. Konačno, poseban pravac budućih istraživanja može biti usmeren ka daljoj operacionalizaciji razvijenih modela kroz integraciju u softverske platforme i sisteme za podršku odlučivanju namenjene krajnjim korisnicima u poljoprivredi.

Može se zaključiti da razvijeni sistem predstavlja metodološki konzistentnu osnovu za razvoj operativnih sistema za digitalni monitoring poljoprivredne proizvodnje i podršku odlučivanju u preciznoj poljoprivredi. Naučni doprinos disertacije ogleda se u razvoju i povezivanju više komplementarnih pristupa za analizu poljoprivrednih površina na nivou piksela, dok se njen praktični značaj ogleda u mogućnosti primene razvijenih metoda u realnim uslovima upravljanja poljoprivrednim parcelama. Na taj način disertacija doprinosi razvoju savremenih digitalnih rešenja za efikasniju, precizniju i održiviju poljoprivrednu proizvodnju.

12.2 Status postavljenih hipoteza

Na osnovu sprovedenih eksperimenata, hipoteza H1 je delimično podržana, jer su rekonstruisani podaci u pojedinim nizvodnim zadacima pokazali uporedivu korisnost, ali njihova dodatna vrednost u odnosu na jednostavne interpolacije nije u potpunosti potvrđena. Hipoteza H2 je podržana u smislu identifikacije relativnih prostornih obrazaca, dok je za tumačenje SAPI kao direktnog pokazatelja prinosa potrebna dodatna validacija. Hipoteza H3 je podržana kao statistička povezanost, ali ne kao uzročna ili operativna detekcija navodnjavanja. Hipoteza H4 je podržana. Validacioni eksperiment pokazao je da kombinovanje spektralne klasterizacije i prostorne homogenizacije omogućava izdvajanje zona koje u značajnom broju slučajeva odgovaraju nezavisno izmerenoj prostornoj varijabilnosti svojstava zemljišta. Algoritamski odabrane OPT lokacije bile su statistički uporedive sa konvencionalnim GRID uzorkovanjem u 79 od 102 poređenja (77,5%), uz približno 61% manji broj terenskih tačaka. Rezultati ipak ukazuju da uspešnost pristupa zavisi od prostorne heterogenosti konkretne parcele, pa je za procenu njegove šire prenosivosti potrebna validacija na dodatnim lokacijama.

12.3 Završna napomena

Razvijeni okvir pokazuje da kombinovanje daljinske detekcije, terenskih podataka, geoprostorne obrade i mašinskog učenja može da podrži više komplementarnih odluka u preciznoj poljoprivredi. Najvažniji doprinos rada nije jedan univerzalni model, već prostorno i vremenski usklađen način povezivanja podataka i zadataka, uz jasno prepoznavanje granica generalizacije svakog pojedinačnog modula.

Bibliografija

Agency, European Space (2021). *Copernicus Sentinel Data Guide*. URL: https://sentinels.copernicus.eu/.

Anuradha, D, Ramu Kuchipudi, B Ashreetha, Janjhyam Venkata Naga Ramesh, and Ayadi Rami (2024). "Enhancing Agricultural Yield Forecasting with Deep Convolutional Generative Adversarial Networks and Satellite Data." In: *International Journal of Advanced Computer Science & Applications* 15.2.

Basso, Bruno, Costanza Fiorentino, Davide Cammarano, and Urs Schulthess (2016). "Variable rate nitrogen fertilizer response in wheat using remote sensing". In: *Precision agriculture* 17, pp. 168–182.

Bellón, Beatriz, Agnès Bégué, Danny Lo Seen, Claudio Aparecido De Almeida, and Margareth Simões (2017). "A remote sensing approach for regional-scale mapping of agricultural land-use systems based on NDVI time series". In: *Remote Sensing* 9.6, p. 600.

