



УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА



**ДАЉИНСКА ДЕТЕКЦИЈА
САДРЖАЈА ОРГАНСКОГ
УГЉЕНИКА У ЗЕМЉИШТУ НА
БАЗИ ДУБОКОГ УЧЕЊА**

ДОКТОРСКА ДИСЕРТАЦИЈА

Ментор:
Андерла др Андраш

Кандидат:
Марко Павловић

Нови Сад, 2024

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА¹

| | |
|--|---|
| Врста рада: | Докторска дисертација |
| Име и презиме аутора: | Марко Павловић |
| Ментор (титула, име, презиме, звање, институција): | др Андраш Андерла, редовни професор, Факултет техничких наука, Универзитет у Новом Саду |
| Наслов рада: | Даљинска детекција садржаја органског угљеника у земљишту на бази дубоког учења |
| Језик и писмо рада: | Српски (хирилица) |
| Физички опис рада: | Унети број: Страница 173 Поглавља 11 Референци 170 Табела 11 Слика 50 Графикона 7 Прилога 2 |
| Научна област: | Индустријско инжењерство и инжењерски менаџмент |
| Ужа научна област (научна дисциплина): | Вештачка интелигенција |
| Кључне речи / предметна одредница: | Даљинска детекција, органски угљеник у земљишту, вештачка интелигенција, машинско учење, дубоко учење |
| Апстракт на језику рада: | Циљ истраживања је да се развије тачан и поуздан модел, на бази дубоког учења, за процену органског угљеника у земљишту коришћењем података добијених путем даљинске детекције и да се упореде његове перформансе са другим постојећим моделима. Модел има за циљ не само да оствари напредак у перформансама које остварују постојеће методе у смислу тачности и поузданости, већ и да буде прилагодљив различитим географским пејзажима и условима животне средине, обезбеђујући широку примену. |
| Датум прихватања теме од стране надлежног већа: | 26.10.2023. |
| Датум одбране: (Попуњава накнадно институција) | |
| Чланови комисије: (титула, име, презиме, звање, институција) | Председник: др Дарко Стефановић, редовни професор, Факултет техничких наука, УНС Члан: др Ненад Филиповић, редовни професор, Факултет инжењерских наука, УКГ Члан: др Душан Јовановић, ванредни професор, Факултет техничких наука, УНС Члан: др Ђорђе Пржуљ, редовни професор, Факултет техничких наука, УНС Члан: др Милан Мирковић, редовни професор, Факултет техничких наука, УНС Члан: др Ранко Чабиловски, ванредни професор, Пољопривредни факултет, УНС Ментор: др Андраш Андерла, редовни професор, Факултет техничких наука, УНС |
| Напомена: | |

¹ Аутор докторске дисертације потписао је и приложио следеће Обрасце:

5б – Изјава о ауторству;

5в – Изјава о истовестности штапане и електронске верзије докторског рада и дозвола за објављивање личних података;

5г – Изјава о коришћењу.

Ове Изјаве се чувају у институцији у штапаном и електронском облику и не корице се са радом.

**UNIVERSITY OF NOVI SAD
FACULTY OR CENTER**

KEY WORD DOCUMENTATION²

| | |
|--|---|
| Document type: | Doctoral dissertation |
| Author: | Marko Pavlovic |
| Supervisor (title, first name, last name, position, institution) | dr Andraš Anderla, full professor, Faculty of technical sciences, University of Novi Sad |
| Thesis title in English: | Remote sensing of soil organic carbon content based on deep learning |
| Language and script: | Serbian language (cyrillic) |
| Physical description: | Number of: Pages 173 Chapters 11 References 170 Tables 11 Illustrations 50 Graphs 7 Appendices 2 |
| Scientific field: | Industrial engineering and engineering management |
| Scientific subfield (scientific discipline): | Artificial intelligence |
| Subject, Key words: | Remote sensing, soil organic carbon, artificial intelligence, machine learning, deep learning |
| Abstract in English: | The goal of the research is to develop an accurate and reliable model, based on deep learning, for estimating soil organic carbon using remote sensing data and to compare its performance with other existing models. The model aims not only to improve the performance achieved by existing methods in terms of accuracy and reliability, but also to be adaptable to different geographical landscapes and environmental conditions, ensuring widespread application. |
| Date of endorsement by the scientific board: | 26.10.2023. |
| Date of defence: (Filled in by the institution) | |
| Thesis defence board: (title, first name, last name, position, institution) | Chair: dr Darko Stefanović, full professor, Faculty of Technical Sciences, UNS Member: dr Nenad Filipović, full professor, Faculty of Engineering Sciences, UKG Member: dr Dušan Jovanović, associate professor, Faculty of Technical Sciences, UNS Member: dr Đorđe Pržulj, full professor, Faculty of Technical Sciences, UNS Member: dr Milan Mirković, full professor, Faculty of Technical Sciences, UNS Member: dr Ranko Čabilovski, associate professor, Faculty of Agriculture, UNS Mentor: dr Andraš Anderla, full professor, Faculty of Technical Sciences, UNS |
| Note: | |

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.

ЗАХВАЛНОСТ

За израду докторске дисертације захвалност дугујем ментору, др Андрашу Андерли, на корисним саветима и сугестијама. Захвалио бих се и проф. др Дарку Стефановићу, председнику комисије, на подршци на сваком кораку приликом рада на докторској дисертацији. Захвалан сам и осталим члановима комисије, проф. др Ненаду Филиповићу, др Душану Јовановићу, ванр. проф., проф. др Ђорђу Пржуљу, др Ранку Чабиловском ванр. проф. и проф. др Милану Мирковићу који су били доступни за мене и пружили ми подршку приликом израде докторске дисертације.

Посебно бих се захвалио проф. др Дубравку Ђулибрку који ме је увео у област вештачке интелигенције, пружио ми шансу и помоћ током мог путовања у тада нову, непознату и узбудљиву област вештачке интелигенције.

Захваљујем се свим својим колегама са којима сам имао привилегију да блиско сарађујем, током свог рада у привреди и после у научно-истраживачком раду. Сваки од вас је на свој начин допринео мом личном и професионалном расту, чинећи ме бољим стручњаком и истраживачем.

Такође, захваљујем се својим пријатељима, јер су ми пријатељи и чине мој живот лепшим.

Највећу захвалност дугујем својој мајци Мирјани и сестри Јелени који су ми били безрезервна подршка и ослонац током читавог живота. Хвала вам што сам на вас могао да рачунам и што сте биле уз мене увек. Без вас нити би било докторске дисертације, нити бих ја био човек какав сам данас.

Свако од нас има прошлост иза и будућност испред себе.

У том духу, докторску дисертацију посвећујем свом покојном оцу Милану, због кога сам желео да једног дана напишем и одбраним своју докторску дисертацију, циљ који сам успео да испуним.

Са погледом напред, дисертацију такође посвећујем и својим сестрићима Растку, Даниелу и Вуку да и они достигну циљеве које себи зацртају.

САДРЖАЈ

| | | |
|--------|---|----|
| 1 | УВОДНА РАЗМАТРАЊА, ПРОБЛЕМ, ПРЕДМЕТ И ЦИЉЕВИ ИСТРАЖИВАЊА | 7 |
| 1.1 | Проблем, предмет и циљ истраживања | 8 |
| 1.2 | Дефинисане хипотезе и истраживачка питања | 9 |
| 2 | ПРЕГЛЕД ЛИТЕРАТУРЕ И ПРОБЛЕМИ У ПРОЦЕНИ ОРГАНСКОГ УГЉЕНИКА У ЗЕМЉИШТУ | 11 |
| 3 | ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА И МАШИНСКО УЧЕЊЕ | 17 |
| 3.1 | Основни концепти у машинском учењу | 19 |
| 3.1.1 | Учење са надзором, учење без надзора и учење путем подстицаја | 20 |
| 3.1.2 | Класификација и регресија | 22 |
| 3.1.3 | Груписање | 24 |
| 3.1.4 | Једноставни и сложени модели | 25 |
| 3.1.5 | Параметарски и непараметарски модели | 26 |
| 3.1.6 | Недовољно и превелико прилагођавање | 26 |
| 3.1.7 | Компромис између пристрасности и варијансе | 27 |
| 3.1.8 | Метрике машинског учења | 29 |
| 3.2 | Општи принципи у машинском учењу | 32 |
| 3.2.1 | Окамова оштрица | 32 |
| 3.2.2 | Теорема непостојања бесплатног obroка | 32 |
| 3.2.3 | Клетва димензионалности | 33 |
| 3.3 | Ток машинског учења | 33 |
| 3.4 | Одабрани алгоритми машинског учења | 36 |
| 3.4.1 | Линеарна регресија | 36 |
| 3.4.2 | Логистичка регресија | 38 |
| 3.4.3 | Метода К најближих суседа | 39 |
| 3.4.4 | Наивни Бајесов класификатор | 42 |
| 3.4.5 | Стабла одлучивања | 43 |
| 3.4.6 | Насумична шума | 46 |
| 3.4.7 | Екстремно насумична стабла | 47 |
| 3.4.8 | Стабла квантилне регресије | 48 |
| 3.4.9 | Машине вектора упоришта | 50 |
| 3.4.10 | Алгоритам повећавања градијента | 52 |
| 3.4.11 | Вештачке неуронске мреже | 53 |
| 4 | ДУБОКО УЧЕЊЕ | 57 |
| 4.1 | Конволуционе неуронске мреже | 58 |
| 4.2 | Рекурентне неуронске мреже | 61 |
| 4.3 | Мреже са дугим и кратким памћењем | 63 |
| 4.4 | Трансформери | 65 |
| 4.5 | Аутоенкодери | 67 |
| 4.6 | Генеративне супростављене мреже | 69 |

| | | |
|-------|---|-----|
| 4.7 | U-Net сегментациони модел | 70 |
| 5 | ДАЉИНСКА ДЕТЕКЦИЈА | 73 |
| 5.1 | Спектар електромагнетног зрачења | 76 |
| 5.2 | Интеракција са Земљином атмосфером (расипање, преламање, апсорпција) | 77 |
| 5.3 | Атмосферски прозори | 81 |
| 5.4 | Интеракција са Земљином површином (рефлексија, трансмисија, дифракција) | 82 |
| 5.5 | Платформе за даљинску детекцију | 86 |
| 5.6 | Sentinel 2 мисија | 89 |
| 5.7 | Индекси даљинске детекције | 90 |
| 5.8 | Атрибути терена | 91 |
| 6 | ОРГАНСКИ УГЉЕНИК У ЗЕМЉИШТУ | 93 |
| 6.1 | Директно мерење промена у залихама SOC | 100 |
| 6.2 | Процењивање промена у залихама SOC путем мерења токова угљеника | 102 |
| 6.3 | Процењивање промена у залихама SOC коришћењем геостатистичких техника | 102 |
| 6.4 | Процењивање промена у залихама SOC путем метода даљинске детекције са земље | 104 |
| 6.5 | Процењивање промена у залихама SOC путем метода даљинске детекције из ваздуха | 104 |
| 6.6 | Процењивање промена у залихама SOC путем метода даљинске детекције из свемира | 107 |
| 6.7 | Статистичко истраживање оквира површине употребе/покривача земљишта - LUCAS | 108 |
| 6.8 | Чилеанска база података о органском угљенику у земљишту - CHLSOC | 109 |
| 6.9 | Координација информација о животној средини - CORINE | 110 |
| 6.10 | EO-learn земљишни покривач | 112 |
| 7 | ПРЕДЛОЖЕНИ ПРИСТУП ПРОЦЕНИ SOC ПУТЕМ ПРИМЕНЕ ДУБОКИХ НЕУРОНСКИХ МРЕЖА И ПОДАТАКА ДОБИЈЕНИХ ДАЉИНСКОМ ДЕТЕКЦИЈОМ | 114 |
| 7.1 | Пристап процени органског угљеника у земљишту | 115 |
| 7.2 | Фаза 1 – сегментациона дубока неуронска мрежа са U-Net архитектуром за процену LC | 116 |
| 7.3 | Тренирање U-Net модела на јавно доступним подацима LULC за Словенију | 117 |
| 7.4 | Прикупљање, обрада и чување сателитских снимака | 118 |
| 7.4.1 | Учитавање, претходна обрада и припрема сателитских снимака, као и кореспондирајућих стварних података о терену (тј. маски стварних вредности) | 121 |
| 7.4.2 | U-Net сегментациони модел | 123 |
| 7.5 | Фаза 2 - SOC естиматор | 129 |

| | | |
|-----|---|-----|
| 8 | РЕЗУЛТАТИ ИСТРАЖИВАЊА И ДИСКУСИЈА | 138 |
| 8.1 | Евалуација LC класификације | 138 |
| 8.2 | SOC процена Експеримент 1 | 139 |
| 8.3 | Додатна евалуација модела са најбољим учинком за процену SOC - Експеримент 2 | 143 |
| 8.4 | Процена садржаја SOC у подручјима са хомогеним типовима вегетацијског покривача - Експеримент 3 | 147 |
| 8.5 | Конструисање SOC мапе: примена модела и мозаик приступ | 151 |
| 8.6 | Генерисање и доступност развијеног производа | 152 |
| 8.7 | Дискусија | 152 |
| 8.8 | Ограничења предложеног приступа | 154 |
| 9 | ЗАКЉУЧНА РАЗМАТРАЊА | 158 |
| 10 | ЛИТЕРАТУРА | 160 |
| 11 | ПРИЛОЗИ | 172 |

Списак слика

| | | |
|----|---|-----|
| 1 | Модел машинског учења | 20 |
| 2 | Учење са надзором | 21 |
| 3 | Учење без надзора | 21 |
| 4 | Учење путем подстицаја | 23 |
| 5 | Регресија и класификација | 24 |
| 6 | Груписање | 25 |
| 7 | Компоненте података за тренирање[40] | 27 |
| 8 | Недовољно, превелико и одговарајуће прилагођавање[40] | 28 |
| 9 | Грешке пристрасности и варијансе[40] | 28 |
| 10 | Компромис између пристрасности и варијансе | 29 |
| 11 | Ток машинског учења | 34 |
| 12 | Линеарна регресија | 36 |
| 13 | Приказ логистичке и линеарне регресије | 39 |
| 14 | К најближих суседа | 41 |
| 15 | Стабло одлучивања | 44 |
| 16 | Машине вектора упоришта | 51 |
| 17 | Алгоритам повећавања градијента | 52 |
| 18 | Вештачка неуронска мрежа | 56 |
| 19 | Конволуциона неуронска мрежа за класификацију[70] | 59 |
| 20 | Рекурентна неуронска мрежа | 61 |
| 21 | LSTM неуронска мрежа[71] | 64 |
| 22 | Трансформер архитектура[72] | 66 |
| 23 | Аутоенкодер неуронска мрежа | 67 |
| 24 | Генеративно супростављена мрежа[74] | 69 |
| 25 | Архитектура U-Net сегментационог модела[75] | 71 |
| 26 | Пасивна и активна даљинска детекција[76] | 73 |
| 27 | Landsat 1 сателит | 74 |
| 28 | Електромагнетни спектар | 77 |
| 29 | Графички приказ расипања, преламања и апсорпције | 81 |
| 30 | Атмосферски прозори[48] | 82 |
| 31 | Графички приказ рефлексије, трансмисије и дифракције | 85 |
| 32 | Слика фиксног земаљског сензора за даљинску детекцију | 86 |
| 33 | Слика беспилотног ваздушног система за даљинску детекцију[48] | 88 |
| 34 | Слика Sentinel 2 сателита за даљинску детекцију[77] | 89 |
| 35 | Циклус кружења угљеника[79] | 94 |
| 36 | Узорковање земљишта за процену садржаја органског угљеника[130] | 101 |
| 37 | Систем вртложне коваријансе[138] | 103 |
| 38 | Мобилни уређај за NIR спектроскопију[147] | 105 |
| 39 | Специјално прерађени авион ДЦ-8 у власништву NASA[48] | 106 |
| 40 | Беспилотни ваздушни систем за даљинску детекцију (NASA)[156] | 107 |
| 41 | Просторна расподела података LUCAS 2015. и 2018. године | 109 |
| 42 | Просторна дистрибуција CHLSOC скупа података | 110 |
| 43 | Подела територије Словеније на плочице | 112 |

| | | |
|----|--|-----|
| 44 | Узорци улазних делова узетих за регион Тоскане | 116 |
| 45 | Предложени приступ процени SOC | 117 |
| 46 | RGB улазна слика дела територије Републике Словеније величине 1500 x 1500 пиксела | 122 |
| 47 | Архитектура модела за LC класификацију | 125 |
| 48 | Графички приказ губитака приликом тренирања и валидације модела за LC класификацију током епоха | 129 |
| 49 | Google Earth Engine Code Editor кориснички интерфејс | 131 |
| 50 | Слика Републике Србије са вредностима BSI, RGB и NDVI | 132 |
| 51 | Слика АП Војводина са вредностима BSI, RGB и NDVI | 133 |
| 52 | Приказ закривљености за територију Републике Србије | 134 |
| 53 | Приказ нагиба за територију Републике Србије | 135 |
| 54 | Приказ аспекта за територију Републике Србије | 136 |
| 55 | Приказ надморске висине за територију Републике Србије | 137 |
| 56 | Евалуација LC класификације: матрица конфузије | 138 |
| 57 | Представљање вредности SOC-а у сивим тоновима са преклапајућом маском валидности за регион Тоскане | 153 |

Списак табела

| | | |
|----|---|-----|
| 1 | Еволуција CORINE базе података | 111 |
| 2 | Класификација земљишта по употреби и земљишном покривачу | 112 |
| 3 | Резултати 10-струке унакрсне валидације за базичне моделе трениране на сва 3 скупа података за базичне моделе 01 и 02 | 141 |
| 4 | Резултати 10-струке унакрсне валидације за базичне моделе трениране на сва 3 скупа података за базичне моделе 03 и 04 | 141 |
| 5 | Резултати 10-струке унакрсне валидације за хибридне моделе трениране на сва 3 скупа података за хибридне моделе 01 и 02 | 142 |
| 6 | Резултати 10-струке унакрсне валидације за хибридне моделе трениране на сва 3 скупа података за хибридне моделе 03 и 04 | 142 |
| 7 | Резултати евалуације базичних 04 модела са алгоритмима са најбољим учинком за процену SOC | 143 |
| 8 | Резултати евалуације хибридних 04 модела са алгоритмима са најбољим учинком за процену SOC | 144 |
| 9 | Резултати евалуације хибридних 04 модела са алгоритмима са најбољим учинком за процену SOC у подручјима са хомогеним типовима вегетацијског покривача | 149 |
| 10 | Класе CORINE скупа података | 152 |
| 11 | Поређење са објављеним истраживањима | 154 |

1 УВОДНА РАЗМАТРАЊА, ПРОБЛЕМ, ПРЕДМЕТ И ЦИЉЕВИ ИСТРАЖИВАЊА

Органски угљеник у земљишту (енгл. Soil Organic Carbon - SOC) је више од индикатора плодности земљишта. SOC је критична компонента глобалног циклуса угљеника, где заузима важну улогу у ублажавању климатских промена путем секвестрације угљеника. Тачно праћење SOC-а је од виталног значаја за разумевање квалитета земљишта, пољопривредне продуктивности и здравља животне средине. Сложеност динамике SOC-а, на коју утиче мноштво фактора, укључујући коришћење земљишта, климатске услове и биолошку активност, представља значајне изазове за конвенционалне методе праћења.

Технологија даљинске детекције појављује се као моћно средство у превазилажењу ових изазова, нудећи потенцијал за развој и примену великих и ефикасних анализу SOC-а. Међутим, замршености превођења података добијених путем даљинске детекције у тачне SOC процене нису тривијалне. Различити спектрални потписи које су ухватили сателити морају бити пажљиво анализирани да би се уочиле суптилне варијације које указују на нивое SOC. Овај задатак је додатно закомпликован нелинеарним и често нејасним односима између посматраних спектралних података и стварног садржаја SOC-а.

Појава машинског учења (енгл. Machine Learning - ML) и дубоког учења (енгл. Deep Learning - DL), посебно дубоких неуронских мрежа (енгл. Deep Neural Networks - DNN), нуди револуционарни приступ овом проблему. Ове технологије се истичу у моделирању сложених, нелинеарних односа, што их чини посебно погодним за тумачење вишедимензионалних података добијених даљинском детекцијом. Њихова способност да уче из огромних скупова података и идентификују обрасце који нису одмах очигледни људским аналитичарима или традиционалним статистичким методама мења традиционалне приступе у применама даљинске детекције.