Bourriz, Mohamed, Hicham Hajji, Ahmed Laamrani, Nadir Elbouanani, Hamd Ait Abdelali, François Bourzeix, Ali El-Battay, Abdelhakim Amazirh, and Abdelghani Chehbouni (2025). "Integration of hyperspectral imaging and AI techniques for crop type mapping: Present status, trends, and challenges". In: *Remote Sensing* 17.9, p. 1574.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.

Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chowdhury, Md Sharafat (2024). "Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting". In: *Environmental Challenges* 14, p. 100800.

Crnojević, Vladimir, Predrag Lugonja, Branko Brkljač, and Borislav Brunet (2014). "Classification of small agricultural fields using combined Landsat-8 and RapidEye imagery: case study of northern Serbia". In: *Journal of Applied Remote Sensing* 8.1, pp. 083512–083512.

Denison, R Ford and Raymond Russotti (1997). "Field estimates of green leaf area index using laser-induced chlorophyll fluorescence". In: *Field Crops Research* 52.1-2, pp. 143–149.

European Environment Agency (2020). *CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly*. Version 2020_20u1. DOI: 10.2909/960998c1-1870-4e82-8051-6485205ebbac. URL: https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac.

Farr, Tom G., Paul A. Rosen, et al. (2007). "The Shuttle Radar Topography Mission". In: *Reviews of Geophysics* 45.2. DOI: 10.1029/2005RG000183.

Feng, Wei, Yapeng Wu, Li He, Xingxu Ren, Yangyang Wang, Gege Hou, Yonghua Wang, Wandai Liu, and Tiancai Guo (2019). "An optimized non-linear vegetation index for estimating leaf area index in winter wheat". In: *Precision Agriculture* 20.6, pp. 1157–1176.

Feyisa, Gudina L, Henrik Meilby, Rasmus Fensholt, and Simon R Proud (2014). "Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery". In: *Remote Sensing of Environment* 140, pp. 23–35.

Fuentes, Sigfredo, Roberta De Bei, Joanne Pech, and Stephen Tyerman (2012). "Computational water stress indices obtained from thermal image analysis of grapevine canopies". In: *Irrigation Science* 30.6, pp. 523–536.

Fuentes-Peñailillo, Fernando, Karen Gutter, Ricardo Vega, and Gilda Carrasco Silva (2024). "Transformative technologies in digital agriculture: Leveraging Internet of Things, remote sensing, and artificial intelligence for smart crop management". In: *Journal of Sensor and Actuator Networks* 13.4, p. 39.

Gao, Bo-Cai (1996). "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space". In: *Remote sensing of environment* 58.3, pp. 257–266.

Gitelson, Anatoly A and Mark N Merzlyak (1996). "Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll". In: *Journal of plant physiology* 148.3-4, pp. 494–500.

— (2022). "Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll". In: *Journal of Plant Physiology* 178, pp. 494–500.

Goel, Narendra S and Wenhan Qin (1994). "Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation". In: *Remote Sensing Reviews* 10.4, pp. 309–347.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2014). "Generative adversarial nets". In: *Advances in neural information processing systems* 27.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2020). "Generative adversarial networks". In: *Communications of the ACM* 63.11, pp. 139–144.

Gowda, V Dankan, R Saranya, SV Ramanan, NMG Kumar, and Priya SACHIN Jadhav (2026). "AI-based solutions for future trends in iot and remote sensing integration for precision agriculture". In: *AI-Driven Smart Industrial Technologies*. IGI Global Scientific Publishing, pp. 19–44.

Hartigan, John A and Manchek A Wong (1979). "Algorithm AS 136: A k-means clustering algorithm". In: *Journal of the royal statistical society. series c (applied statistics)* 28.1, pp. 100–108.

Hengl, Tomislav, Jorge Mendes de Jesus, Gerard BM Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, Marvin N Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, et al. (2017). "SoilGrids250m: Global gridded soil information based on machine learning". In: *PLoS one* 12.2, e0169748.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. (2020). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049.

Huang, Zhongling, Xidan Zhang, Zuqian Tang, Feng Xu, Mihai Datcu, and Junwei Han (2024). "Generative artificial intelligence meets synthetic aperture radar: A survey". In: *IEEE Geoscience and Remote Sensing Magazine*.

Huete, Alfredo R (1988). "A soil-adjusted vegetation index (SAVI)". In: *Remote sensing of environment* 25.3, pp. 295–309.

Huete, AR, HQ Liu, KV Batchily, and WJDA Van Leeuwen (1997). "A comparison of vegetation indices over a global set of TM images for EOS-MODIS". In: *Remote sensing of environment* 59.3, pp. 440–451.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros (2017). "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134.

Jabed, Md Abu and Masrah Azrifah Azmi Murad (2024). "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability". In: *Heliyon* 10.24.

Jain, Anil K. (2010). "Data clustering: 50 years beyond K-means". In: *Pattern Recognition Letters* 31.8, pp. 651–666.

Jiang, Zhangyan, Alfredo R Huete, Kamel Didan, and Tomoaki Miura (2008). "Development of a two-band enhanced vegetation index without a blue band". In: *Remote sensing of Environment* 112.10, pp. 3833–3845.

Kataoka, Takashi, Toshihiro Kaneko, Hiroshi Okamoto, and S Hata (2003). "Crop growth estimation system using machine vision". In: *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*. Vol. 2. IEEE, b1079–b1083.

Kaufman, Yoram J and Didier Tanre (1992). "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS". In: *IEEE transactions on Geoscience and Remote Sensing* 30.2, pp. 261–270.

Khaki, S. and L. Wang (2019). "Crop Yield Prediction using Deep Neural Networks". In: *Frontiers in Plant Science* 10, p. 621.

Khaki, Saeed, Lizhi Wang, and Sotirios V Archontoulis (2020). "A CNN-RNN framework for crop yield prediction". In: *Frontiers in plant science* 10, p. 1750.

Khan, Atiya, Chandrashekhar Himmatrao Patil, Amol D Vibhute, and Shankar Mali (2026). "Crop type classification using Sentinel-2 images and AI-enabled methods for precision agriculture". In: *Applied Geomatics* 18.1, p. 8.

Khanal, Sami, John Fulton, Andrew Klopfenstein, Nathan Douridas, and Scott Shearer (2018). "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield". In: *Computers and electronics in agriculture* 153, pp. 213–225.

Kooistra, Lammert, Katja Berger, Benjamin Brede, Lukas Valentin Graf, Helge Aasen, Jean-Louis Roujean, Miriam Machwitz, Martin Schlerf, Clement Atzberger, Egor Prikaziuk, et al. (2023). "Reviews and syntheses: Remotely sensed optical time series for monitoring vegetation productivity". In: *Biogeosciences Discussions* 2023, pp. 1–67.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton (2017). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems* 30.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep Learning". In: *Nature* 521.7553, pp. 436–444.

Lin, Fudong, Summer Crawford, Kaleb Guillot, Yihe Zhang, Yan Chen, Xu Yuan, Li Chen, Shelby Williams, Robert Minvielle, Xiangming Xiao, et al. (2023). "Mmst-vit: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5774–5784.

Liu, Peng, Jun Li, Lizhe Wang, and Guojin He (2022). "Remote sensing data fusion with generative adversarial networks: State-of-the-art methods and future research directions". In: *IEEE Geoscience and Remote Sensing Magazine* 10.2, pp. 295–328.

Lloyd, Stuart (1982). "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137.

Louppe, Gilles (2014). *Understanding random forests: From theory to practice*. Universite de Liege (Belgium).

Lu, Yuchen et al. (2023). "A survey of uncertainty in deep neural networks". In: *Artificial Intelligence Review*.

Ma, Xianping, Yiming Huang, Xiaokang Zhang, Man-On Pun, and Bo Huang (2023). "Cloud-egan: Rethinking cyclegan from a feature enhancement perspective for cloud removal by combining cnn and transformer". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16, pp. 4999–5012.

MacQueen, James (1967). "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1, pp. 281–297.