Интеграција DNN-а у даљинску детекцију за процену SOC-а није само постепено побољшање, него представља промену парадигме у мониторингу животне средине. Коришћењем моћи напредних алгоритама, могуће је извући значајније увиде из сателитских података, што доводи до прецизнијег, детаљнијег и свеобухватнијег SOC мапирања. Овај приступ има потенцијал да трансформише наше разумевање плодности земљишта на глобалном нивоу, информишући о пракси одрживог управљања земљиштем и значајно доприносећи нашим напорима у очувању животне средине и ублажавању климатских промена.

Штавише, потенцијал ових технологија превазилази процену SOC-а. Развијене методологије се могу прилагодити и применити на низ задатака праћења животне средине, постављајући терен за шире примене у еколошким истраживањима и одрживом управљању ресурсима. Фузија даљинске детекције са напредним рачунарским техникама означава нову еру у науци о животној средини, где увиди засновани на подацима могу довести до ефикаснијег и одрживијег управљања природним ресурсима наше планете.

1.1 Проблем, предмет и циљ истраживања

У основи потребе за овим истраживањем је проблем развоја тачних и робусних модела даљинске детекције за процену садржаја SOC на бази дубоког учења. Директно мерење угљеника у земљишту се суочава са значајним изазовима због инхерентне сложености и варијабилности SOC-а, на који утичу различити фактори као што су промене коришћења земљишта, климатски услови и биолошка активност. Ова сложеност се погоршава када се преведе у домен даљинске детекције, где мултиспектрална природа сателитских података захтева напредне технике интерпретације. Садашњи модели често не успевају у прецизном хватању ове динамике, што доводи до празнина у процени SOC-а и може имати значајне импликације на управљање животном средином и креирању политика и законских регулатива.

Предмет истраживања је усредсређен на искоришћавање потенцијала дубоких неуронских мрежа како би се остварио значајан напредак у пољу даљинске детекције за процењивање садржаја SOC. Ово подразумева темељно истраживање о томе како се DNN могу оптимално дизајнирати и истренирати за обраду сложених, мултиспектралних сателитских података, и како се могу искористити да се из ових података разазнају суптилни, али критични показатељи SOC-а. Истраживање се бави испитивањем учинка ML модела DNN-а како би се одговорило на специфичне изазове које поставља SOC процена, укључујући руковање великим скуповима података, прилагођавање нелинеарности у подацима и интеграцију различитих извора података ради побољшања тачности предвиђања.

Циљ истраживања је да се развије тачан и поуздан DNN модел за процену SOC коришћењем података добијених путем даљинске детекције и да се упореде његове перформансе са другим постојећим моделима. Овај модел има за циљ не само да оствари напредак у перформансама које остварују постојеће методе у смислу тачности и поузданости, већ и да буде прилагодљив различитим географским пејзажима и условима животне средине, обезбеђујући широку примену. Једна од амбиција је и да се истраживање преведе у комерцијално одржив алат, који нуди напредно решење за доносиоце регулатива, агенције за заштиту животне средине и пољопривредни сектор за ефикасно праћење и управљање SOC-ом. Поред тога, циљ се протеже на суштински допринос глобалном дијалогу о очувању животне средине, ублажавању климатских промена и одрживом управљању земљиштем.

Да би се постигао овај циљ, истраживање ће укључивати неколико корака. Поглавље 1 садржи уводна разматања на основу којих је проистекла мотивација за истраживање, изложени и дефинисани су проблем, предмет и циљ дисертације, као и полазне хипотезе.

Поглавље 2 пружа исцрпан преглед тренутне литературе, идентификујући недостатке у методологијама процене SOC. Истиче се потреба за побољшаним моделима даљинске детекције за прецизну процену SOC, наглашавајући сложеност овог задатка с обзиром на различите еколошке и климатске факторе који утичу на SOC.

У поглављу 3 је изложен кратак приказ на историјат вештачке интелигенције (енгл. Artificial Intelligence - AI), кључне моменту у развоју AI-а, као и осврт на успон метода дубоког учења.

Поглавље 4 пружа увид у основне концепте, опште принципе, сам ток машинског учења, као и детаљна објашњења функционисања алгоритама машинског учења. У де-

лу 4.1 описани су основни концепти машинског учења попут врста машинског учења, недовољног и претераног прилагођавања и метрика које се употребљавају приликом евалуације учинка ML модела. акође, представљен је и сам ток процеса машинског учења са његовим компонентама. Део 4.4. пружа детаљан опис најчешће коришћених ML алгоритама са објашењем принципа њиховог функционисања и наведеним предностима и недостацима сваког алгорита.

Поглавље 5 фокусира се на пружање увида у различите архитектуре дубоког учења, њихову примену уз осврт на потенцијалне недостатке сваке архитектуре.

У поглављу 6 дат је историјат развоја поља даљинске детекције. Даље, представљени су принципи технологије даљинске детекције, њене интеракције са Земљином атмосфером и површином, као и различите платформе које се користе за даљинску детекцију. Ово поставља основу за разумевање извора података и методологија коришћених у истраживању.

У поглављу 7 разматра се утицај угљеника на процесе који се одвијају у животној средини са приказом циклуса кружења угљеника. Посебно се објашњава значај органског угљеника у земљишту за плодност земљишта и његов потенцијал у секвестрацији угљеника. Даје се детаљан преглед стања SOC, као и приказ различитих метода за процену промена у залихама SOC-а, од директних мерења до техника даљинског детекције.

Поглавље 8 представља кључан део истраживања, у којем се детаљно описује предложени иновативни приступ који комбинује дубоке неуронске мреже са подацима добијеним путем даљинске детекције за процену SOC-а. У њему се разрађују специфичности предложене методологије, укључујући изворе података и укључене процесе корак по корак.

У поглављу 9 представљени су и анализирани налази истраживања. Поглавље укључује процене класификације земљишног покривача, експерименте процене SOC-а и генерисање SOC мапа, а завршава се дискусијом о резултатима, укључујући ограничења предложеног приступа.

У поглављу 10 дат је резиме докторске дисертације и наводе се потенцијалне практичне примене налаза истраживања, наглашавајући његов шири утицај.

На крају у поглављу 11 наведене су све библиографске референце и литература која коришћена у истраживању.

1.2 Дефинисане хипотезе и истраживачка питања

Основне постављене хипотезе које ће бити потврђене или оповргнуте у истраживачком раду у оквиру докторске дисертације наведене су у наставку:

X1: Коришћење дубоких неуронских мрежа у комбинацији са подацима добијеним даљинском детекцијом ће резултовати прецизнијим проценама садржаја SOC у поређењу са традиционалним регресионим моделима.

X2: Укључивање додатних обележја добијених даљинском детекцијом као што су закривљеност, висина или нагиб земљишта ће побољшати тачност дубоких неуронских мрежа у процени SOC.

X3: Употреба дубоких неуронских мрежа биће најефикаснија у процени садржаја SOC у подручјима са хомогеним типовима вегетацијског покривача.

Посебно изведена истраживачка питања која се тичу конкретне теме одабране за анализу примена дубоких неуронских мрежа за даљинску детекцију су:

ИП1: Каква је тачност дубоких неуронских мрежа у поређењу са традиционалним моделима регресије у процењивању SOC коришћењем података добијених даљинском детекцијом?

ИП2: Какав је ефекат укључивања додатних обележја добијених даљинском детекцијом, као што су закривљеност, висина или нагиб земљишта, на тачност дубоких неуронских мрежа у процени SOC?

ИП3: Која су ограничења и фактори који утичу на ефикасност дубоких неуронских мрежа у процени SOC у различитим типовима вегетације?

2 ПРЕГЛЕД ЛИТЕРАТУРЕ И ПРОБЛЕМИ У ПРОЦЕНИ ОРГАНСКОГ УГЉЕНИКА У ЗЕМЉИШТУ

Упркос постојању сагласности у оквиру научне заједнице о потенцијалу и многобројним користима које могу донети развој и примена техника складиштења и секвестрације органског угљеника у земљишту, оне остају ограничене у пракси. Основно питање које утиче на усвајање оваквих методологија је недостатак тачних и исплативих начина мерења садржаја SOC у горњем слоју земљишта (пошто на то највише утиче коришћење земљишта, пољопривредне праксе, итд.).

Велике варијације у проценама глобалних залиха SOC произилазе из разлика у периоду узорковања, интензитета и просторне резолуције база података профила земљишта, као и из разлика у приступима израчунавању самих процена [1]. Неравномерна дистрибуција геореференцираних профила земљишта широм света је још један разлог за тако велике варијације у проценама [2]. Поред тога, не постоји консензус када је у питању укључивање неорганског угљеника, различитих нивоа садржаја стена [3] и ефеката природних или антропогених појава (као што су поплаве, ерозија, пожар, ђубрење земљишта и орање [4]) приликом процене залиха угљеника.

Ако желимо да напредујемо у разумевању и управљању SOC-ом за добробит човечанства у целини, потребне су ефикасне и ефективне методологије за континуирано праћење SOC-а на глобалном нивоу.

Нажалост, методе за директно мерење угљеника у земљишту које се користе у праћењу SOC-а имају тенденцију да буду радно интензивне, скупе и непрактичне [5]. Ове процедуре подразумевају свеобухватне кампање узорковања земљишта, накнадна лабораторијска испитивања и опсежну обраду података [6] да би се добили подаци који се заиста могу означити као „темељна истина“ (енгл. ground truth).

Последњих година, даљинска детекција се показала као посебно ефикасан метод за праћење пољопривредних и еколошких промена [7][8] [9]. Технологија се ослања на различите сензоре и платформе, као што су сателитске констелације и беспилотни ваздушни системи за прикупљање података који се затим обично обрађују помоћу напредних алгоритама, често у домену машинског учења и дубоког учења [10].

Текући напредак у даљинском детектовању представља обећавајућу алтернативу конвенционалном праћењу SOC-а. Тот и Жожков пружају актуелан преглед различитих платформи и сензора за даљинску детекцију доступних данас [11].

У истраживању представљеном овде, фокус је на процењивању садржаја SOC само из сателитских података. Већина студија која се фокусира на одређивање SOC-а, међутим, ослања се на податке (спектограме) прикупљене од ручних сензора. Иако је прецизност постигнута на овај начин обично већа од употребе сателитских снимака, такви приступи се тешко могу скалирати да би омогућили континуирано праћење залиха угљеника на глобалном нивоу.

Гомез и сарадници [12] су представили рану, иако ограничену студију (засновану на само 146 узорка земљишта), која је упоређивала резултате који се могу постићи применом ML метода на VIS-NIR (енгл. visible - VIS, опсег електромагнетног спектра који је видљив људском оку, енгл. Near-Infrared - NIR, блиски инфрацрвени опсег) мерења са терена у односу на примену ML метода на хиперспектралне сателитске снимке. Сlike

су добијене са сензора Хиперион постављеног на сателиту EO-1, који, нажалост, више није функционалан и више не постоји активан хиперспектрални сателит који снима слике у VNIR-SWIR (енгл. Short-Waved Infrared - SWIR, краткоталасни инфрацрвени опсег) региону, што отежава реплицирање њиховог рада. Поред покушаја да моделирају цео скуп података који се користи у студији, аутори су покушали да се фокусирају на специфичне класе земљишног покривача (земљиште са усевима, земљиште за пашњаке) и одлучили су се за делимичну регресију најмањег квадрата (енгл. Partial Least Square Regression) као свој SOC предиктор. Gomez и сарадници су приметили да се SOC у њиховим земљиштима са усевима кретао између 0,54% и 1% и био је нижи него на земљишту за пашњаке, где је SOC био у распону од 1,08% до 5,1%. Они су проценили своју методологију на основу коефицијента детерминације R^2 (енгл. coefficient of determination) и корена средње квадратне грешке (енгл. Root Mean Square Error - RMSE). Модели засновани на сателитским снимцима нису имали добре резултате за земљиште са усевима (R^2 од 0,04 и RMSE од 0,11) и значајно су заостајали за моделима заснованим на ручним сензорима у смислу R^2 (R^2 од 0,16 и RMSE од 0,1). Међутим, када су процењени на земљишту за пашњаке и читавом скупу података, ова два приступа су постигла упоредиве и много боље перформансе. Приступ заснован искључиво на сателитским подацима у својој изворној резолуцији постигао је R^2 од 0,51, али је RMSE био прилично висок (0,73% SOC). Дакле, студија је показала да је земљишни покривач веома важан када је у питању моделирање и процена SOC-а на даљину.

Недавно су Ванг и сарадници [13] покушали да користе ML технике за процену залиха SOC у полусушним пашњацима источне Аустралије применом различитих техника машинског учења, са фокусом на процену утицаја узимајући у обзир сезонску фракцију покривеност земљишта на перформансе модела. Ова обележја су коришћена за проширење других ручно израђених обележја изведених из сателитских снимака, као и других климатских карактеристика које се детектују на даљину као што су падавине и температура и подаци о литологији. Они су обучавали и процењивали своје моделе користећи ограничену количину узорака земљишта (705). Користили су насумичне шуме (енгл. Random Forest - RF) [14], стабла побољшане регресије (енгл. Boosted Regression Trees - BRT)[15] и машине за векторе подршке (енгл. Support Vector Machines - SVM)[16] да би моделирали своје податке. RF приступ се показао најбољим и постигао је R^2 од 0,47 на њиховом скупу података.

Неколико студија је покушало да процени ефикасност хиперспектралних података добијених од сензора постављених на платформе за даљинску детекцију из ваздуха и прошири своје налазе како би проценили очекиване перформансе следеће генерације сензора за које се очекује да ће бити коришћени у будућности [17] [18]. Без обзира што је фокус овог истраживања на мултиспектралне податке, вреди напоменути да, иако се ослањајући на веома ограничен скуп узорака земљишта (81) добијених за 7 км² подручје у Луксембургу, од којих је 40% било коришћено као тест сет, Стајнберг и сарадници су постигли релативно висок R^2 (0,74) и RMSE од 0,22% за SOC користећи autoPSLR примењен на хиперспектралне податке са сензора постављеног на платформу за даљинску детекцију из ваздуха [18]. Када буде доступно довољно хиперспектралних података, методологија која је предложена у овом истраживању може се лако прилагодити том домену, што доводи до још бољих перформанси.

Током последње деценије дубоко учење је револуционисало област машинског учења

и вештачке интелигенције и постало је доминантна парадигма у том домену. Кључни напредак у односу на претходно коришћене методе је то што се приступ ослања на учење од краја до краја, што омогућава моделима ML да науче обележја на којима доносе своје одлуке и процењују директно из сирових улазних података, уместо да се ослањају на обележја произведена од стране људског стручњака [19].

Јуан и сарадници пружају преглед примена и класичних неуронских мрежа и DL модела за праћење параметара животне средине помоћу података добијених од даљинског сензора[20]. Они показују да је DL надмашио традиционалне ML моделе и довео до значајних побољшања у многим применама, укључујући мапирање земљишног покривача, параметара вегетације, влажности земљишта, евапотранспирације, предвиђање приноса у пољопривреди итд.. Аутори исправно истичу ограничење DL приступа, које је повезано са релативно ограниченим количинама доступних података за тренирање, као и потенцијалом да се примени трансфер учења како би се заобишао овај проблем. Они помињу два типа трансферног учења: засновано на региону и базирано на подацима. Први се односи на претходни тренинг на географском региону за који је доступно довољно података и прилагођавање модела другом региону са ограниченим подацима. У ML заједници ово се обично назива финим подешавањем. Други тип је више у складу са значењем трансферног учења у домену ML и односи се на преношење модела обучених на подацима добијеним од сензора или групе сензора на други сензор(е). У овом истраживању ми примењујемо трећу врсту трансферног учења, уобичајену у заједници која се бави рачунарским видом [21], где се почетни модел обучава на истом типу улазних података (Sentinel-2), али за другачији визуелни задатак (класификација земљишног покривача) и користи се као екстрактор обележја за коначни модел (који у овом истраживању врши процену SOC).

Док је прва примена о којој су Јуан и сарадници расправљали јесте покривач земљишта, у овој студији се не помињу никакви приступи процени SOC. Поред тога, док су разматрани приступи засновани на различитим DNN архитектурама (већина се ослањају на конволуционе неуронске мреже), ниједан није идентификован у студији који користи U-Net модел.

Рахлин и сарадници су, међутим, успешно применили U-Net са Ловаш-софтмакс губитком за класификацију земљишног покривача користећи RGB податке који су доступни као део DeepGlobe Challenge-a [22].

Јанг и сарадници су користили CNN да покушају да добију процене SOC-а за централну локацију на основу улазних података који покривају околни регион [23]. Улаз за њихов модел биле су варијабле средине у комбинацији са MODIS MCD12K2 фенолошким варијаблама. Они су обучили и проценили свој приступ на ограниченом скупу узорака (733), прикупљених у кинеској провинцији Анхуи. Ово је ограничило сложеност CNN-а коју су могли да користе, пошто у студији није коришћено трансфер учење, али је CNN прошао боље од модела насумичних шума, постигавши скромни резултат R^2 од 0,26.

Емади и сарадници [24] су се фокусирали на северни Иран и користили су велики број улазних обележја (105). Већина су били индекси које су израдили људски експерти извучени из сателитских снимака Landsat-8 и MODIS, али њихов унос је укључивао и параметре везане за топографију, као што су закривљеност, нагиб итд. Користећи скуп података од 1879 композитних узорака тла и ослањајући се на десетоструку унакрсну

валидацију, упоредили су перформансе неколико традиционалних ML алгоритама (машине за векторе подршке, вишеслојни перцептрон, регресиона стабла одлучивања, насумичне шуме и екстремно повећање градијента) са DL моделом приликом предвиђања SOC. DL модел који је показао најбоље резултате у студији био је прилично једноставна потпуно повезана неуронска мрежа, са 7 скривених слојева и 50 неурона у сваком од њих, али је ипак победио друге тестиране методе. Аутори наводе релативно велику вредност R^2 од 0,65, са RMSE од 0,75% SOC.

У недавној студији Касталди и сарадници [25] су евалуирали способност временске серије Sentinel-2 да процени садржај органског угљеника и глине у земљишту на локалном нивоу у обрадивим површинама. ML ток који су предложили у великој мери се ослања на људски инжењеринг, како у погледу обележја које произилазе из Sentinel-2 снимака (енгл. Normalized Difference Vegetation Index - NDVI, енгл. Normalized Burn Ratio 2 - NBR2, енгл. Bare Soil Index - BSI, енгл. Sentinel-2 Water Index - S2WI), тако и у смислу начина на који се користе за креирање улаза за њихове моделе машинског учења. Што се тиче моделирања, нису се определили за дубоке неуронске мреже, већ за алгоритам QRF (енгл. Quantile Regression Forest), QRF са додатном географском дужином и ширином као коваријатима, као и за хибридни приступ, LMEM (енгл. Linear Mixed Effect Model), који је укључивао просторну аутокорељацију својстава тла. Док последњи модел узима у обзир просторне информације до одређене тачке, њихов приступ суштински базиран на пикселима, што се разликује од оног који је предложен у овом истраживању. Поред тога, аутори студије су имали за циљ да процене способност свог приступа у веома ограниченом сценарију, тако што су креирали и процењивали моделе за свако од својих тестних места посебно. Није учињен ниједан покушај да се направи јединствен модел који би се могао применити глобално, или бар за велики део Земљине површине. Према томе, резултати које су постигли могли би се посматрати као нека врста перформансе плавог неба до које би могао доћи глобални модел користећи Sentinel-2 снимке као улаз. R^2 најбољих модела Касталди и сарадника се кретао од 0,26 до импресивних 0,96 за различите локације, са просечним R^2 од 0,67, RMSE (у % SOC) се кретао од 0,09 до 0,22 и био је 0,152 у просеку.

У овом моменту постоји ограничен број пројеката који су фокусирани на органски угљеник из земљишта у контексту регулаторне усклађености или добровољних тржишта угљеника. Наведена тржишта представљају механизме за ублажавање емисије угљен-диоксида, омогућавајући субјектима да купују или продају карбонске кредите као начин надокнађивања свог угљеничног отиска. Тренутно постоји мање од 60 SOC пројеката који обезбеђују мање од 50 Кт CO_2 - еквивалентног (CO_2e) уклањања годишње који се односе на примену у пољопривреди и на пашњацима [26], што указује на искоришћеност малог дела процењеног потенцијала за ублажавање утицаја емисије угљен-диоксида (прецизније износи 0.0001% процењеног потенцијала)[27].

Поређења ради, постоји много више пројеката ублажавања емисије угљен-диоксида у сектору шумарства, са преко 1.500 таквих пројеката који покривају 12 милиона хектара земље [28]. Ово сугерише да, иако SOC има значајан потенцијал као стратегија за ублажавање емисија угљеника, још увек није широко прихваћен на тржиштима угљеника у поређењу са пројектима угљеника заснованим на шумама. Ово ограничено прихватање SOC пројеката на тржиштима угљеника може бити последица низа фактора, укључујући изазове повезане са прецизним мерењем стопе секвестрације SOC,

недостатка робусних методологија за квантификацију SOC у различитим контекстима и релативно високе трансакцијске трошкова повезаних са имплементацијом SOC пројеката.

Политичке тежње за бављење и решавање проблема секвестрације угљеника у земљишту и побољшавање здравља земљишта још увек нису у потпуности реализоване у смислу саме имплементације. Само осам земаља је укључило циљеве везане за земљиште у оквиру ублажавања утицаја у своје национално одређене доприносе оквирној конвенцији Уједињених нација о климатским променама (енгл. United Nations Framework Convention on Climate Change - UNFCCC) [29]. Међутим, UNFCCC сада почиње да се фокусира на улогу пољопривреде и земљишта у ублажавању климатских промена, укључујући важност побољшања органског угљеника у земљишту, здравља земљишта и плодности у системима усева и пашњака, као и на интегрисане системе који укључују управљање водама [30].

Један од примера представља програм Организације за храну и пољопривреду Уједињених Нација (енгл. Food and Agriculture Organization - FAO) под називом Глобална процена потенцијала секвестрације органског угљеника у земљишту (енгл. Global Assessment of Soil Organic Carbon Sequestration Potential - GSOCsek).

Иницијатива има за циљ процену глобалног потенцијала земљишта за секвестрацију угљеника. Програм је део напора FAO-а да промовише праксе одрживог управљања земљиштем и смањи емисију гасова стаклене баште. GSOCsek пружа свеобухватну анализу потенцијала различитих пракси управљања земљиштем за секвестрацију угљеника и даје креаторима политике и управницима земљишта информације које су им потребне да донесу информисане одлуке о управљању њиховим земљиштем ради секвестрације угљеника.

Програм такође подржава земље у развијању и имплементацији националних система за обрачунавање угљеника, који се користе за извештавање о емисијама гасова стаклене баште и везивању угљеника кроз употребу земљишта, промене коришћења земљишта и шумарске активности у складу са UNFCCC.

Такође, постоји и међународна иницијатива „4 на 1000“ коју је 2015. покренула француска влада на климатској конференцији COP21 са циљем повећања секвестрације угљеника у земљишту за 0,4% годишње како би се ублажиле климатске промене. Назив „4 на 1000“ односи се на циљ повећања залиха органског угљеника у земљишту за 0,4% годишње, што је еквивалентно годишњем повећању од око 4 милијарде тона CO₂ у земљишту.

Иницијатива има за циљ да промовише праксе одрживог коришћења земљишта које побољшавају секвестрацију угљеника у земљишту, укључујући агрошумарство, конзервациону пољопривреду и употребу покровних усева.

Иницијатива такође настоји да побољша здравље тла, повећа сигурност хране и промовише рурални развој. Подржале су је многе земље и организације широм света. Овај повећан фокус на секвестрацију угљеника у земљишту указује на све веће препознавање значаја здравља земљишта у контексту ублажавања климатских промена и одрживе пољопривреде.

Постоји растући тренд међу компанијама да усвоје праксе које промовишу повећање органског угљеника у земљишту у њиховим пољопривредним ланцима вредности као начин да се побољша њихова дугорочна одрживост и профитабилност [29]. Повећано ин-

тересовање за унапређење SOC-а је вођено идејом да такве праксе могу пружити бројне заједничке користи. На пример, повећање SOC може побољшати плодност земљишта, смањити ерозију земљишта и побољшати способност земљишта да задржи воду. Штавише, праксе побољшања SOC-а могу се применити без потребе за додатном површином, имају минималан водени отисак и не захтевају промене у начину коришћења земљишта, што их чини лаким за имплементацију [27] [31].

Све наведено је довело до тога да праксе побољшања SOC-а постану атрактивна опција за компаније које желе да побољшају одрживост и отпорност својих пољопривредних операција.

3 ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА И МАШИНСКО УЧЕЊЕ

За већину људи, када чују појам „вештачка интелигенција“, вероватно прва асоцијација која им падне на памет су поуздани и услужни робот-адроид СЗРО или R2D2 из серијала Ратови Звезда или смртоносни Терминатор из познатог научно фантастичног серијала. За истраживаче који се баве вештачком интелигенцијом кључно питање представља могућност конструисања интелигентних машина које се понашају као људска бића и која приказују интелигентно понашање [32]. Вештачка интелигенција не представља неку футуристичку фантазију, већ мултидисциплинарну област која се све више примењује и која продире у све аспекте модерног друштва кроз имплементације у области медицине, образовања, пољопривреде, аутономних возила, финансијама, производњи [33] и читавом низу других области.

Развој вештачке интелигенције започео је почетком 40-тих година 20. века, тачније 1942. године са кратком причом Исака Асимова „Коло-наоколо“ [34] где су дефинисана 3 закона Роботике:

- Робот не сме повредити људско биће нити, својом неактивношћу, дозволити да људско биће буде повређено,
- Робот мора поштовати наређења људских бића, осим ако се та наређења не косе са Првим законом,
- Робот мора да штити сопствену егзистенцију, осим ако се то не коси са Првим и Другим законом.

Прича је инспирисала бројне генерације истраживача и научника из области вештачке интелигенције, роботике и рачунарских наука [35]. Прву дефиницију вештачке интелигенције дао је један од пионира у области Џон Макарти 1955. године: „Циљ вештачке интелигенције је да развије машине које се понашају као да су интелигентне“ [32]. Дефиниција вештачке интелигенције по Оксфордском речнику енглеског језика гласи: „Вештачка интелигенција представља капацитет рачунара или других машина да покажу или симулирају интелигентно понашање.“ У Encyclopedia Britannica дефиниција гласи: „Вештачка интелигенција је способност дигиталних рачунара или рачунаром контролисаних робота да решавају проблеме који су обично повезани са вишим интелектуалним способностима обраде људи“ [36]. Имајући у виду да вештачка интелигенција представља веома широко поље дефиниција скована од стране Елејн Рич делује прикладно: “Вештачка интелигенција је студија о томе како натерати рачунаре да раде ствари у којима су људи у овом тренутку бољи“ [37].

Проналажење свеобухватне и потпуне дефиниције вештачке интелигенције представља велики изазов услед жеље да се опишу рачунари који опонашају когнитивне функције повезане са људским умом, као што су учење, перцепција, расуђивање и решавање проблема [38].

У почетном периоду развоја вештачке интелигенције преовладавали су симболички приступи, који су се заснивали на великом броју „ако-онда“ изјава које су омогућавале

рачунару да предузме одређене радње под одређеним условима. Изјаве „ако-онда“ типа су се називале правилима. Сва правила у одређеном систему вештачке интелигенције су колективно названа база знања. Базе знања су прављене од стране стручњака из одређених области у домену проблема који су требали да буду решени од стране система вештачке интелигенције. Симболички приступи су били доминантни у раној фази вештачке интелигенције јер су алгоритми закључивања били засновани на провереним математичким методама и јер су могли да се користе да се дође до веома објашњивих резултата кроз транспарентан процес одлучивања, као што су експертски системи популарни 1970-их и 1980-тих година [39].

Симболички приступи проблемима из домена вештачке интелигенције су подразумевали да се прате принципи рачунарског програмирања, јер се сматрало да може да се напише велики програм који би научио рачунар да изврши било који сложен задатак [38]. Међутим, ту се наишло на круцијалне проблеме у виду количине правила која требају да се напишу како би се обухватила сва комплексност реалног света, као и последично одржавање свих конструисаних правила и њихових међусобних релација [40]. Јавила се потреба за дефинисањем другачијег приступа заснованог на идеји дизајнирања алгоритама са способношћу учења, помоћу којих су рачунари способни да аутоматски побољшавају своје способности за решавање било каквог проблема из домена вештачке интелигенције путем искуства [41].

Претходно искуство се убацује у модел са способношћу учења у виду „података за тренирање“ из којих ће алгоритам да учи. Од 80-тих година 20. века поље изучавања наведених алгоритама за аутоматско учење брзо је заузело своје место као истакнута подобласт вештачке интелигенције под називом машинско учење [41]. Сам термин машинско учење сковао је Артур Самуел, истраживач у компанији International Business Machines (IBM) и пионир у области вештачке интелигенције [42].

Алгоритми машинског учења су се показали као способни да науче нетривијалне, понекад сложене класификације за разне врсте примена, али је било потребно добро дефинисати обележја (карактеристике) података који се користе за тренирање [32]. Посао истраживача, тј. људског фактора, је да идентификује и направи разумно-мали скуп карактеристика података. Наведени приступ подразумева спор, дуготрајан и веома захтеван поступак од стране истраживача како би се направили скупови података који би се користили као улаз за алгоритме машинског учења. Поред тога наметнуло се додатно питање како одабрати праве карактеристике података на основу којих ће се правити улазни скупови података? Што је довело до проблема под називом „клетва димензионалности“ уведеним од стране Ричард Белмана [43] који описује динамичну природу повећања димензија података и резултирајућег експоненцијалног повећања рачунарске моћи потребних за њихову обраду и/или анализу.

У машинском учењу, карактеристика објекта може бити одлика или својство које га дефинише. Свака карактеристика представља димензију, а група димензија ствара инстанцу односно појединачни пример на основу којег алгоритам учи. Самим додавањем нових карактеристика података повећавамо и број димензија података. Како се димензионалност повећава, број инстанци потребних да би се постигао задовољавајући ниво перформанси било ког алгоритма машинског учења расте експоненцијално. Такође, у својој студији из 1968. године, Гордон Хјуз је закључио да са фиксираним бројем примера или инстанци за тренирање алгоритма, предиктивна моћ било ког класификатора

прво расте како се број карактеристика података (димензија) повећава, али да се након одређене вредности броја димензија, перформансе погоршавају [44]. Тако да је феномен „клетве димензионалности“ познат и под називом Хјузов феномен.

Нова етапа у развоју вештачке интелигенције наступила је 1995. године са појавом нових алгоритама са способношћу учења у оквиру подобласти вештачке интелигенције под називом дубоко учење, који су показивали обећавајуће резултате у решењу проблема „клетве димензионалности“. Нови приступ је омогућио рачунарима да уче из искуства и разумеју свет у смислу хијерархије концепата, при чему је сваки концепт дефинисан кроз своје односе према једноставнијим концептима. Прикупљањем знања из искуства, овај приступ избегава потребу да људски оператери формално специфицирају сва знања која су потребна рачунару. Хијерархија концепата омогућава рачунару да научи компликоване концепте тако што их гради од једноставнијих.

Наведене карактеристике алгоритама дубоког учења су довеле до тога да дубоко учење буде идентификовано као најбрже растући тренд у анализи великих количина података и да се сматра за једну од десет револуционарних технологија у 2013. години [45]. Решења заснована на дубоком учењу су у великој мери примењивана и промовисана од стране великих технолошких компанија попут Google-а, Microsoft-а, Facebook-а. На таласу ових успеха методе дубоког учења су нашле своју примену у широком спектру области. Једно од поља где се показао јасан потенцијал за примену дубоког учења је и област даљинске детекције (енгл. Remote Sensing) [46].

У општем смислу даљинска детекција представља прибављање физичких података о објекту без додира или контакта [47]. У смислу истраживања презентованог у оквиру докторске дисертације прикладнија би била дефиниција: “Даљинска детекција представља праксу прикупљања информација о копненим и воденим површинама Земље коришћењем слика добијених из „птичје перспективе“, употребом електромагнетног зрачења из једног или више региона електромагнетног спектра, које се рефлектује или емитује са површине Земље.” [48]. У области даљинске детекције, конкретно код сателитских снимака, постоје специфичности као што су мултимодалност података, просторна одређеност, временска одређеност и сама количина доступних података [46]. Наведене карактеристике представљају изазове за примену алгоритама дубоког учења.

3.1 Основни концепти у машинском учењу

Концепт машинског учења може се схватити као процес откривања вођен подацима, где је кључ у разликовању образаца унутар обимних података. Овај процес почиње са улазним подацима: сложеним низом варијабли и мерења, често обимним и вишедимензионалним, прикупљених из различитих извора и пажљиво претпроцесираних да би се осигурала униформност и употребљивост.

У срцу овог процеса је модел машинског учења - попут математичког маистра који оркестрира ансамбл алгоритама за интерпретацију симфоније података. Модели се крећу од једноставне линеарне регресије, која тежи проналажењу директног односа између улаза и излаза, до комплексних мрежа дубоког учења које емулирају синаптичку повезаност људског мозга. Архитектура модела се пажљиво подешава да идентификује темељне везе унутар података, користећи статистичку инференцију за претварање сирових информација у дестиловану суштину разумевања.

Кулминација овог процеса су излазни подаци, који представљају хипотезе које је модел научио - инференцију функције која мапира улазе на излазе. То се може манифестовати као предиктивне вредности, класификације, или чак генеративни модели који симулирају тачке података. Излаз није само пасиван резултат већ сведочанство о способности модела да учи, генерализује и пружа применљиве увиде.

Ова сложена интеракција између улаза, модела и излаза наглашава се у теоријским основама статистичке теорије учења, која пружа оквир за квантификацију и анализу перформанси модела. Она испитује компромисе између пристрасности и варијансе, истражује сложеност структура модела и тежи успостављању равнотеже између тежње за тачношћу и императива једноставности.

Сваки модел машинског учења је дефинисан улазом, алгоритмом машинског учења, моделом знања и излазом приказаном на Слици 1.



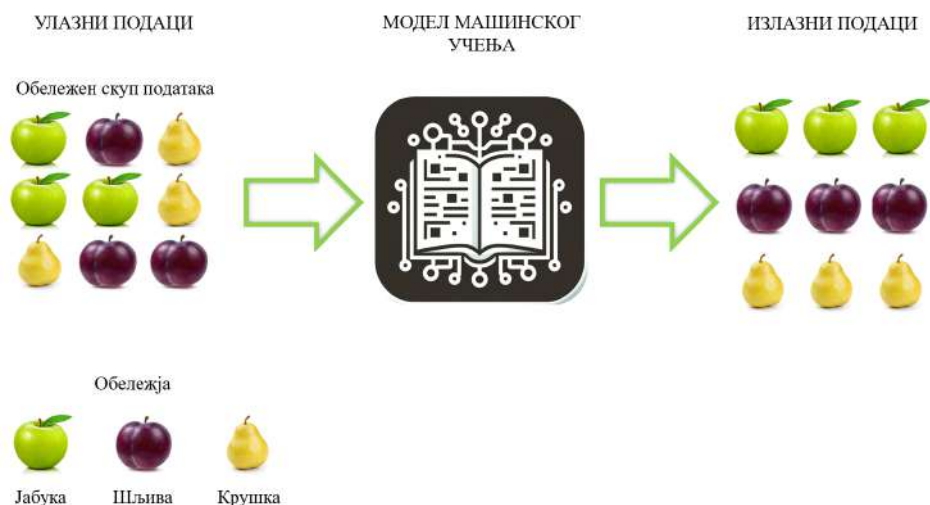
Слика 1: Модел машинског учења

3.1.1 Учење са надзором, учење без надзора и учење путем подстицаја

Све методе машинског учења захтевају претходно прикупљање података који ће се користити приликом тренирања.

Учење са надзором се бави оним проблемима из домена машинског учења где прикупљање података за тренирање (или скуп тренинг опсервација) садржи податке, како о вредностима обележја тако и о зависним променљивим. Другим речима, скуп тренинг опсервација се састоји од улазно-излазних парова. За сваки улаз у скупу тренинг опсервација, знамо његов одговарајући излаз, што се може искористити за навођење алгоритма машинског учења ка тачном излазу, односно представља вид надзора при учењу. Методе учења са надзором су добро проучене у пољу машинског учења и обично гарантују добре перформансе модела, све док на располагању постоји довољан број улазно-излазних парова. Међутим, прикупљање и сачињавање скупа тренинг опсервација који ће садржавати адекватан број улазно-излазних парова за учење са надзором често захтева обележавање скупова од стране људи, што у пракси може захтевати пуно труда. Приказ учења са надзором - Слика 2.

Насупрот томе, методе учења без надзора се баве оним проблемима из домена машинског учења где прикупљање података за тренирање (или скуп тренинг опсервација)



Слика 2: Учење са надзором

садржи податке само о вредностима обележја. Дobar алгоритам за учење без надзора требало би да буде у стању да открије неке критеријуме за груписање сличних улаза користећи само информације из свих доступних улаза који се налазе у скупу тренинг опсервација, при чему се два улаза сматрају за слична само када се очекује да дају исту излазну ознаку (класу). Основна потешкоћа у учењу без надзора лежи у томе како знати који су улази слични када су њихове излазне ознаке недоступне. Учење без надзора представља много тежи проблем због недостатка информација које би се користиле за надзор рада алгоритма машинског учења. У учењу без надзора, обично је једноставније прикупити скупове тренинг опсервација јер то не захтева додатне људске напоре да се свака опсервација означи са одговарајућим излазом. Услед свих наведених карактеристика учење без надзора углавном остаје и даље отворен проблем у домену машинског учења. Приказ учења без надзора - Слика 3.



Слика 3: Учење без надзора

Између учења са надзором и учења без надзора, постоје и комбиновани приступи, где можемо комбиновати малу количину означених података са великом количином неозначених података током тренирања модела машинског учења. Ове методе учења се често називају полу-надзираним учењем.

У другим случајевима, ако су праве вредности зависних променљивих сувише тешке

или скупе за добијање, можемо користити друге лакше доступне информације, које су само делимично релевантне за праве излазе, које можемо користити као вид слабијег сигнала за тренирање алгорита машинског учења. Ове методе се називају учење са слабом надзором.

Посебну област машинског учења представља приступ учење путем подстицаја. Учење путем подстицаја се састоји из 4 главна елемента:

- Агента – програма који може да се тренира и који извршава задатке који су му додељени
- Окружења – реалне или виртуалне средине где агент извршава своје задатке
- Акције – потез агента који резултује у промени статуса у окружењу
- Награде – Негативне или позитивне накнаде базиране на акцији.

Учење путем подстицаја се фокусира на проналажење оптималних акција агената које максимизирају награду у оквиру одређеног окружења. Агент се тренира да усаврши своје акције како би остварио највишу могућу кумулативну награду [49].

За разлику од учења са надзором, агент не добија никакав јак надзор од околине у погледу тога која је најбоља акција на сваком кораку. Уместо тога, агент само повремено прима одређене нумеричке награде (позитивне или негативне). Циљ учења путем подстицаја је да алгоритам научи које акције треба предузети у сваком одређеном стању система како би максимизирао кумулативну награду у дужем периоду. Приказ учења путем подстицаја на примеру робота за усисавање - Слика 4.

Учење путем подстицаја се сматра изузетно изазовним задатком јер агент мора научити како да истражује потенцијално огромне просторе за претрагу само на основу слабих сигнала награде.

3.1.2 Класификација и регресија

Класификација и регресија представљају технике машинског учења које се користе у оквиру учења са надзором. Упоредни приказ класификације и регресије на предвиђању температуре - Слика 5.

Предиктивно моделирање представља проблем развоја модела путем коришћења историјских података за предвиђање на новим подацима за које немамо одговоре. Предиктивно моделирање се може описати као математички проблем апроксимирања функције мапирања (f) од обележја (x) до зависних променљивих (y). Задатак алгорита за моделирање је да пронађе најбољу функцију мапирања коју може, узимајући у обзир потребно време и расположиве ресурсе.

Предиктивно моделирање типа класификације је задатак апроксимирања функције мапирања (f) од обележја (x) до дискретних зависних променљивих (y). Зависне променљиве се често називају ознакама или категоријама. Функција мапирања предвиђа класу или категорију за дату опсервацију.