Maimaitijiang, Maitiniyazi, Vasit Sagan, Paheding Sidike, Ahmad M Daloye, Hasanjan Erkbol, and Felix B Fritschi (2020). "Crop monitoring using satellite/UAV data fusion and machine learning". In: *Remote Sensing* 12.9, p. 1357.

Marko, Oskar, Sanja Brdar, Marko Panic, Predrag Lugonja, and Vladimir Crnojevic (2016). "Soybean varieties portfolio optimisation based on yield prediction". In: *Computers and Electronics in Agriculture* 127, pp. 467–474.

Meraner, Andrea, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt (2020). "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166, pp. 333–346.

Mitrović, Slavka, Aleksandra Mitrović, and Maja Cogoljević (2017). "Contribution of Agriculture to the Development of Serbia". In: *Economic of Agriculture* 64.2, pp. 805–819.

Motohka, Takeshi, Kenlo Nishida Nasahara, Hiroyuki Oguma, and Satoshi Tsuchida (2010). "Applicability of green-red vegetation index for remote sensing of vegetation phenology". In: *Remote Sensing* 2.10, pp. 2369–2387.

Najjar, Hiba, Miro Miranda, Marlon Nuske, Ribana Roscher, and Andreas Dengel (2025). "Explainability of subfield level crop yield prediction using remote sensing". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18, pp. 4141–4161.

Ngugi, Habiba N, Absalom E Ezugwu, Andronicus A Akinyelu, and Laith Abualigah (2024). "Revolutionizing crop disease detection with computational deep learning: a comprehensive review". In: *Environmental Monitoring and Assessment* 196.3, p. 302.

Nikolaev, Evgeny, Nataly Zakharova, and Vladimir Zakharov (2025). "Cloud Removal in Satellite Images Using a Generative Adversarial Approach". In: *2025 International Russian Smart Industry Conference (SmartIndustryCon)*. IEEE, pp. 614–618.

Orynbaikyzy, Aiym, Ursula Gessner, and Christopher Conrad (2019). "Crop type classification using a combination of optical and radar remote sensing data: A review". In: *international journal of remote sensing* 40.17, pp. 6553–6595.

Patil, Yashraj, Harikrishnan Ramachandran, Sridhevi Sundararajan, and P Sridevipon-malar (2025). "Comparative analysis of machine learning models for crop yield prediction across multiple crop types". In: *SN Computer Science* 6.1, p. 64.

Pejak, Branislav, Predrag Lugonja, Aleksandar Antić, Marko Panić, Miloš Pandžić, Emmanouil Alexakis, Philip Mavrepis, Naweiluo Zhou, Oskar Marko, and Vladimir Crnojević (2022). "Soya yield prediction on a within-field scale using machine learning models trained on Sentinel-2 and soil data". In: *Remote sensing* 14.9, p. 2256.

Peng, Min, Yunxiang Liu, Asad Khan, Bilal Ahmed, Subrata K Sarker, Yazeed Yasin Ghadi, Uzair Aslam Bhatti, Muna Al-Razgan, and Yasser A Ali (2024). "Crop monitoring using remote sensing land use and land change data: Comparative analysis of deep learning methods using pre-trained CNN models". In: *Big Data Research* 36, p. 100448.

Pettorelli, Nathalie (2013). *The normalized difference vegetation index*. Oxford University Press.

Radulović, M., S. Brdar, B. Pejak, P. Lugonja, I. Athanasiadis, N. Pajević, and V. Crnojević (2023). "Machine learning-based detection of irrigation in Vojvodina (Serbia) using Sentinel-2 data". In: *GIScience & Remote Sensing* 60.1. DOI: 10.1080/15481603.2023.2262010.

Radulović, M., M. Marković, S. Brdar, et al. (2025). "The first geospatial dataset of irrigated fields (2020–2024) in Vojvodina (Serbia)". In: *Scientific Data* 12, p. 105. DOI: 10.1038/s41597-025-04443-9.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, pp. 234–241.