Најједноставнији облик класификације представља двокласни проблем, који укључује додељивање инстанце једној од две класе. Овај тип проблема је такође познат као



Слика 4: Учење путем подстицаја

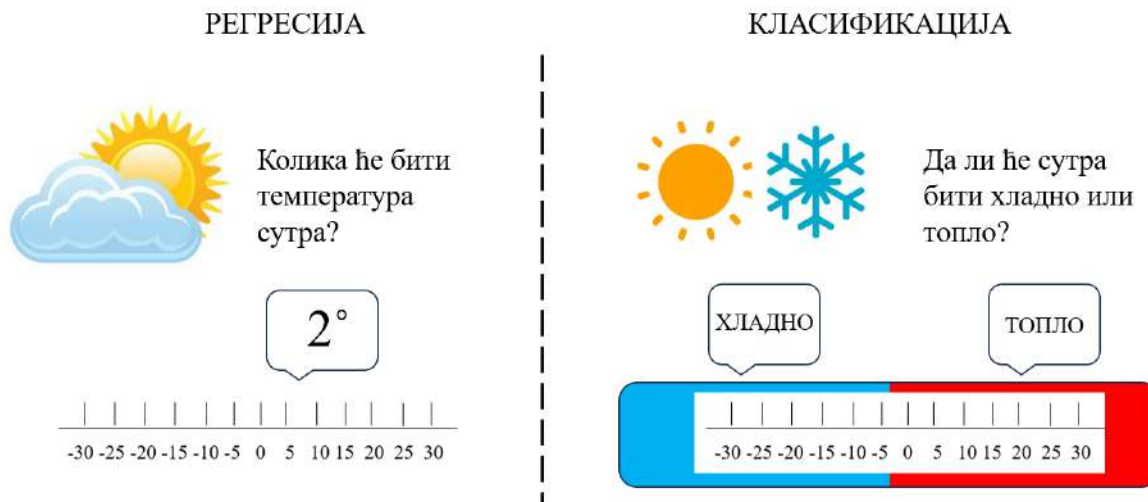
проблем бинарне класификације. Примери проблема бинарне класификације укључују откривање нежељене електронске поште, као и откривање преваре у трансакцијама кредитним картицама.

Насупрот томе, проблем са више од две класе се често назива проблемом вишекласне класификације. У овом случају, опсервација мора бити класификована у једну од неколико класа. Пример проблема са вишекласном класификацијом представља класификовање слика животиња у различите врсте животиња.

Коначно, у неким случајевима, опсервација може бити сврстана у више од једне класе. Овај тип проблема се назива проблем класификације са више ознака. Пример проблема са класификацијом са више ознака представља додељивање жанровске ознаке филмовима.

Предиктивно моделирање типа регресије је задатак апроксимације функције мапирања (f) од обележја (x) у континуалну зависну променљиву (y). Континуална зависна променљива је реална вредност, као што је цео или децимални број. У питању су често квантитативне вредности, као што су количине и величине. На пример ако желимо да предвидимо цену некретности на одређеној локацији.

Вредности обележја у проблему регресије могу бити реалне или дискретне. Реалне вредности обележја могу да поприме било коју вредност унутар одређеног опсега, као што је старост особе или тежина објекта. Дискретне вредности обележја могу по-



Слика 5: Регресија и класификација

примити ограничен број вредности, као што је број спаваћих соба у кући или марки аутомобила.

У неким случајевима може постојати више обележја које се користе за предвиђање зависне променљиве. Ово се зове проблем мултиваријантне регресије. Пример за то би било предвиђање цене куће на основу њене величине, броја спаваћих соба, локације и других карактеристика.

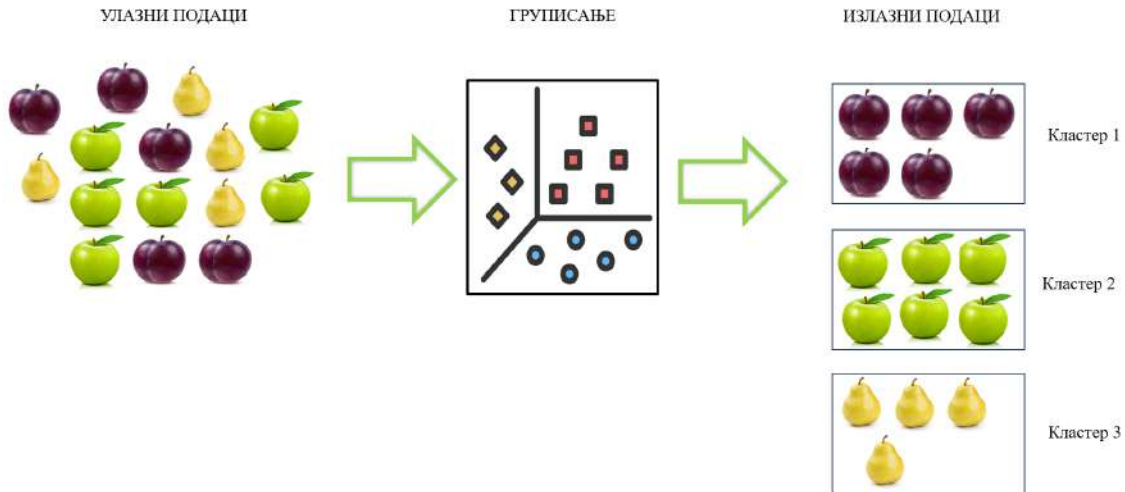
Посебан тип проблема регресије представља проблем предвиђања временске серије, где су вредности обележја поређане по времену. У овој врсти проблема, циљ је предвиђање будућих вредности на основу прошлих вредности. Пример за то би било предвиђање цене акција компаније на основу њених прошлих перформанси.

3.1.3 Груписање

Груписање у машинском учењу представља технику која се употребљава у оквиру проблема из домена учења без надзора, а примарно се бави идентификацијом образаца или структура у скуповима података који немају унапред дефинисане ознаке или категорије. Суштина груписања представља сврставање опсервација или тачака података у подскупове, познате као кластери, на основу заједничких карактеристика, осигуравајући да чланови унутар сваког кластера показују висок степен сличности, док чланови различитих кластера исказују значајне разлике.

Процес почиње одабиром одговарајућих мера сличности, које служе као основа за груписање тачака података. Две најчешће коришћене мере су косинусна удаљеност и Еуклидска удаљеност. Косинусна удаљеност се типично користи за процену сличности између два вектора, што је идеално за сценарије у којима је величина вектора мање важна. Насупрот томе, Еуклидска удаљеност се користи за израчунавање стварне просторне удаљености између тачака у вишедимензионалном простору, што је погодно за случајеве када је величина разлике важна. Приказ груписања је дат на Слици 6.

Методологије и претпоставке које леже у основи груписања могу значајно утицати на



Слика 6: Груписање

учинак и прикладност наведене технике на одређени скуп података. На пример, унутрашња компактност може се користити за процену блискости чланова унутар кластера, док мере раздвајања процењују степен разликовања између различитих кластера. Неке напредне технике груписања такође користе методе засноване на процењеној густини и повезаности графа за одређивање структуре и расподеле тачака података унутар скупа података.

Анализа кластера није само статички процес већ укључује динамичко доношење одлука у вези са бројем кластера. Хеуристичке технике попут лакат методе (енгл. elbow) могу се користити за одређивање оптималног броја кластера, што је кључно јер директно утиче на гранулацију и корисност резултата груписања. Метода се састоји од цртања графика објашњене варијације као функције броја кластера и одабира лакта криве као броја кластера које треба користити.

У практичној примени, груписање представља изузетно вредан алат у различитим областима. На пример, у маркетингу се може користити за сегментацију купаца на основу понашања при куповини или демографије, омогућавајући циљаније и ефикасније маркетиншке стратегије. У анализи података, кластеринг помаже у идентификацији природних груписања у подацима, што може бити кључно за детекцију аномалија, препознавање образаца или поједностављење сложених скупова података.

3.1.4 Једноставни и сложени модели

У машинском учењу, извршавамо алгоритме учења над подацима за тренинг да бисмо направили одређене математичке моделе за доношење одлука. У смислу избора конкретног модела који ће се користити у учењу, обично морамо да направимо разуман избор између једноставних модела и сложених модела. Генерално, линеарни модели се третирају као једноставни модели, док се нелинеарни модели посматрају као сложени модели, јер нелинеарни модели имају способност да препознају много компликованије обрасце у дистрибуцијама података од линеарних.

Једноставан модел захтева много мање рачунарских ресурса и може се поуздано

истренирати на основу много мањег скупа за тренирање. У многим случајевима можемо извести потпуну теоријску анализу за једноставне моделе, што нам даје боље разумевање основног процеса учења. Међутим, перформансе једноставних модела често брзо дођу до границе засићења, тј. не долази више до побољшања перформанси, с порастом количине података за тренирање. У многим практичним случајевима, једноставни модели могу постићи само осредње перформансе јер не успевају да се носе са компликованим обрасцима, што је норма у скоро свим практичним применама.

Са друге стране, сложени модели захтевају много више рачунарских ресурса приликом учења и потребно је да припремимо много више података за тренирање да бисмо их поуздано научили. Због сложених функционалних облика, за многе сложене моделе не постоји никаква теоријска анализа. Стога је учење сложених модела често веома незгодан процес такозване црне кутије (енгл. black-box) и обично захтева много необјашњивих трикова да би се постигли оптимални резултати.

3.1.5 Параметарски и непараметарски модели

Када бирамо модел за проблем машинског учења, постоје два различита типа. Такозвани параметарски модели (тзв. коначно-димензионални модели) представљају моделе који имају претпостављену функционалну форму и потпуно су одређени фиксираним скупом параметара модела. На пример ако се одлучимо да користимо за тренирање модела метод линеарне регресије он се може у потпуности спецификовати са два параметра (константом и нагибом). По правилу линеарни модели су параметарски модели.

Насупрот њима, такозвани непараметарски модели (тзв. модели без дистрибуције) не претпостављају функционални облик основног модела, и што је још важније, сложеност таквог модела није фиксирана и може зависити од доступних података. Другим речима, непараметарски модел не може бити у потпуности спецификован фиксним бројем параметара. На пример, ако се одлучимо да за тренирање модела користимо метод стабла одлучивања, нећемо претпоставити функционални облик модела, а ни величина стабла одлучивања обично није фиксирана. У случају да имамо више података за тренирање то нам може омогућити да направимо веће стабло одлучивања.

Генерално, лакше је руковати параметарским моделима него непараметарским моделима јер се увек можемо фокусирати на процењивање фиксираних параметара за било који параметарски модел. Процена параметара је увек много једноставнији проблем од процењивања произвољног модела без познавања његовог облика.

3.1.6 Недовољно и превелико прилагођавање

Све методе машинског учења се ослањају на податке за тренирање. Интуитивно говорећи, подаци за тренирање алгоритама машинског учења садрже важне информације о одређеним правилностима које желимо да научимо помоћу модела, које неформално називамо компонентом сигнала. С друге стране, подаци за тренирање такође неизбежно садрже неке небитне или чак ометајуће информације, које се називају компонентом шума. Главни извор шума су варијације узорковања приказане у било ком коначном скупу насумичних узорака. Ако насумично извучемо неке узорке, чак и из исте дистрибуције, два пута, нећемо добити идентичне узорке. Ова варијација се посматра као компонен-

та шума у прикупљеним подацима. Шум такође може настати због грешака приликом мерења или самог забележавања података.

Уопштено говорећи, све прикупљене податке за тренирање можемо представити као комбинацију две компоненте, односно збир компоненте сигнала и компоненте шума, као што је приказано на Слици 7. где можемо видети да компонента сигнала представља неке правилности у подацима, док компонента шума представља неке непредвидиве, веома флукутирајуће остатке.



Слика 7: Компоненте података за тренирање[40]

Односи између компоненте сигнала и компоненте шума у подацима за тренирање алгоритама машинског учења нас доводе до два веома важна концепта у машинском учењу – недовољног прилагођавања (енгл. *underfitting*) и превеликог прилагођавања (енгл. *overfitting*).

Претпоставимо да учимо једноставан модел из скупа података за тренирање. Ако је коришћени модел исувише једноставан да би обухватио све правилности у компоненти сигнала, научени модел ће дати веома лоше резултате на подацима за тренирање, а поготово би лоши резултати били на подацима које модел „види“ први пут. Ова појава се назива андерфитинг.

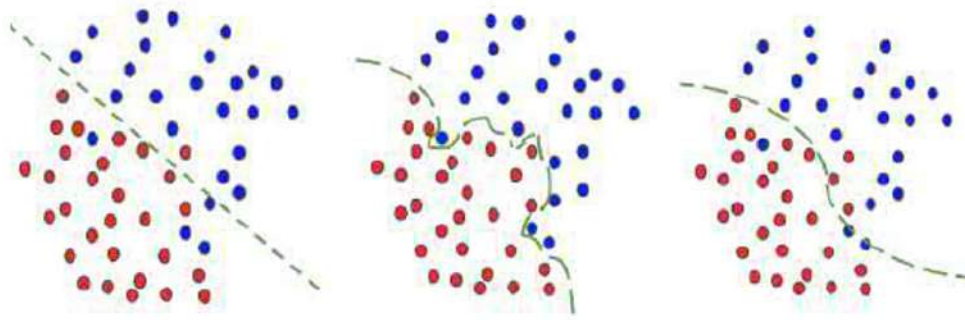
С друге стране, ако је коришћени модел превише сложен, процес учења може приморати моћан модел да се савршено прилагоди насумичној компоненти шума док покушава да ухвати правилности у компоненти сигнала. Штавише, савршено прилагођавање компоненте шума може ометати модел да ухвати све правилности у компоненти сигнала јер јако флукутирајући шум може више да утиче на исход учења када се користи сложени модел.

Додатно, савршено прилагођавање компоненте шума је бескорисно јер ћемо се суочити са потпуно другом компонентом шума у другом скупу опсервација. Ово ће довести до непожељног феномена у машинском учењу под називом оверфитинг. Проблеми недовољног и превеликог прилагођавања се јављају како код проблема класификације тако и код проблема регресије.

На Слици 8. имамо приказане случајеве недовољног, превеликог и одговарајућег прилагођавања.

3.1.7 Компромис између пристрасности и варијансе

Укупна очекивана грешка алгорита машинског учења на скупу података које „види“ први пут може се разложити на два извора, на пристрасност модела услед недо-



Недовољно
прилагођавање

Превелико
прилагођавање

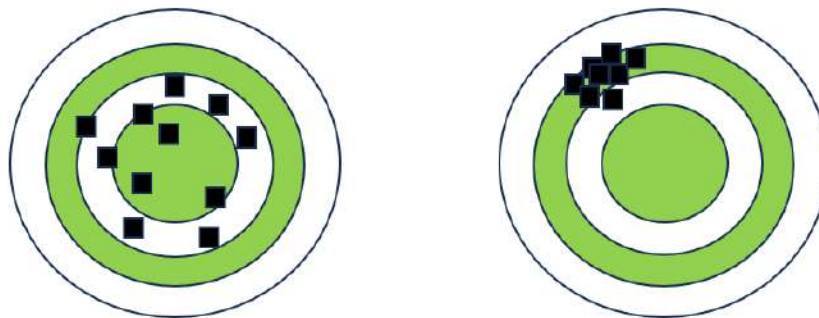
Одговарајуће
прилагођавање

Слика 8: Недовољно, превелико и одговарајуће прилагођавање[40]

вољног прилагођавања и на варијансу услед превеликог прилагођавања.

Грешка пристрасности модела квантификује немогућност наученог модела да „ухвати“ све правилности у компоненти сигнала због погрешних претпоставки у коришћеном моделу. Високе вредности грешке пристрасности указују на то да научени модел константно пропушта неке важне правилности у подацима због инхерентних слабости основног метода учења.

Као што је приказано на Слици 9. сваки црни квадрат означава научени модел добијен коришћењем истог метода учења на случајном скупу за тренирање једнаке величине. Висока вредност грешке пристрасности имплицира да научени модел даје лоше поклапање са правилностима у компоненти сигнала које су заиста релевантне за циљ учења.



Висока варијанса

Висока пристрасност

Слика 9: Грешке пристрасности и варијансе[40]

Варијанса је грешка која произилази из осетљивости учења на мале флукуације у

мерила која процењују ефикасност модела, кључно усмеравајући његов развој. Избор одговарајуће метрике није само питање мерења, већ фундаментална одлука која утиче на правац и фокус тренирања и валидације модела.

Метрике служе за неколико кључних функција у процесу машинског учења. Обезбеђују средства за валидацију и упоређивање ефикасности различитих модела или алгоритама, помажући у избору најпогоднијег приступа за дати проблем. Кроз квантификацију перформанси, метрике идентификују области које захтевају побољшање, подстичући итеративно усавршавање модела. Ово је од суштинског значаја за побољшање предиктивне тачности и поузданости модела током времена.

Различите метрике бацају светло на различите аспекте понашања модела и карактеристика перформанси. Ово разумевање је кључно у тумачењу могућности и ограничења модела, осигуравајући да су његове перформансе усклађене са специфичним потребама дефинисане примене.

Штавише, метрике играју главну улогу у балансирању компромиса који се често јављају у развоју модела. На пример, у сценаријима где су прецизност и опозив у супротности, одговарајућа метрика помаже при балансирању компромисима како би се постигла оптимална равнотежа. Овај чин балансирања је од виталног значаја за прилагођавање перформанси модела нијансираним захтевима различитих примена.

На крају, важност метрика превазилази техничко подручје развоја модела. Оне су инструменталне у повезивању перформанси модела са дефинисаним циљевима и резултатима у стварном свету, обезбеђујући да су увиди добијени из модела машинског учења релевантни, делотворни и да додају опипљиву вредност њиховим намераваним применама. Стога, метрике нису само алати за процену, оне су саставни део успешне примене машинског учења у различитим доменима и изазовима.

Најчешће метрике које се употребљавају у проблемима типа класификације су:

$$\text{Тачност} = \frac{\text{Број тачних предвиђања}}{\text{Укупан број предвиђања}}$$

Тачност (енгл. accuracy) мери пропорцију тачних предвиђања у укупном броју испитаних случајева. Иако је једноставна и интуитивна, може наводити на погрешне закључке у неуравнотеженим скуповима података.

$$\text{Прецизност} = \frac{\text{Стварни позитивни}}{\text{Стварни позитивни} + \text{Лажни позитивни}}$$

Прецизност (енгл. Precision) процењује удео стварних позитивних резултата међу позитивним предвиђањима. Стварни позитивни су случајеви у којима модел тачно предвиђа позитивну класу. На пример у сценарију медицинске дијагнозе, ако модел тачно идентификује 90 од 100 пацијената са болешћу као да имају болест, тих 90 су стварни позитивни. Лажни позитивни су случајеви у којима модел погрешно предвиђа позитивну класу када је она заправо негативна. На пример у истом сценарију медицинске дијагнозе, ако модел предвиђа да 20 од 100 здравих пацијената има болест када је немају, тих 20 су лажно позитивни. То је кључно када је цена лажних позитивних резултата висока, али не узима у обзир лажне негативне.

$$\text{Опозив} = \frac{\text{Стварни позитивни}}{\text{Стварни позитивни} + \text{Лажни негативни}}$$

Опозив (енгл. Recall) мери пропорцију од стварно позитивних који су тачно идентификовани. Што је од виталног значаја када су погрешно идентификовани позитивни скупи, али не узима у обзир лажне позитивне. Код опозива рачунају се лажни негативни случајеви у којима модел погрешно предвиђа негативну класу када је она заправо позитивна. На пример у истом сценарију медицинске дијагнозе, ако модел не успе да идентификује 10 од 100 пацијената са болешћу, тих 10 су лажно негативни.

$$\text{F1 резултат} = 2 \times \frac{\text{Прецизност} \times \text{Опозив}}{\text{Прецизност} + \text{Опозив}}$$

F1 резултат (енгл. F1 score) комбинује прецизност и опозив, балансирајући лажно позитивне и негативне. То је корисно за неуједначене дистрибуције класа, али може бити мање разумљиво за интерпретацију од појединачних вредности.

Област испод ROC криве (енгл. Area Under the ROC Curve -AUC-ROC) је метрика за мерење перформанси која узима у обзир све могуће прагове класификације. То је корисно за процену укупног учинка на различитим праговима, али може бити мање информативно у присуству високо неуравнотежених класа.

Најчешће метрике које се употребљавају у проблемима типа регресије су:

$$\text{Средња апсолутна грешка} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Средња апсолутна грешка (енгл. Mean Absolute Error - MAE) даје просек апсолутних грешака, то јест разлику између стварних вредности (y_i) и предвиђања модела (\hat{y}_i). Лака је за разумевање и тумачење. Све грешке третира подједнако, не наглашавајући велике грешке.

$$\text{Средња квадратна грешка} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Средња квадратна грешка (енгл. Mean Squared Error - MSE) строжије кажњава веће грешке, чинећи метрику осетљивијом на одступања. Ово може бити и предност и недостатак.

$$\text{Корен средње квадратне грешке} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Корен средње квадратне грешке (енгл. Root Mean Squared Error - RMSE) мери стандардну девијацију резидуала. Наведена метрика је осетљива на одступања, али изражава грешку у јединицама променљиве одговора.

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

Коефицијент детерминације (енгл. Coefficient of Determination - R^2) показује колико добро независне променљиве објашњавају варијансу у зависној променљивој. Корисно за процену прилагођавања, али не указује на адекватност модела или неопходност сваког предиктора.

$$\text{Средња апсолутна процентуална грешка} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

Средња апсолутна процентуална грешка (енгл. Mean Absolute Percentage Error - MAPE) изражава грешку у процентима, нудећи интерпретацију независну од скале. Може да доведе у заблуду са нултим или скоро нултим вредностима у стварним подацима.

3.2 Општи принципи у машинском учењу

У овом делу биће покривено неколико општих принципа машинског учења, пружајући важне увиде неопходне за разумевање спроведеног истраживања.

3.2.1 Окамова оштрица

Окамова оштрица представља општи принцип решавања проблема у филозофији и науци. Понекад се парафразира изјавом „најједноставније решење је највероватније право“. У контексту машинског учења, Окамова оштрица значи преференцију за једноставност приликом избора модела. У случају да се примети да два различита модела дају сличне перформансе на истим подацима за тренирање, требало би да преферирамо једноставнији модел него сложенији.

Штавише, принцип минималне дужине описа представља формализацију принципа Окамове оштрице у машинском учењу, који наводи да све методе машинског учења имају за циљ да пронађу правилности у подацима, а најбољи модел за описивање правилности у подацима је такође онај који највише може да сажме податке [50].

3.2.2 Теорема непостојања бесплатног оброка

У контексту машинског учења, теорему непостојања бесплатног оброка први објашњава Волперт у раду из 1996. године [51]. Теорема каже да ће ако имамо скуп података без компонентне шума за било која два алгорита машинског учења А и Б, просечне перформансе А и Б бити исте за све могуће опсервације узорковане на основу униформне дистрибуције вероватноће.

Наведена тврдња се ослања на чињеницу да сваки алгоритам машинског учења има претходне претпоставке о правилностима између обележја података и зависне променљиве за проблем машинског учења. Ове претпоставке се често називају а priori претпоставкама.

Перформансе алгорита за машинско учење на било ком датом проблему зависе од тога колико су претпоставке алгорита усклађене са стварношћу проблема. Алгоритам може да оствари веома добре перформансе за један проблем, али то нам не даје разлога

да верујемо да ће исто тако добро радити и на другом проблему где исте претпоставке можда неће функционисати.

3.2.3 Клетва димензионалности

Клетва димензионалности се односи на проблем који се јавља при раду са високодимензионалним подацима. Димензија скупа података одговара броју обележја који постоје у скупу података. Алгоритми машинског учења су добри у анализи података са много димензија. Људска интуиција тродимензионалног физичког света често за-каже у високим димензијама[52]. Људи нису добри у проналажењу образаца који се могу раширити на много димензија, посебно ако су те димензије међусобно повезане на контраинтуитиван начин.

С друге стране, како додајемо више димензија, ми такође повећавамо рачунарску снагу која нам је потребна како бисмо анализирали високодимензионалне податке. Уз то неопходно је да повећамо количину података за тренирање како бисмо остварили задовољавајуће перформансе модела машинског учења. Најгори сценарио који може настати путем клетве димензионалности је у случају када су подаци униформно дистрибуирани у високодимензионалним просторима. Већина проблема из домена машинског учења који долазе из стварног света укључује високодимензионалне податке који нису униформно распоређени по високодимензионалним просторима. Ово запажање се често назива благословом неуниформности [52].

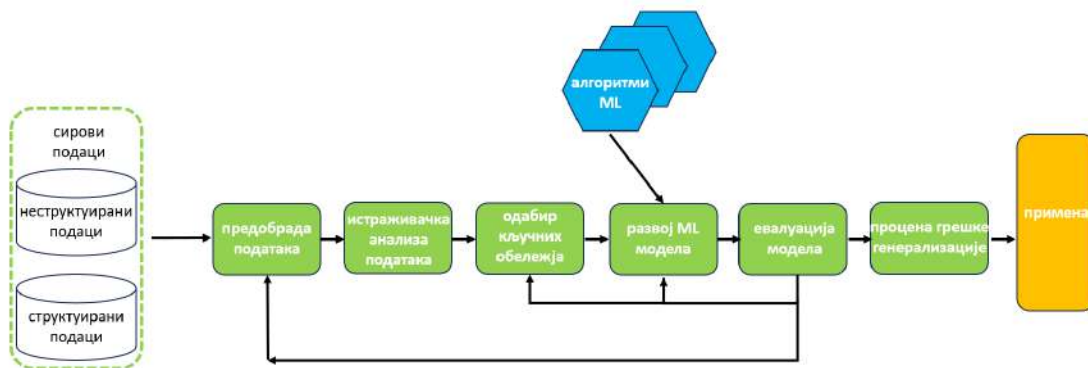
Благослов неуниформности нам у суштини омогућава да будемо у могућности да ефикасно тренирамо алгоритме машинског учења за решавање проблема у којима се јављају високодимензионални подаци користећи разумну количину података за тренирање и рачунарских ресурса. Неуниформна дистрибуција података сугерише да све димензије података нису независне, већ су у високој корелацији на такав начин да су многе димензије сувишне. Другим речима, многе димензије се могу одбацити без губитка много информација о дистрибуцији самих података.

3.3 Ток машинског учења

Ток машинског учења има за циљ да помогне у аутоматизацији примене операција машинског учења, како би се из великих скупова података добили корисни увиди (знања). Циљ машинског учења је да се истренира модел машинског учења који ће имати способност доброг решавања дефинисаног проблема, тј. да оствари задовољавајуће перформансе у решавању одређеног проблема.

Сам термин ток машинског учења наводи нас на погрешне закључке у вези природе одвијања операција машинског учења, јер се многи кораци из тока операција машинског учења могу понављати итеративно како би се унапредила и побољшала тачност модела машинског учења. Циклична архитектура тока машинског учења је приказана на Слици 11.

На почетку се, пре извођења било какве анализе, извршава припремање (предобрада) улазних (или прикупљених) података који се налазе у „сировом“ облику. Овај корак укључује активности као што су чишћење података, импутацију података (процес



Слика 11: Ток машинског учења

измене недостајућих података са заменским подацима), скалирање / стандардизацију / нормализацију података, трансформисање категоријских варијабли у нумеричке, узорковање података итд.. Наведене активности се предузимају како би се решила потенцијална питања као што су шум у подацима, постојање ексцентричних вредности, неуравнотежени (или пристрасни) скупови података, а све у циљу што боље припреме „сирових“ података пред остале кораке у току машинског учења.

Затим следи корак истраживачке анализе података. У овом кораку подаци се анализирају да би се боље разумеле њихове карактеристике нпр. да ли је у питању нормална или искривљена (енгл. *skewed*) дистрибуција. Искривљеност података утиче на перформансе статистичког модела, посебно у случају модела заснованих на регресији. Да би се спречила корупција резултата услед искривљености, уобичајена пракса је да се примени трансформација на цео скуп вредности и да се користе трансформисани подаци за статистички модел.

Такође, у овом кораку се извршава и откривање корелација између обележја скупа података како би се идентификовале независне варијабле које се после користе у процесу тренирања модела. На пример, ако је обележје a_1 у великој корелацији са обележјем a_2 , онда само једно од тих обележја треба узети у обзир за тренирање модела. У скуповима података где постоји линеарна веза између обележја и зависних променљивих, важно је схватити односе између њих као што су позитивна корелација (када се одређено обележје повећава/смањује тако се зависна варијабла повећава/смањује) и негативна корелација (када се одређено обележје повећава/смањује тако се зависна променљива смањује/повећава) или да уопште нема корелације између обележја и зависних променљивих. Технике визуелизације као што је матрица корелације могу показати односе између различитих парова комбинација обележја и зависних променљивих из скупа података.

Наредни корак је одабир кључних обележја података за тренирање модела машинског учења. Предности које произилазе из доброг одабира кључних обележја података су смањивање проблема претераног прилагођавања, побољшање тачности модела и скраћивање времена трајања тренирања модела. Одабир обележја се може извршити на различите начине. Искоришћење препознатих односа између променљивих може послужити као водила приликом одабира обележја. Међутим, када број обележја расте,

технике одабира засноване на подацима проналазе своју примену. Неке од најчешћих техника за смањење димензионалности укључују Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), Independent Component Analysis (ICA), и алгоритме груписања (нпр. Gaussian mixture model). Реални скупови података садрже велик број обележја, међу којима само један подскуп помаже у решавању дефинисаног проблема. На пример, за детекцију саобраћајне траке у примени за аутономну вожњу, важна обележја могу укључити ивицу, градијент и интензитет [53], јер се ослањају на различит интензитет између површине пута и ознака саобраћајне траке који омогућава детекцију.

Даље следи корак развоја модела машинског учења и одабир одговарајућег алгорита машинског учења за решавање дефинисаног проблема. С'обзиром да је развијено много различитих алгоритама машинског учења за решавање различитих проблемских задатака суочавамо се са проблемом избора алгоритама машинског учења. Теорема непостојања бесплатног оброка нам говори да не постоји један универзални алгоритам машинског учења који ће најбоље решавати све могуће проблеме из машинског учења. Овај концепт можемо повезати са популарном изреком конструисаном од стране Абрахама Маслоуа: „Верујем да је примамљиво, да ако је једини алат који имамо чекић, да све третирамо као да је ексер“ [54]. То у машинском учењу није случај. У пракси је стога од суштинског значаја да упоредимо бар неколико различитих алгоритама машинског учења како бисмо тренирали и изабрали модел са најбољим перформансама за дефинисани проблем.

Након развоја модела машинског учења потребно је да упоредимо различите моделе, односно да извршимо евалуацију модела. Евалуација модела је неопходна за креирање ефективних модела који не само да добро раде на подацима који су коришћени за тренирање модела, већ и на новим подацима које модел „види“ први пут. Евалуација модела машинског учења представља суштински део сваког тока машинског учења. Када припремимо податке и омогућимо нашем моделу да путем алгоритама машинског учења учи из података за тренирање, следећи корак је да измеримо перформансе модела. Неопходно је да се пронађе метрика која не само да нам може рећи колико су тачна предвиђања модела, већ нам такође омогућава да упоредимо перформансе бројних модела, како бисмо могли да изаберемо онај који је најприкладнији за дефинисани проблем.

Потребно је напоменути да ће метрике евалуације бити различите за предиктивно моделирање типа класификације и типа регресије, пошто су излазне вредности код првог типа категоричке, док су излазне вредности код другог типа континуалне.

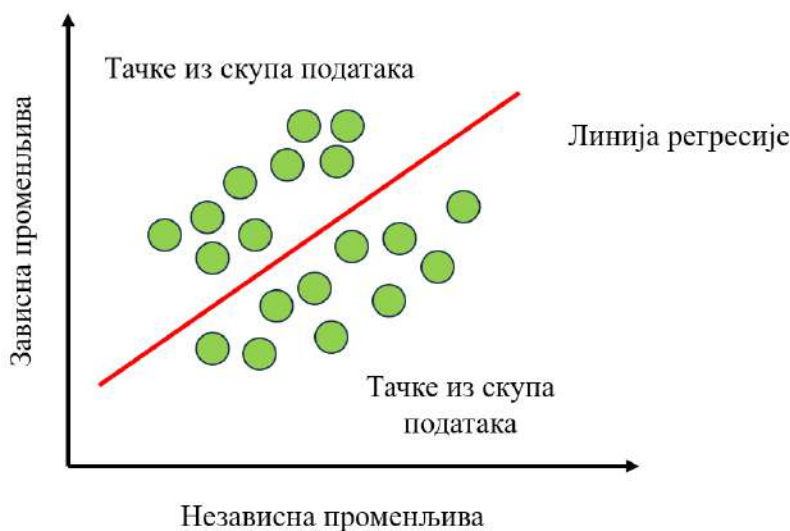
Завршни корак представља процена грешке генерализације, односно какве перформансе модел остварује на подацима које модел „види“ први пут. У случају да смо задовољни са перформансама модела сада можемо користити овај модел за предвиђање нових, будућих података, тј. да пређемо у домен примене модела машинског учења.

3.4 Одабрани алгоритми машинског учења

3.4.1 Линеарна регресија

Линеарна регресија је врста алгоритма машинског учења из домена учења са надзором који се користи за предвиђање вредности континуалне зависне променљиве на основу вредности једне или више независних променљивих. У линеарној регресији, циљ је пронаћи линију или раван која најбоље одговара подацима и минимизира разлике између предвиђених и стварних вредности зависне променљиве.

Да би се пронашло најбоље поклапање, линеарна регресија користи метод најмањих квадрата, који укључује проналажење коефицијената линеарне једначине који минимизирају збир квадрата грешака између предвиђених и стварних вредности зависне променљиве. Ова линија или раван се сматра најбољим поклапањем јер је то линија или раван која је најближа свим тачкама података из скупа за тренирање. Када су коефицијенти линеарне једначине процењени методом најмањих квадрата, једначина се може користити за предвиђање нових, тј. података које модел види „први пут“. Ова предвиђања се затим могу упоредити са стварним вредностима зависне варијабле да би се проценио учинак модела линеарне регресије. Приказ линеарне регресије - Слика 12.



Слика 12: Линеарна регресија

У случају једноставне линеарне регресије алгоритам моделира однос између две променљиве: једне независне (X) и једне зависне (Y). Однос се моделује путем линеарне једначине:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- β_0 (одсечак): Код једноставне линеарне регресије има вредност 0 и назива се још регресија кроз координатни почетак. То је тачка у којој се линија регресије укршта са Y осом.

- β_1 (нагиб): Овај коефицијент представља промену у Y за једну јединичну промену у X . Он одређује нагиб линије регресије.
- ε (термин грешке): Ово представља разлику између посматраних вредности и вредности предвиђених моделом. То је кључни део модела, узимајући у обзир варијабилност која није објашњена линеарним односом.

Када се ради са више независних променљивих, једначина се проширује на:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Овде сваки X_i представља другу независну променљиву, а сваки β_i је коефицијент који квантификује однос између те специфичне независне променљиве и зависне променљиве.

Коефицијенти (β) се обично процењују коришћењем методе најмањих квадрата. Овај метод минимизира збир квадрата резидуала (разлике између посматраних вредности и вредности предвиђених моделом). За једноставну линеарну регресију, формуле за коефицијенте су:

- Нагиб (β_1):

$$\beta_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Ова формула израчунава нагиб као ратио коваријансе X и Y према варијанси X .

- Одсечак (β_0):

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

Ово се израчунава узимањем средње вредности Y и одузимањем производа нагиба и средње вредности X .

Кључне претпоставке за употребу линеарне регресије су:

- Линеарност: однос између X и Y мора бити линеаран.
- Хомоскедастичност: грешке (резидуали) треба да имају константну варијацију у различитим вредностима X .
- Независност: операције треба да буду независне једна од друге.
- Нормална расподела грешака: грешке би у идеалном случају требале да прате нормалну расподелу.

Једна од предности линеарне регресије је то што пружа једноставну и разумљиву математичку формулу за прављење предвиђања. Пошто је линеарна регресија заснована на добро утврђеним статистичким принципима, својства алгоритама линеарне регресије су добро позната и алгоритми се могу брзо тренирати. Поред тога, треба напоменути да се линеарна регресија може применити само на податке где су зависне и независне променљиве квантитативне, што значи да су представљене нумеричким вредностима.

Линеарна регресија је једноставан и широко коришћен алгоритам машинског учења који се може применити на много различитих поља. Неки примери примене линеарне регресије укључују предвиђање цена акција на берзи, цена некретнина и других континуалних вредности. Линеарна регресија се такође користи у многим различитим областима, укључујући биологију, испитивање понашања, животну средину, друштвене науке и пословање.

3.4.2 Логистичка регресија

Логистичка регресија је врста алгоритма машинског учења из домена учења са надзором који се користи за бинарну класификацију.

За разлику од линеарне регресије, која се користи за предвиђање континуалних вредности, логистичка регресија се користи за предвиђање бинарног исхода, као што је да ли је електронска пошта нежељена пошта или није нежељена пошта, или у медицини да ли је уочени тумор малигни или бенигни, или да ли је мачка на слици или није.

Логистичка регресија функционише тако што се логистичка крива (такозвана S крива) прилагођава подацима, што је нелинеарна функција која се може користити за моделовање вероватноће бинарног исхода. Логистичка крива се одређује кроз процес тренирања модела, где алгоритам прилагођава коефицијенте криве да би минимизирао грешку између предвиђених и стварних вредности зависне променљиве. Када је логистичка крива одређена, може се користити за предвиђање нових, тј. података које модел види „први пут“.

Централна компонента логистичке регресије је логистичка функција, такође позната као сигмоидна функција. Ова функција мапира сваки реални број у вредност између 0 и 1, што га чини идеалним за тумачење у виду вероватноће. Функција је дефинисана као:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

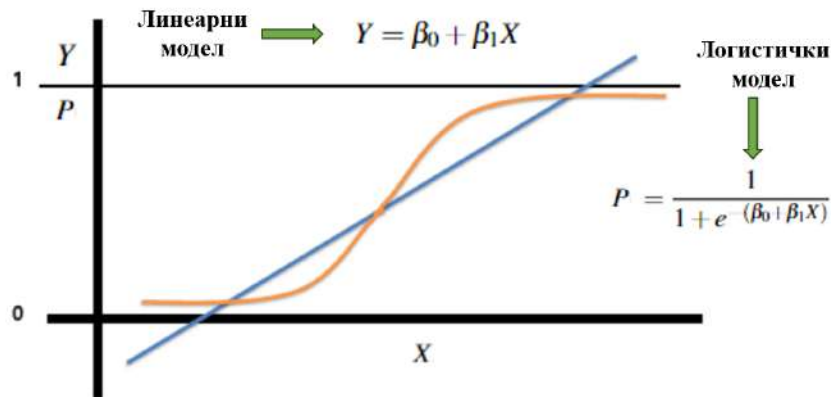
Где:

- $P(Y = 1)$: представља вероватноћу да је зависна променљива Y у класи 1, узимајући у обзир независну променљиву X .
- e : Ојлеров број, основа природних алгоритама.
- β_0, β_1 : ово су коефицијенти слични онима код линеарне регресије. β_0 је одсечак, а β_1 је коефицијент за независну променљиву X .
- израз $e^{-(\beta_0 + \beta_1 X)}$ је кључан. Када је ова вредност велика, $P(Y = 1)$ се приближава 0, а кад је мала, $P(Y = 1)$ се приближава 1.

Процена максималне вероватноће (енгл. Maximum Likelihood Estimation - MLE) се користи за процену коефицијената. Она бира вредности β_0 и β_1 које максимизирају вероватноћу опсервирања података узорка. Наведени процес укључује проналажење скупа параметара који чине опсервиране исходе највероватнијим.

Кључне претпоставке за употребу логистичке регресије су:

- Бинарни исход: зависна променљива треба да буде у бинарном облику (0/1, тачно/нетачно).
- Без мултиколинearности: независне варијабле не би требало да буду у великој корелацији једна са другом.
- Линеарност у log вероватноћама: требало би да постоји линеарна веза између независних променљивих и log вероватноћа зависне променљиве.



Слика 13: Приказ логистичке и линеарне регресије [55]

Једна од предности логистичке регресије је у томе што пружа једноставну и разумљиву математичку формулу за прављење предикција. Пошто је логистичка регресија заснована на добро утврђеним статистичким принципима, својства алгоритама логистичке регресије су добро схваћена и алгоритми се могу брзо тренирати. Додатно, логистичка регресија се може применити само на податке где је зависна променљива бинарна, што значи да може имати само две могуће вредности (нпр. 0 или 1, да или не, итд.).

Алгоритам логистичка регресија представља драгоцену средство за решавање проблема бинарне класификације и често се користи као полазна тачка за сложеније алгоритме машинског учења, као што су неуронске мреже.