Rouse, John Wilson, Rüdiger H Haas, John A Schell, Donald W Deering, et al. (1974). "Monitoring vegetation systems in the Great Plains with ERTS". In: *NASA special publication* 351.1974, p. 309.

Sebastianelli, Alessandro, Artur Nowakowski, Erika Puglisi, Maria Pia Del Rosso, Jamila Mifdal, Fiora Pirri, Pierre Philippe Mathieu, and Silvia Liberata Ullo (2021). "Spatio-Temporal SAR-Optical Data Fusion for Cloud Removal via a Deep Hierarchical Model". In: *arXiv preprint arXiv:2106.12226*.

Sharma, Abhinav, Arpit Jain, Prateek Gupta, and Vinay Chowdary (2020). "Machine learning applications for precision agriculture: A comprehensive review". In: *IEEE access* 9, pp. 4843–4873.

Stow, D, M Niphadkar, and J Kaiser (2005). "MODIS-derived visible atmospherically resistant index for monitoring chaparral moisture content". In: *International Journal of Remote Sensing* 26.17, pp. 3867–3873.

Thakur, Poornima Singh, Shubhangi Chaturvedi, Pritee Khanna, Tanuja Sheorey, and Aparajita Ojha (2023). "Vision transformer meets convolutional neural network for plant disease classification". In: *Ecological Informatics* 77, p. 102245.

Tillack, Adina, Anne Clasen, Birgit Kleinschmit, and Michael Förster (2014). "Estimation of the seasonal leaf area index in an alluvial forest using high-resolution satellite-based vegetation indices". In: *Remote Sensing of Environment* 141, pp. 52–63.

Wang, Jinyu, Haitao Yang, Yu He, Fengjie Zheng, Zhengjun Liu, and Hang Chen (2024). "An unpaired SAR-to-optical image translation method based on Schrödinger bridge network and multi-scale feature fusion". In: *Scientific Reports* 14.1, p. 27047.

Wang, Lei, Xin Xu, Yue Yu, Rui Yang, Rong Gui, Zhaozhuo Xu, and Fangling Pu (2019). "SAR-to-optical image translation using supervised cycle-consistent adversarial networks". In: *Ieee Access* 7, pp. 129136–129149.

Wang, Lingli and John J Qu (2007). "NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing". In: *Geophysical Research Letters* 34.20.

Wang, Peng, Yongkang Chen, Bo Huang, Daiyin Zhu, Tongwei Lu, Mauro Dalla Mura, and Jocelyn Chanussot (2025). "MT_GAN: A SAR-to-optical image translation method for cloud removal". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 225, pp. 180–195.

Wei, Juan, Huanxin Zou, Li Sun, Xu Cao, Shitian He, Shuo Liu, and Yuqing Zhang (2023). "CFRWD-GAN for SAR-to-optical image translation". In: *Remote Sensing* 15.10, p. 2547.

Woebbecke, David M, George E Meyer, Kenneth Von Bargen, and David A Mortensen (1995). “Color indices for weed identification under various soil, residue, and lighting conditions”. In: *Transactions of the ASAE* 38.1, pp. 259–269.

Wu, Renzhe, Guoxiang Liu, Jichao Lv, Yin Fu, Xin Bao, Age Shama, Jialun Cai, Baikai Sui, Xiaowen Wang, and Rui Zhang (2023). “An innovative approach for effective removal of thin clouds in optical images using convolutional matting model”. In: *Remote Sensing* 15.8, p. 2119.

Xu, Hanqiu (2006). “Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery”. In: *International journal of remote sensing* 27.14, pp. 3025–3033.

You, Jie, Wei Liu, and Joonwhoan Lee (2020). “A DNN-based semantic segmentation for detecting weed and crop”. In: *Computers and Electronics in Agriculture* 178, p. 105750.

Zhong, Liheng, Lina Hu, and Hang Zhou (2019). “Deep learning based multi-temporal crop classification”. In: *Remote sensing of environment* 221, pp. 430–443.