3.4.3 Метода К најближих суседа

Метода К најближих суседа (енгл. K Nearest Neighbors - KNN) је врста алгоритама машинског учења из домена учења са надзором који се користи како за задатке класификације тако и за задатке регресије. У случају класификације, циљ је да се предвиди ознака класе нове опсервације на основу њене сличности са примерима из скупа података за тренирање. У случају регресије, циљ је да се предвиди нумеричка вредност за нове опсервације на основу њихове сличности са примерима из скупа података за тренирање.

KNN алгоритам је непараметарски алгоритам за „лењо учење“, што значи да не захтева тренирање на скупу података за тренирање, и не прави никакве претпоставке о основној дистрибуцији података. Уместо тога, структура модела се одређује из скупа података, а сви подаци о тренирању се користе у фази тестирања. KNN је заснован на меморији и није потребна да се ради изградња модела.

KNN алгоритам се заснива на идеји да слични примери вероватно припадају истој класи. Да би направио предвиђање за нове опсервације, алгоритам прво израчунава растојање између нове опсервације и свих примера у скупу података за тренирање користећи метрику удаљености, као што је на пример Еуклидска удаљеност. Затим се бира K примера у скупу података за тренирање који су најближи новој опсервацији, а класа већине међу ових K примера се враћа као предвиђање за нову опсервацију.

Процес KNN алгоритма се одвија на следећи начин:

- Нормализација: први корак у KNN алгоритму је нормализација карактеристика података. Ово је важно јер се растојање између две опсервације израчунава на основу карактеристика података, а ако карактеристике нису у истој скали, прорачун удаљености можда неће имати смисла. Нормализација обично укључује скалирање карактеристика тако да имају нулту средњу вредност и јединичну варијансу, што осигурава да свака карактеристика података подједнако доприноси израчунавању удаљености.
- Иницијализација K : следећи корак је иницијализација вредности хиперпараметра K , који одређује број најближих суседа који ће се користити за предвиђање. Вредност K се може одредити кроз унакрсну валидацију или коришћењем хеуристичког приступа. Уобичајени хеуристички приступ за избор ' K ' је $K = \sqrt{n}$, где је n број тачака података у скупу за тренирање или избор вредности која је мали непаран број.
- Израчунавање растојања: за сваку нову опсервацију, растојање између опсервације и свих примера из скупа података за тренирање се израчунава коришћењем изабране метрике удаљености. Кључна одлика KNN-а лежи у израчунавању удаљености између тачака података. Најчешће коришћене метрике удаљености су:

- Еуклидска удаљеност за тачке x и y , која се израчунава као:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- Менхетн удаљеност која се израчунава као:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

- Минковски удаљеност генерализована метрика која се израчунава као:

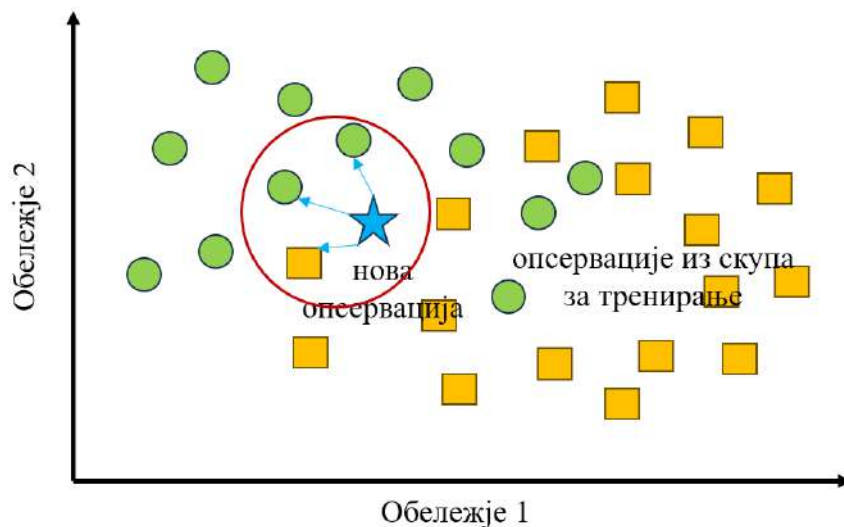
$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

- Сортирање удаљености: растојања се затим сортирају узлазним редоследом, тако да су најмања растојања на почетку листе.
- Одабир првих K примера: Првих K примера на сортираној листи се одабира.
- Преузимање ознака: ознаке K одабраних примера се преузимају из скупа података за тренирање.
- Предвиђање: у случају да се KNN алгоритам користи за класификацију, најбројнија K ознака се враћа као предвиђање за нову опсервацију. Ако се KNN алгоритам користи за регресију, средња вредност или медијана K ознака се враћају као предвиђање.

KNN алгоритам има неколико предности, укључујући његову једноставност, ефикасност и флексибилност. Може да обрађује и континуиране и категоричке променљиве и може да ради са задацима класификације у више класа. Такође није осетљив на специфичне карактеристике података за тренирање, што га чини робусним алгоритмом.

Међутим, KNN алгоритам такође има нека ограничења. Потребна је велика количина меморије за складиштење скупа података за тренирање и може бити рачунски захтевно израчунавање удаљености између нове опсервације и свих примера у скупу података за тренирање. Поред тога, избор вредности K и метрике удаљености може утицати на перформансе алгоритма, па ови хиперпараметри морају бити пажљиво одабрани.

Приказ KNN алгоритма је дат на Слици 14.



Слика 14: K најближих суседа

Кључне претпоставке за употребу KNN алгоритма су:

- Нормализација: пошто се KNN ослања на мерење удаљености, нормализација или стандардизација обележја је од кључног значаја како би се осигурало да свако обележје подједнако доприноси прорачунима удаљености.

- Клетва димензионалности: са повећањем димензија, растојање између тачака постаје мање значајно, што може смањити способност алгоритма да прави тачна предвиђања. Одабир обележја или технике смањења димензионалности могу бити од помоћи у таквим случајевима.

KNN алгоритам је једноставан и ефикасан начин за обављање задатака класификације и регресије. Често се користи као мерило за сложеније алгоритме због своје једноставности и флексибилности.

KNN алгоритам може бити користан у различитим применама, посебно код система препорука. У систему препорука, циљ је препоручити артикле (као што су филмови, музика или производи) корисницима на основу њихових интересовања и преференција. KNN алгоритам се може користити за идентификацију K најближих суседа корисника на основу њихових прошлих интеракција са системом, а затим препоручити ставке које су сличне ставкама које су сличне групе корисника одабрале. Овај приступ може дати персонализоване препоруке у поређењу са другим методама, које не узимају у обзир сличност међу корисницима.

3.4.4 Наивни Бајесов класификатор

Наивни Бајесов класификатор је пробабилистички алгоритам машинског учења који користи Бајесову теорему за предвиђање. Назива се „наивним“ јер чини поједностављујућу претпоставку да су сва улазна обележја у скупу података независна једна од других.

Бајесова теорема каже да је вероватноћа да се догоди неки догађај (A) узимајући у обзир то да се догодио други догађај (B) једнака вероватноћи да се догоди догађај B за дато A помноженој са вероватноћом да се A догоди, подељеној са вероватноћом да се догоди B.

Ово се може математички написати као:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

У случају наивног Бајесовог класификатора, догађаји у питању су класе проблема класификације и посматране карактеристике тачке из улазних података. Класификатор израчунава вероватноћу сваке класе за нову тачку података множењем појединачних вероватноћа сваке променљиве обележја која припада тој класи. Ово је омогућено претпоставком да су поменљиве обележја независне једна од друге. Да би класификовао нову тачку из скупа података, наивни Бајесов класификатор прво израчунава вероватноћу сваке класе с обзиром на посматране карактеристике података користећи Бајесову теорему. Затим бира класу са највећом вероватноћом као предвиђену класу за нову тачку података. Ово је познато као правило одлуке *maximum a posteriori* (MAP).

Кључне претпоставке за употребу наивног Бајесовог класификатора су:

- Независност обележја: највећа претпоставка наивног Бајесовог класификатора је да су сва обележја независна једна од других с обзиром на класу. Ово је поједностављење које можда не важи у стварним сценаријима, али изненађујуће, наивни Бајесов класификатор често ради добро упркос томе.

- Избор модела: у зависности од типа података, користе се различити наивни Бајесови модели:
 - Гаусов: претпоставља нормално распоређена обележја. Што је погодно за континуиране податке.
 - Мултиноминални: користан за дискретне податке. Уобичајено у класификацији текста где број речи или њихова фреквентност представљају обележја.
 - Бернулијев: најбољи за бинарне векторе обележја.
- Претходно знање: претходне вероватноће класа могу значајно да утичу на перформансе наивног Бајесовог класификатора, посебно у скуповима података са неуравнотеженим класама.
- Технике изглађивања: у пракси, технике као што је Лапласово изглађивање се користе за решавање проблема нулте вероватноће у случају да се дата комбинација класа и обележја нису опсервиране у скупу за тренирање.
- Претходна обрада података: ефикасна претходна обрада као што је нормализација (за Гаусов НБ модел) или токенизација (за текстуалне податке) може значајно побољшати перформансе наивног бајесовог алгоритма.

Наивни Бајесов класификатор се обично користи за задатке као што су анализа сентимента, класификација текста, филтрирање нежељене електронске поште и за системе препорука јер је једноставан и ефикасан. Међутим, његова претпоставка о независности међу предикторским варијаблама може ограничити његов учинак у неким случајевима. Када предиктори зависе једни од других, наивни Бајесов класификатор можда неће моћи тачно да ухвати однос између варијабли, што може смањити његову тачност.

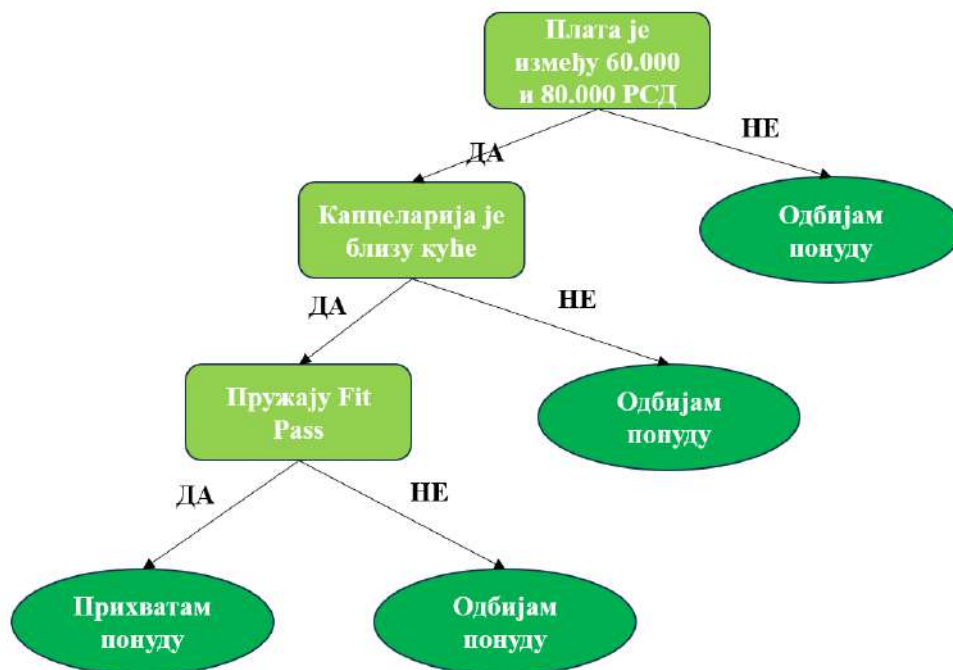
3.4.5 Стабла одлучивања

Стабла одлучивања су тип алгоритма машинског учења који се може користити и за задатке класификације и за задатке регресије. Они функционишу тако што деле податке на све мање и мање подскупе на основу вредности променљивих обележја, са циљем стварања скупа „лисних“ чворова на дну стабла од којих сваки садржи тачке података са истом предвиђеном класом.

На Слици 15. приказан је пример стабла одлучивања на примеру одлуке да ли да се прихвати понуда за посао.

Можемо да користимо стабла одлучивања како бисмо одговорили на секвенцијална питања, која нас воде одређеном рутом низ стабло на основу одговора. Услов „ако ово, онда оно“ је понашање модела које на крају даје резултат.

Изградња стабла одлучивања захтева алгоритме који могу да одреде оптималан избор у сваком чвору. Један од алгоритама који се користи је Хантов алгоритам [56] који је добро познат. Наведени алгоритам доноси оптималну одлуку на сваком кораку, притом не узимајући у обзир укупан оптимум. То значи да је на сваком кораку алгоритам бирао најбољи одговор.



Слика 15: Стабло одлучивања

Мада, избор најбољег резултата у датом кораку не гарантује да ћете донети оптималну одлуку од почетка стабла до коначног чвора стабла, тзв. „лисног“ чвора [57]. Оптимизација може да се ради и помоћу метрика као што су пораст информација или нечистоћа. Алгоритам затим дели податке на основу изабране променљиве обележја и понавља овај процес на сваком следећем чвору у стаблу. Процес се наставља све док се не постигне жељени ниво хомогености или док алгоритам не достигне унапред дефинисане критеријуме заустављања, као што је максимална дубина стабла.

Када је стабло одлучивања изграђено, може се користити за предвиђање нових тачака података. Да би класификовао нову тачку података, алгоритам једноставно прати одлуке донете у сваком чвору у стаблу све док не дође до „лисног“ чвора, у ком тренутку додељује предвиђену класу тачке података да буде иста као класа тачака података у том „лисном“ чвору.

Процес формирања стабла одлучивања се одвија на следећи начин:

- Подела чворова: Процес почиње у основном чвору и укључује поделу података на основу одређених услова. Подела се врши према обележјима која најбоље раздвајају податке у различите класе или исходе, на основу одређених критеријума као што су Гинијева нечистоћа, информацијска добит или смањење варијансе. Да би се изградило стабло одлучивања, алгоритам треба да буде у стању да процени квалитет подела у сваком кораку. Избор метрике за оптимизацију зависиће од специфичног задатка и врсте података са којима се ради.

– Гинијева нечистоћа: користи се у класификацији и мери учесталост у којој

би случајно одабрани елемент био погрешно лабелиран. Рачуна се као:

$$Gini = 1 - \sum (p_i)^2$$

где p_i представља вероватноћу да се објекат сврста у одређену класу.

- Информацијска добит: базирана на основу концепта ентропије из теорије информација. Стабло одлучивања ће изабрати поделу која максимизира информацијску добит:

$$ID(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j)$$

где I представља функцију ентропије, D_p и D_j су скупови података родитељског чвора j^{th} је скуп података дететовог чвора, и N_p и N_j су бројеви елемената у родитељском и дететовом скупу података.

- Смањење варијансе: користи се у регресији, овај метод бира поделе које резултирају највећим смањењем варијансе.
- Прављење стабла одлучивања: алгоритам стабла одлучивања ће затим изградити стабло правећи узастопне поделе на улазним варијаблама. У сваком кораку, алгоритам ће изабрати поделу која резултира најнижом нечистоћом, у складу са изабраном метриком. Процес се наставља све док стабло одлучивања не израсте у потпуности, или док се не постигне неки други критеријум заустављања.
- Предвиђање: када је стабло одлучивања изграђено, може се користити за предвиђање нових тачака података. За нову тачку података, алгоритам ће пратити путању низ стабло која одговара вредностима улазних променљивих. Предвиђена вредност циљане променљиве биће вредност која је повезана са „лисним“ чвором на крају путање. У класификацији, листни чвор представља одлуку о класи. У регресији, представља предвиђену вредност, обично средњу вредност циљних вредности у том листу.

Кључне претпоставке за употребу стабала одлучивања су:

- Претерано прилагођавање: један од највећих изазова са стаблима одлучивања је њихова склоност претераном прилагођавању, посебно код сложених стабала. Различите технике резивања се користе да би се ово ублажило.
- Избор обележја: ефикасност стабла одлучивања у великој мери зависи од избора обележја које се користе за поделу.
- Рад са недостајућим вредностима: стаблима одлучивања су потребни механизми за руковање недостајућим вредностима у тачкама података како би се оствариле робусне перформансе.

Неке од главних области примене стабала одлучивања укључују класификацију медицинских података, процену кредитног ризика, откривање превара, сегментацију купаца и предиктивно одржавање.

Стабла одлучивања су моћан и разумљив алгоритам машинског учења који се може користити за широк спектар задатака. Међутим, као што је већ наведено, они могу бити склони претераном прилагођавању, посебно када је дрвеће дубоко и сложено. Да би се решио овај проблем, могу се користити методе ансамбла које комбинују предвиђања вишеструких стабала одлучивања да би смањили претерано прилагођавање и побољшали перформансе модела.

3.4.6 Насумична шума

Насумична шума (енгл. Random Forest) представља ефикасну технику ансамбл учења у области машинског учења, посебно познату по својој робустности и тачности. Ова метода користи стабла одлучивања као своје основне грађевинске блокове, користећи њихову једноставност и интерпретативност, док истовремено превазилази њихова ограничења кроз ансамбл учење.

У срцу алгоритма насумична шума налази се примена бутстрап методе (енгл. bootstrap), где се свако стабло одлучивања у шуми тренира на бутстрап узорку података. Бутстрап узорци су јединствени по томе што се извлаче са заменом из тренинг скупа података, што значи да се иста тачка података може појавити више пута у скупу података за тренирање једног стабла. Овај приступ уноси разноврсност међу појединачна стабла, што је кључно за укупну ефикасност модела.

Математички, ако је скуп за тренирање D и састоји се од N опсервација, сваки бутстрап узорак D_i (за $i^{\text{о}}$ стабло) је насумично извучен из скупа D са заменом и величином N .

Разноликост међу основним моделима, који су у суштини појединачна стабла одлучивања, кључни је аспект алгоритма насумичне шуме. Ако су стабла превише слична или су у великој корелацији, ансамбл учење не побољшава значајно предиктивне перформансе изнад онога што једно дрво самостално може да постигне. Насумичне шуме решавају овај проблем кроз узорковање и редова и колона. Док се узорковање редова постиже процесом паковања (енгл. bagging), узорковање колоне уводи додатни слој стохастичности. Овде се бира насумични подскуп обележја за свако стабло одлучивања, обезбеђујући да различита стабла разматрају различите скупове обележја за раздвајање.

Типично, ако има M обележја, \sqrt{M} обележја се користи за класификацију и $\frac{M}{3}$ за регресију у свакој подели.

Процес изградње стабала такође доприноси разноликости модела. У ономе што је познато као субоптимално раздвајање, алгоритам бира најбољу поделу из насумичног подскупа обележја у сваком чвору, уместо да узима у обзир сва доступна обележја. Ово не само да доприноси разноликости међу стаблима, већ омогућава и укључивање потенцијално слабијих обележја која би се могла предвидети у детерминистичком приступу.

Једном када се изгради шума стабала одлучивања, насумичне шуме праве предвиђа-

ња агрегирајући резултате појединачних стабала. За задатке класификације, најчешћи облик агрегације је већинско гласање, где се коначно предвиђање заснива на најчешћем исходу међу свим стаблима:

$$\text{Финална класа} = \text{mod}\{f_1(x), f_2(x), \dots, f_B(x)\}$$

где $f_i(x)$ представља предвиђање $i^{\text{ог}}$ стабла.

У случају задатака регресије, алгоритам обично даје средњу вредност резултата:

$$\text{Финално предвиђање} = \frac{1}{B} \sum_{i=1}^B f_i(x)$$

где B представља број стабала у шуми и $f_i(x)$ је предвиђање за $i^{\text{то}}$ стабло.

Кључне претпоставке за употребу насумичне шуме су:

- Претерано прилагођавање: насумичне шуме су мање склоне претераном прилагођавању од појединачних стабала одлучивања због процеса усредњавања.
- Број стабала (B): генерално, више стабала повећава прецизност, али и рачунске захтеве.
- Величина подскупа обележја: избор m (број обележја која се разматрају при свакој подели) је кључан за перформансе модела.

Једна од најзначајнијих предности алгоритма насумична шума је његова отпорност на прекомерно прилагођавање, што је уобичајен проблем са појединачним стаблима одлучивања, посебно када се ради са сложеним, високодимензионалним подацима. Они су у стању да рукују великим скуповима података и дају процену важности обележја, која је изведена из учесталости избора обележја за раздвајање у стаблима.

Алгоритми насумичне шуме налазе примену у различитим доменима, од геномских студија у биологији до процене ризика у финансијама и препознавања објеката у рачунарском виду. Њихова свестраност, лакоћа коришћења и способност да обезбеде моћ предвиђања и увид у податке чине их популарним избором у области машинског учења.

3.4.7 Екстремно насумична стабла

Екстремно насумична стабла (енгл. *Extremely Randomized Trees*), која се често називају екстра стаблима (енгл. *Extra Trees*), је метода ансамбл учења која је уско повезана са насумичним шумама, али са посебним приступом конструисању стабала. Познат је по својој ефикасности и делотворности, посебно у сценаријима где су случајност и брзина приоритет.

Алгоритам екстра стабала разликује се од насумичних шума првенствено по начину на који су стабла конструисана. Док насумичне шуме траже оптималну поделу у сваком чвору стабала одлучивања, екстра стабла у овај процес уносе додатни слој случајности. У екстра стаблима, за свако обележје која се разматра, бира се насумична тачка пресека

да би се чвор поделио, уместо да се тражи најоптималнија тачка пресека. То значи да се поделе бирају потпуно насумично, а користе се најбоље поделе међу њима. Овај приступ повећава разноликост између појединачних стабала у ансамблу, што може довести до побољшане робусности модела.

Штавише, за разлику од насумичних шума, које обично користе бутстрап узорак (случајни подскуп са заменом) скупа података за тренирање за прављење сваког стабла, алгоритам екстра стабала користи цео оригинални скуп података. Ово понекад може довести до смањења пристрасности јер свако дрво у ансамблу разматра све узорке за тренирање.

Због овог процеса насумичне поделе, екстра стабла могу бити бржа од насумичних шума, посебно када се ради о великим скуповима података и када се подаци за тренирање одликују високодимензионалним простором обележја. Насумична природа подела такође често чини екстра стабла робуснијим на прекомерно прилагођавање и буку у подацима.

Кључне претпоставке за употребу екстра стабала су:

- Степен насумичности: додатни слој насумичности у екстра стаблима (у поређењу са насумичном шумом) може довести до више разноврсних стабала и потенцијално боље генерализације.
- Рачунски захтеви: због насумичне природе подела, екстра стабла могу бити бржа за тренирање од насумичних шума.

Што се тиче перформанси, алгоритам екстра стабла понекад може надмашити алгоритам насумичне шуме, у зависности од скупа података и проблема. Посебно је користан у ситуацијама када је број обележја велики, а простор за обележја оскудан. Као и насумичне шуме, алгоритам екстра стабла се може користити за задатке класификације и регресије.

Алгоритам екстра стабала је такође познат по својој једноставности и лакоћи имплементације. Његова способност да ефикасно рукује великим скуповима података, заједно са његовом отпорношћу на претерано прилагођавање, чини га вредним алатом за широк спектар практичних примена у областима у распону од биоинформатике до финансијског моделирања.

3.4.8 Стабла квантилне регресије

Стабла квантилне регресије (енгл. Quantile Regression Trees - QRT) су специјализовани облик регресионих стабала који се користе у машинском учењу за сценарије где је циљ предвиђање специфичних квантила дистрибуције варијабле одговора, а не њене средње вредности. Овај приступ је посебно користан у разумевању условне дистрибуције варијабле одговора и погодан је за примене где су квантили информативнији од средње вредности.

Традиционална стабла регресије, попут оних која се користе у насумичним шумама или екстра стаблима, фокусирају се на предвиђање средње вредности или медијане циљне променљиве. Међутим, ово понекад може бити ограничавајуће, посебно када је дистрибуција променљиве искривљена или када су присутне обсервације које се налазе

изван очекиваног опсега вредности (енг. оутлиерс). Стабла квантилне регресије решавају ово ограничење тако што омогућавају предвиђање различитих квантила, пружајући свеобухватнији поглед на потенцијалне исходе и њихове вероватноће.

У стаблима квантилне регресије, процес изградње стабла је сличан оном код стандардних регресионих стабала, али са кључном разликом у критеријуму који се користи за прављење подела. Уместо да минимизира укупну варијансу или средњу квадратну грешку, алгоритам минимизира функцију губитка која је осетљивија на одређени квантил. Ово укључује израчунавање квантила унутар сваког чвора и њихово коришћење за одређивање најбоље поделе.

За дати квантил τ (где су $0 < \tau < 1$), функција губитка за предвиђање \hat{y} и стварна вредност y дефинисане су као:

$$L_{\tau}(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{ако је } y \geq \hat{y} \\ (1 - \tau)(\hat{y} - y) & \text{ако је } y < \hat{y} \end{cases}$$

Наведена функција губитка асиметрично кажњава превелика и премала предвиђања, у зависности од изабраног квантила.

Кључне претпоставке за употребу стабла квантилне регресије су:

- одабир квантила τ : Одабир квантила за предвиђање је битан. Уобичајени избори подразумевају медијану ($\tau = 0.5$), доњи квантил ($\tau = 0.25$), или горњи квантил ($\tau = 0.75$).
- Руковање вредностима изван очекиваног опсега: QRT-ови могу бити робуснији на вредности изван очекиваног опсега у поређењу са стандардним регресионим стаблима, јер се не фокусирају само на предвиђање средње вредности.

Значајна предност QRT-а је његова способност да пружи нијансираније разумевање података. На пример, може да моделира понашање доњег или горњег дела дистрибуције, што је корисно у сценаријима процене ризика, као што су финансије или осигурање, где су екстремне вредности од посебног интереса. Такође може дати увид у то како независне променљиве утичу на различите делове дистрибуције зависне променљиве, а не само на њену централну тенденцију.

Стабла квантилне регресије су такође отпорна на опсервације које се налазе изван очекиваног опсега вредности. Пошто се фокусирају на квантиле, а не на средње вредности, екстремне вредности у опсервацијама из сетова за тренирање имају мањи утицај на модел, што QRT чини робуснијом опцијом у скуповима података где постоји бојазан у вези са опсервацијама које се налазе изван очекиваног опсега вредности.

У пракси, QRT се може имплементирати коришћењем различитих оквира за машинско учење који подржавају квантилну регресију. Флексибилност у избору квантила омогућава аналитичарима и истраживачима да прилагоде модел својим специфичним потребама, било да су заинтересовани за медијане исхода, екстремне вредности или било коју другу тачку у дистрибуцији.

3.4.9 Машине вектора упоришта

Машине вектора упоришта (енгл. Support Vector Machines - SVM) представљају тип алгоритма за учење са надзором који се може користити за задатке класификације или регресије. Основна идеја иза SVM-а је да се пронађе оптимална линија (или хипер-раван) која максимално одваја тачке података једне класе од оних друге класе.

Циљ SVM алгоритма је да пронађе хипер-раван у N димензионалном простору која јасно раздваја тачке података. N се односи на димензије обележја података. Да би се одвојиле две класе тачака података, постоји много могућих хипер-равни које се могу изабрати. Циљ SVM-а је да пронађе раван која има максималну маргину. Другим речима, циљ SVM-а је да пронађе раван која има максимално растојање између тачака података обе класе. Максимизовање удаљености маргине даје одређену резерву сигурности тако да се будуће тачке података могу класификовати са више поузданости. Хипер-равни представљају границе одлучивања које омогућавају класификацију опсервација.

Опсервације које падају са обе стране хипер-равни могу се приписати различитим класама. Такође, димензија хипер-равни зависи од броја обележја. Ако је број улазних обележја једнак два, онда је хипер-раван само права. Ако је број улазних обележја једнак три, онда хипер-раван постаје дводимензионална раван. Постаје тешко замислити када број обележја прелази три. Вектори подршке представљају тачке података које су ближе хипер-равни и утичу на положај и оријентацију хипер-равни. Користећи векторе упоришта, маргина класификатора је максимизована. Брисањем вектора упоришта промениће се положај хипер-равни.

Када се пронађе хипер-раван максималне маргине, SVM алгоритам може да је користи да направи предвиђања за нове опсервације. Са обзиром на нову опсервацију, алгоритам ће је доделити класи која се налази на исправној страни границе одлучивања. На Слици 16. видимо приказ машине вектора упоришта.

У случају линеарно одвојиве бинарне класификације, SVM проналази хипер-раван која раздваја класе са максималном маргином. Ова хипер-раван се може описати једначином:

$$w * x - b = 0$$

где w представља вектор тежина, а b је пристрасност.

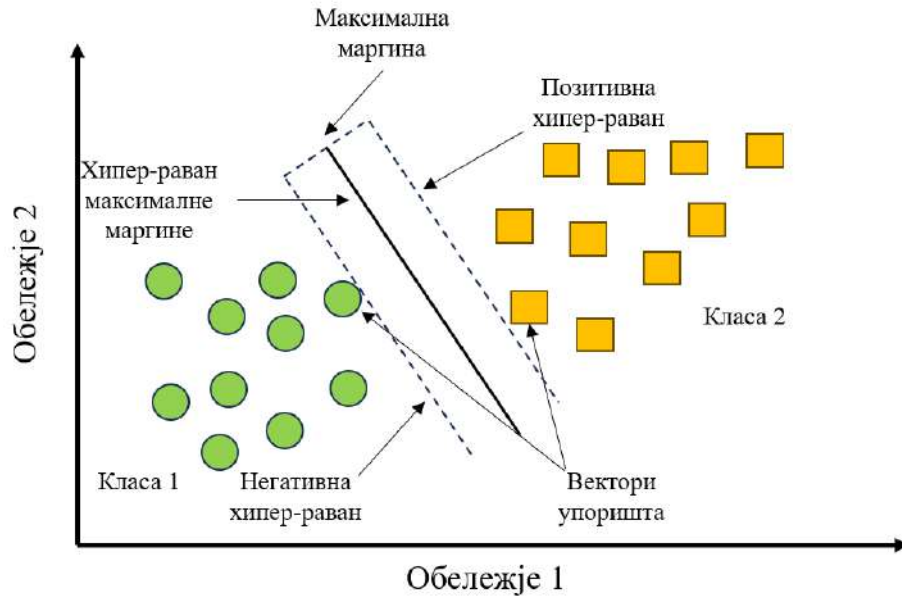
Циљ SVM је да минимизира $\|w\|^2$ субјекат на:

$$y_i(w * x_i - b) \geq 1 \quad \forall i$$

где y_i представља ознаку примера из опсервација за тренирање x_i , а i индексе примера за тренирање.

За нелинеарно одвојиве податке, SVM користи кернел функцију да трансформише улазни простор у простор веће димензије где је могуће пронаћи раздвајајућу хипер-раван.

Типичне кернел функције су:



Слика 16: Машине вектора упоришта

- Линеарна: $\langle x, x' \rangle$
- Полиномиална: $(\gamma \langle x, x' \rangle + r)^d$, где d представља степен полиномиалности
- Радијална базна функција (RBF): $\exp(-\gamma \|x - x'\|^2)$, где је γ параметар.

Кључне претпоставке за употребу SVM алгоритама су:

- Одабир кернел функције: избор кернел функције је кључан и зависи од проблема и података.
- Параметар регулације (C): контролише компромис између постизања мале грешке на подацима за тренирање и минимизирања норме тежина.
- Хиперпараметри: поред параметра C , хиперпараметри специфични за кернел функцију као што је γ у RBF кернелу морају бити пажљиво подешени.

Једна од кључних предности SVM-а је та што могу да рукују високодимензионалним подацима. То је зато што је граница одлучивања одређена само векторима упоришта, који су подскуп података за тренирање. То значи да SVM алгоритам може пронаћи границу одлучивања на коју не утиче „шум“ који се налази у остатку података за тренирање.

Још једна предност SVM-а је та што се регуларизација може извршити једноставно, што значи да се могу прилагодити како би се избегло претерано прилагођавање на подацима за тренирање.

SVM алгоритам се може користити за широк спектар задатака класификације и регресије. Неке од главних области примене SVM-а укључују класификацију текста

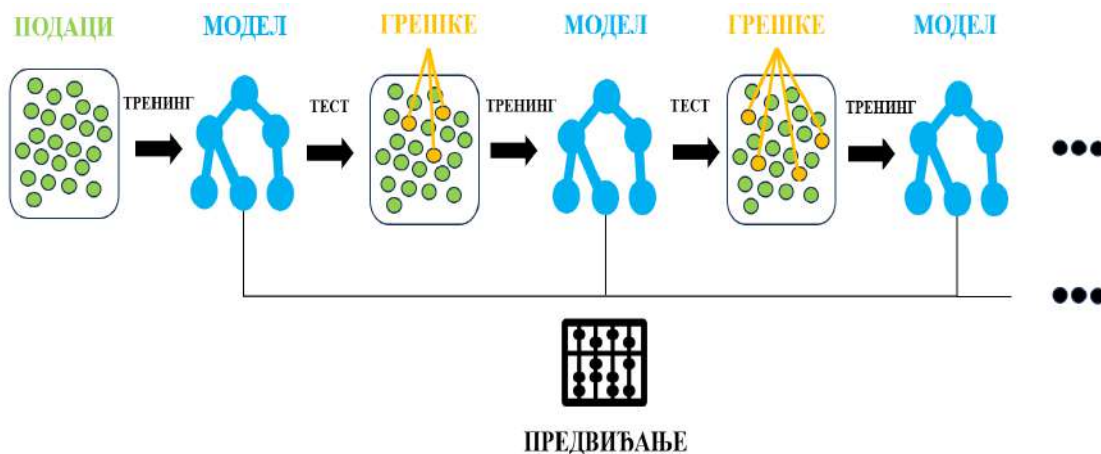
и докумената, препознавање рукописа, медицинску дијагнозу, финансијску анализу. Ово је само неколико примера многих могућих примена SVM алгоритма. Генерално, SVM алгоритми су погодни за било који задатак где је потребно класификовати или предвидети на основу великог броја карактеристика.

3.4.10 Алгоритам повећавања градијента

Алгоритам повећавања градијента (енгл. Gradient Boosting Algorithm) представља моћну и свестрану технику машинског учења, познату по својој ефикасности у задацима класификације и регресије. Припада породици ансамбл метода, што значи да гради модел у фазама, и представља облик појачавања перформанси модела, где сваки нови модел постепено исправља грешке које су направили претходни модели.

У суштини, алгоритам повећавања градијента укључује изградњу низа тзв. слабих ученика, односно модела са слабом моћи учења, обично стабала одлучивања. Свако стабло у низу се фокусира на исправљање остатака или грешака претходног стабла. Израз градијент у његовом називу потиче од његове употребе градијентног спуштања, алгоритма за оптимизацију, како би се минимизирао губитак или разлика између стварних и предвиђених вредности.

За разлику од алгоритма насумичне шуме, који паралелно гради стабла, алгоритам повећавања градијента гради једно по једно стабло. Свако ново стабло је постављено на модификовану верзију оригиналног скупа података, што прогресивно смањује остатке. Алгоритам комбинује ове слабе ученике у једног јаког ученика на начин да збир њихових резултата или предвиђања даје коначни резултат. Овај процес секвенцијалног додавања модела, уз исправљање остатака, доводи до моћног модела предвиђања. Процес рада алгоритма повећавања градијента приказан је на Слици 17.



Слика 17: Алгоритам повећавања градијента

Претпоставимо да имамо скуп података са обележјима X као и циљану променљиву y . Циљ је да се пронађе модел $F(X)$ који минимизира очекивану вредност функције губитка $L(y, F(X))$.

Модел $F(X)$ се гради итеративно. Почевши од константног модела $F_0(X)$, сваки следећи модел $F_i(X)$ се ажурира у складу са:

$$F_i(X) = F_{i-1}(X) + v * h_i(X)$$

где v представља стопу учења и $h_i(X)$ је слаб ученик трениран на негативном градијенту функције губитка.

Стопа учења v контролише колико снажно свако стабло покушава да исправи грешке својих претходника. Мање вредности чине модел робуснијим, али захтевају више стабала да би конвергирала.

Кључне претпоставке за употребу алгоритма повећавања градијента су:

- Број стабала: више стабала може довести до прецизнијег модела, али повећава ризик од претераног прилагођавања.
- Стопа учења: кључни хиперпараметар који балансира брзину учења са ризиком од претераног прилагођавања.
- Дубина стабала: плића стабла смањују варијансу и сложеност, али могу повећати пристрасност.

Једна од кључних карактеристика алгоритма повећавања градијента је његова способност да рукује различитим типовима функција губитка, што га чини прилагодљивим широком спектру проблема, укључујући регресију, класификацију и рангирање. Флексибилност у избору функције губитка омогућава да се она прилагоди специфичним потребама података или задатку.

Још једна предност алгоритма повећавања градијента је његова отпорност на претерано прилагођавање, посебно са структурираним подацима. Коришћење параметара као што су број стабала, дубина стабала и брзина учења обезбеђује контролу над сложенošћу модела, што омогућава ефикасно балансирање пристрасности и варијансе.

Међутим, алгоритам повећавања градијента може бити рачунарски интензиван и спорији за тренирање, пошто се стабла граде узастопно. То такође значи да може бити мање погодан за веома велике скупове података или у сценаријима где је брзина критични фактор.

Алгоритам повећавања градијента се нашироко користи у различитим применама, од побољшања алгоритама за претрагу интернета до победа на такмичењима у машинском учењу. Његова популарност произилази из његове ефикасности у широком спектру подешавања и његове способности да производи високо прецизне моделе.

3.4.11 Вештачке неуронске мреже

Вештачке неуронске мреже (енгл. Artificial Neural Networks - ANN) су рачунарски модели који су инспирисани структуром и функционисањем људског мозга. Састоје се од многих међусобно повезаних јединица за обраду, или „неурона“, који су организовани у слојевима. Неурони у ANN примају улазне сигнале из спољашњег окружења или од других неурона у мрежи и користе те сигнале за израчунавање излазног сигнала. Излазни сигнал се затим преноси на друге неуроне или у спољашње окружење.

Везе између неурона, назване "ивице", имају тежине (пондере) које одређују снагу везе. Излаз неурона се израчунава применом нелинеарне функције на пондерисани збир

његових улаза. Јачина сигнала на свакој вези може се повећати или смањити променом тежине. Излаз неурона се затим преноси на друге неуроне у мрежи.

Неурони примењују активационе функције на свој улаз. Уобичајене активационе функције укључују сигмоидалну (енгл. sigmoid), ReLU (енгл. Rectified Linear Unit) и софтмакс (енгл. Softmax). Активационе функције уводе нелинеарност, омогућавајући мрежи да моделира сложене односе.

Поред, улаза и тежина, пристрасност (енгл. bias) представља критичну компоненту која, која одређује излаз сваког неурона у мрежи. Улога пристрасности је да сваком неурону обезбеди подесиву константну вредност (поред нормалних улаза које неурон прима). Пристрасност омогућава да се активациона функција помера. Ова промена помаже неурону да боље уклопи податке са већом флексибилношћу. Без пристрасности, излаз неурона је стриктно зависан од пондерисане суме његових улаза, што ограничава опсег вредности које неурон може да избацаи.

Излаз у из неурона ANN-а се представља формулом:

$$y = f \left(\sum_{i=1}^n (w_i \cdot x_i) + b \right)$$

Где је:

- y је излаз из неурона.
- f је активациона функција примењена на неуроноу улаз.
- $\sum_{i=1}^n (w_i \cdot x_i)$ представља пондерисану суму улаза, где w_i представља тежине i -ог улаза, а x_i је i -и улаз у неурон.
- b је пристрасност везана за неурон.

Тежине и пристрасности су параметри који се могу тренирати. Током процеса тренирања, ANN прилагођава своје тежине и пристрасности на основу грешке свог излаза у поређењу са стварним резултатом. Ово подешавање се врши коришћењем алгоритама као што је пропација уназад у комбинацији са градијентом спуштања.

Функција губитка мери колико добро мрежа ради. Функција губитка квантификује разлику између предвиђених вредности и стварних вредности. За скуп примера за тренирање, обично се дефинише као просечни губитак у свим примерима. Уобичајене функције губитка укључују средњу квадратну грешку за задатке регресије и унакрсну ентропију за задатке класификације. Циљ тренирања је да се минимизира ова функција губитка.

MSE се дефинише као:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

где N представља број опсервација за тренирање, y_i представља стварне вредности, и \hat{y}_i су предвиђене вредности.

Унакрсна ентропија се дефинише као:

$$\text{унакрсна ентропија} = - \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$$

Пропагација уназад је метода која се користи за израчунавање градијента функције губитка у односу на сваку тежину у мрежи, применом ланчаног правила рачуна.

Са L је означена функција губитка (нпр. MSE за регресију). Градијент од L у односу на тежину w_i је:

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial y} * \frac{\partial y}{\partial w_i}$$

Градијент од L у односу на пристрасност b је:

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y} * \frac{\partial y}{\partial b}$$

Применом ланчаног правила рачуна се добија, пошто је $y = f(z)$ и $z = \sum_{i=1}^n w_i x_i + b$, добија се:

$$\frac{\partial y}{\partial w_i} = \frac{\partial f(z)}{\partial z} * x_i$$

као и,

$$\frac{\partial y}{\partial b} = \frac{\partial f(z)}{\partial z} * 1.$$

Градијентно спуштање или његове варијанте се затим користе за ажурирање тежи-на и предрасуда, постепено побољшавајући перформансе модела.

Ажурирање правила за тежине:

$$w_i = w_i - \eta * \frac{\partial L}{\partial w_i}$$

где η представља стопу учења.

Ажурирање правила за пристрасност:

$$b = b - \eta * \frac{\partial L}{\partial b}$$

Архитектура ANN, укључујући број слојева и број неурона у сваком слоју, игра кључну улогу у перформансама модела. Број слојева и неурона у ANN-у може да варира у зависности од сложености проблема за чије решавање се користи.

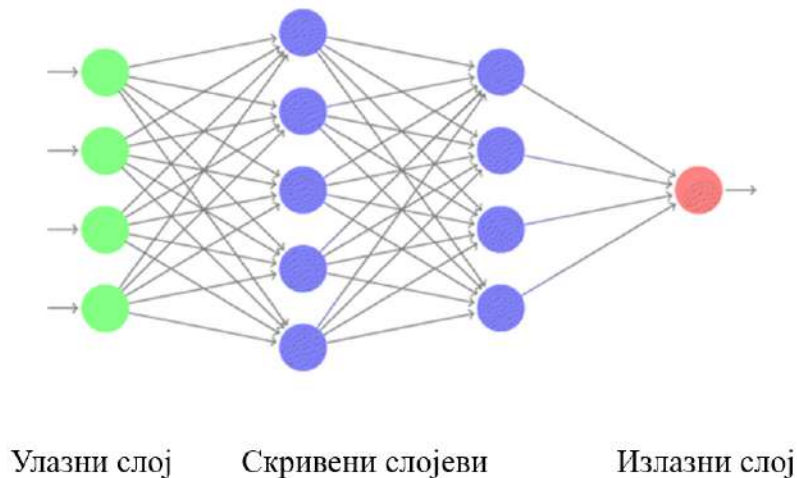
Први слој у ANN је улазни слој, који прима улазне сигнале из спољашњег окружења. Ови сигнали могу бити необрађени подаци, као што су вредности пиксела на слици, или

могу бити претходно обрађена обележја података извучена из података за тренирање. Број неурона у улазном слоју једнак је броју улазних обележја.

Следећи слојеви у ANN су скривени слојеви, који представљају међуслојеве између улазног и излазног слоја. Број скривених слојева у ANN-у може да варира и може имати значајан утицај на перформансе модела. Уопштено говорећи, више скривених слојева може омогућити моделу да научи сложеније обрасце у подацима, али такође може повећати ризик од претераног прилагођавања.

Последњи слој у ANN-у је излазни слој, који производи коначни излаз модела. Број неурона у излазном слоју је одређен бројем класа у проблему, у случају да се ANN користи за класификацију. На пример, ако се ANN користи за класификацију слика мачака и паса, излазни слој би имао два неурона, по један за сваку класу.

На Слици 18. приказана је вештачка неуронска мрежа.



Слика 18: Вештачка неуронска мрежа

Када се сама архитектура ANN одреди, модел се може тренирати коришћењем алгорита учења, као што је пропација грешке уназад, да би се подесиле тежине ивица како би се произвео жељени излаз за дати улаз. Током тренирања, ANN приказује скуп улазно-излазних парова, а тежине ивица се прилагођавају како би се минимизирала разлика између предвиђеног излаза модела и правог излаза. Овај процес се наставља све док модел не достигне задовољавајући ниво перформанси на подацима за тренирање.

Када је модел истрениран, може се користити за предвиђање на новим подацима. Улазни подаци се преносе кроз мрежу, а излаз модела се користи за доношење одлука или предвиђања на основу података.

Вештачке неуронске мреже су моћно средство за решавање широког спектра проблема и широко су коришћене у многим различитим областима, укључујући рачунарски вид, обраду природног језика и роботiku.

4 ДУБОКО УЧЕЊЕ

Дубоко учење (енгл. Deep Learning - DL) представља подскуп машинског учења који користи вештачке неуронске мреже за учење из података. У случајевима када се вештачка неуронска мрежа састоји из више скривених слојева она се назива дубока неуронска мрежа. Обично, ако је број скривених слојева већи од три, сматра се да је у питању дубока неуронска мрежа [58].

ML алгоритми се могу класификовати на алгоритме за учење са надзором, полу-надзирано учење, лабаво учење и учење без надзора [59], у зависности од тога како су тренирани. Ови алгоритми се могу ослањати на различите архитектуре дубоких неуронских мрежа (енгл. Deep Neural Networks - DNN), као што су на пример рекурентне неуронске мреже или конволуционе неуронске мреже. DL алгоритми уклањају инжењерске и људске напоре за обављање посебних задатака на подацима, тј. отклањају потребу да људски експерт ручно креира најважнија обележја података, тако што трансформишу податке у скривене слојеве и аутоматски издвајају значајна обележја. DL алгоритми се такође могу користити за проблеме учења без надзора када означени подаци нису доступни или означени подаци нису довољни. На пример, аутоенкодери [60] се могу користити за проблеме учења без надзора. Генеративне супростављене мреже (енгл. Generative Adversarial Networks - GAN), такође могу да генеришу синтетичке податке на основу скупа ограничених узорака података. Дистрибуција генерисаних података је слична стварним подацима, али сами подаци нису исти [61].

DL алгоритми се примењују у различитим доменима као што су рачунарски вид [62], локализација заснована на отиску прста [63], препознавање говора [64], анализа медицинских снимака [65], филтрирање садржаја друштвених мрежа [66], биоинформатика [67], итд.. DL алгоритми су показали супериорне перформансе у реалним практичним применама, чак и у поређењу са људима. Модел АлфаГо научен техникама дубоког учења да игра игру Го, која се сматра много компликованијом игром у односу на шах, победио је људског шампиона [68].

У дубоком учењу, слојеви представљају суштинску компоненту у трансформацији улазних података у излазне податке. Сваки слој је истрениран да трансформише свој улаз у нову репрезентацију, познату као обележје података. На пример, у обради слике, улазни подаци могу бити 2Д или 3Д низ или вектор пиксела. У таквим случајевима, конволуциони слојеви се обично користе у слојевима ближим почетку мреже, док се потпуно повезани слојеви користе у излазном слоју да би се одредила вероватноћа да улаз припада свакој класи. На пример, у проблему класификације лица, излаз првог скривеног слоја могу бити ивице лица, док други слој може издвојити карактеристике као што су очи, трепавице, нос и уши. Излазни слој затим додељује вероватноћу на основу детаљне репрезентације сваког претходног слоја. Важно је напоменути да не постоји фиксно правило за одређивање броја слојева и неурона у сваком слоју.

Иако у литератури можемо пронаћи различите архитектуре неуронских мрежа, све се ослањају на исте фундаменталне елементе, као што су фактори поларизације, тежине, неуроне и активационе функције, које су инспирисане понашањем неурона у људском мозгу. DL алгоритми се могу обучити на сличан начин као и други алгоритми машинског учења. На пример, ако је циљ DL модела да разликује пилиће на сликама, скуп података означених слика пилића (означених вредношћу 1) се доставља моделу и

користи се за тренирање модела.

Дубоке неуронске мреже су способне да апроксимирају сложене нелинеарне функције. Ови модели трансформишу улазне карактеристике у детаљније приказе кроз узастопне слојеве, на крају омогућавајући процену на основу ових детаљних карактеристика података у излазном слоју. Скривени слојеви су одговорни за учење ових детаљних репрезентација, при чему слојеви који су ближи улазном слоју садрже апстрактније карактеристике, а они који су ближи излазном слоју садрже детаљније карактеристике [69].

Иако архитектуре модела дубоког учења могу имати много различитих облика, све оне деле неке основне архитектонске принципе. На пример, генеративне супростављене мреже, имају много варијанти и развијених верзија, али су све засноване на оригиналном GAN моделу. Свака архитектура је прилагођена одређеном домену и можда неће бити директно упоредива са другима јер је дизајнирана за специфичне задатке. DNN тежине се обично иницијализују насумичним вредностима, али ове вредности нису оптималне и морају се оптимизовати и ажурирати на основу података за тренирање. Једном обучен, модел дубоког учења може препознати обрасце у новим улазним подацима, познатим као први пут „виђени“ подаци.

4.1 Конволуционе неуронске мреже

Конволуционе неуронске мреже (енгл. Convolutional Neural Networks - CNN) су тип неуронских мрежа дизајнираних за обраду и анализу података са топологијом налик 2Д мрежи, као што су на пример слике (фотографије). CNN су посебно корисне за задатке као што су класификација слика и детекција објеката, јер су у стању да науче карактеристике и обрасце слика (фотографија) директно из података.

Тренирање конволуционе неуронске мреже подразумева коришћење алгоритма оптимизације, као што су стохастички градијентни спуст (енгл. Stochastic Gradient Descent - SGD) како би се подесиле тежине и фактори поларизације мреже како би се минимизирала грешка. Грешка мери разлику између предвиђеног излаза CNN-а и правих вредности улазних података.

За тренирање CNN-а, потребни су велики скупови података означених слика (фотографија). У CNN се затим убацују улазни подаци, а излаз се пореди са правим ознакама употребом функције губитка. Градијенти грешке у односу на тежине и факторе поларизације се затим израчунавају коришћењем алгоритма пропагације грешке уназад. Вредности тежина и фактора поларизације се затим ажурирају у супротном смеру од градијената, користећи унапред дефинисану стопу учења, како би се смањила грешка. Овај процес се понавља током више епоха, све док вредност функције губитка не дође до минимума.

У машинском учењу, параметри и хиперпараметри представљају две врсте променљивих које се користе за тренирање и подешавање модела.

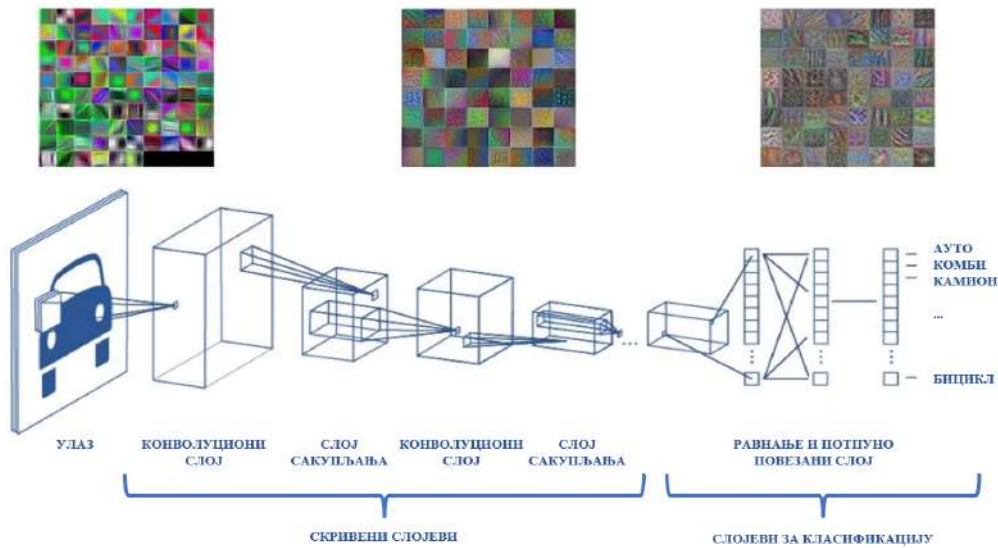
Параметри су варијабле које модел учи током тренирања да би направио предвиђања на основу нових података. То су интерне поставке модела које се оптимизују током тренирања како би се минимизирала грешка између предвиђених и стварних излаза. У неуронским мрежама, параметри су тежине и фактори поларизације које одређују снагу и правац веза између неурона.

Хиперпараметри су, с друге стране, екстерне поставке модела које се задају пре тренирања и одређују како се модел обучава. Они се не уче током тренирања, већ их ручно поставља инжењер машинског учења. Хиперпараметри могу утицати на перформансе модела, сложеност и способност генерализације на новим подацима.

Кључна разлика између параметара и хиперпараметара је у томе што се параметри уче током тренирања, док се хиперпараметри постављају пре тренирања и одређују како се модел обучава. Процес проналажења оптималних хиперпараметара за модел познат је као подешавање хиперпараметара или оптимизација хиперпараметара.

Постоји неколико кључних хиперпараметара који могу утицати на перформансе CNN-а током тренирања. Ово укључује стопу учења, која одређује величину корака за ажурирање тежина и фактора поларизације, и величине подскупа за евалуацију градијента, која одређује број примера обрађених у свакој итерацији. Избор алгорита за оптимизацију и употреба техника регуларизације, као што су изостављање или ерозија тежина, такође могу утицати на перформансе CNN-а.

CNN се састоје од низа слојева који обрађују и трансформишу улазне податке. Слојеви се могу поделити у три главна типа: конволуциони слојеви, слојеви скупљања и потпуно повезани слојеви. Приказ конволуционе неуронске мреже је дат на Слици 19.



Слика 19: Конволуциона неуронска мрежа за класификацију[70]

Конволуциони слојеви примењују скуп филтера на улазне податке да би издвојили карактеристике података. Ови филтери су обично мали, често 3x3 или 5x5 пиксела, и примењују се на улазне податке коришћењем приступа „клизећег прозора“. Како се филтер креће по улазним подацима, израчунава се скаларни производ између улаза и филтера, што резултира 2Д мапом активације или мапом обележја. Тежине филтера се тренирају коришћењем одабраног алгорита оптимизације да би се минимизирала грешка, која мери разлику између предвиђених резултата класе и правих ознака података за тренирање.

Један кључни аспект конволуционих слојева је употреба локалних веза и заједничких тежина. У вишеслојном перцептрону, сваки неурон прима улаз од свих неурона из

претходног слоја. Ово може бити рачунски скупо и захтева велики број параметара. У конволуционом слоју, сваки неурон прима само улаз из малог региона улазних података, познатог као поље утицаја. Ово смањује број параметара и чини модел ефикаснијим за тренирање. Поред тога, исти скуп тежина се дели за цео улаз, што значи да се исти филтери примењују на сваки регион улазних података. Ово омогућава моделу да се иста карактеристика може открити без обзира на њену позицију у улазним подацима.

Други важан аспект конволуционих слојева је употреба активационих функција. Активационе функције уводе нелинеарност у модел и омогућавају му да научи сложене карактеристике. Уобичајене функције активације за конволуционе слојеве укључују ReLU, која примењује функцију $f(x) = \max(0, x)$ и сигмоидалну, која примењује функцију $f(x) = 1 / (1 + \exp(-x))$.

Слојеви скупљања се користе како би се смањила висина и ширина активационих мапа применом функције као што је максимално скупљање или просечно скупљање на мали део улаза. Максимално скупљање бира максималну вредност из улазног региона, док просечно скупљање израчунава просечну вредност. Слојеви скупљања се обично примењује на излаз конволуционог слоја и служе за смањење величине података и броја параметара у моделу. Такође помажу да се модел учини робуснијим, јер агрегира активације у малом региону.

Потпуно повезани слојеви су исти као у другим неуронским мрежама и користе се на крају CNN-а за мапирање екстрахованих карактеристика у коначни излаз, као што је ознака класе или гранични оквир у случају задатка детекције објеката.

Конволуционе неуронске мреже су тип неуронских мрежа која су посебно дизајниране за обраду и анализу дводимензионалних података, као што су слике. CNN су веома успешне у широком спектру примена, као што су класификација слика, детекција објеката и генерисање слика. Међутим, као и сваки други алгоритам машинског учења, CNN такође имају нека ограничења и недостатке.

Нека од главних ограничења CNN-а укључују:

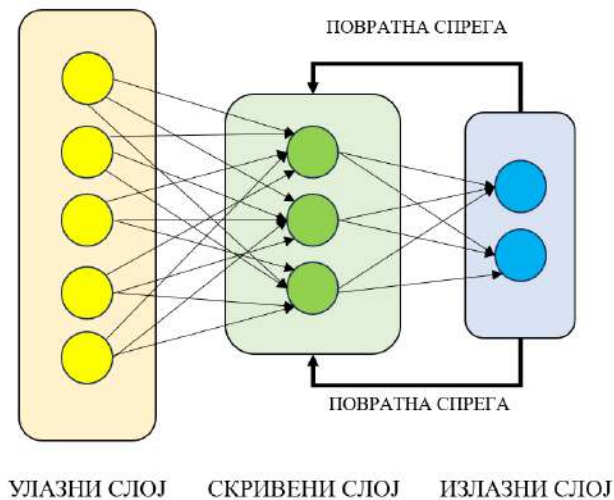
- Комплексност тренирања: CNN-ове може бити тешко истренирати, посебно када се ради са великим или сложеним скуповима података. Ово је делимично због великог броја параметара које треба научити, што може захтевати много рачунарских ресурса и времена.
- Захтеви за меморијом: CNN-овима може бити потребна значајна количина меморије за складиштење тежина и активација мреже. Ово може бити проблем када се ради са веома великим скуповима података или када се тренирају веома дубоке мреже.
- Ограничени контекст: CNN су генерално у стању да размотре само локални контекст унутар слике, а не глобални контекст. Ово може бити ограничење у случају да се покушава моделовање сложенијих зависности унутар података.
- Ограничена применљивост: Иако су CNN-ови добро прилагођени за обраду дводимензионалних података, као што су слике, они нису толико ефикасни у обради других типова података, као што су текст или аудио. За ове врсте података, друге

архитектуре неуронских мрежа, као што су рекурентне неуронске мреже (енгл. Recurrent Neural Networks - RNN), могу бити прикладније.

4.2 Рекурентне неуронске мреже

Рекурентне неуронске мреже су тип неуронских мрежа које су погодне за обраду секвенцијалних података, као што су временске серије, природни језик или говор. Називају се „понављајућим“ јер користе повратну спрегу за преношење информација из једног временског корака у други, што им омогућава да моделују зависности између временских корака и да предвиђају на основу читаве улазне секвенце.

RNN имају способност да запамте и уграде информације из прошлих временских корака у своје тренутно стање. RNN имају различите примене, укључујући машинско превођење, предвиђање кретања на берзанском пословању и препознавање говора. Често их користе познате компаније у својим алатима, као што су Apple виртуелни асистент Сири или Google гласовна претрага. За разлику од конволуционих неуронских мрежа, које немају сачуване податке о прошлим улазима, RNN имају интерну меморију и погодне су за задатке машинског учења који укључују секвенцијалне податке. RNN-ови су играли значајну улогу у успеху дубоког учења последњих година. На слици 20. имамо поједностављен приказ архитектуре рекурентне неуронске мреже.



Слика 20: Рекурентна неуронска мрежа

Типичне архитектуре RNN у употреби су:

- Архитектура један према један: Овај тип RNN-а мапира један улаз у један излаз. Пример овога би било предвиђање следеће речи у реченици, где је улаз текућа реч, а излаз следећа реч. Када је више RNN-ова један према један везано заједно, они се могу користити за генерисање реченица или чак дужег текста, што је процес познат као моделирање језика.
- Архитектура један према много: У овом типу RNN-а, улаз је један елемент, а излаз је низ елемената. Пример овога представља додавање објашњења сликама, где је једна слика улаз, а натпис који се састоји од више речи је излаз.
- Архитектура много према један: У овом типу RNN-а, улаз је низ елемената, а излаз је један елемент. Пример овога је анализа осећаја (енгл. sentiment analysis), где је преглед или реченица која се састоји од више речи улаз, а излаз је једна ознака која указује на осећање (нпр. позитивно или негативно).
- Архитектура много према много: У овом типу RNN-а, и улаз и излаз су низови елемената. Пример за то је машински превод, где је реченица на једном језику улаз, а преведена реченица на другом језику је излаз. Број елемената у улазној и излазној секвенци не мора бити исти или у истом редоследу.

У рекурентним неуронским мрежама градијенти тежина у односу на грешку могу или експлодирати или нестати. Ово може отежати или онемогућити ефикасно тренирање модела.

Експлодирајући градијенти настају када вредности градијената постану изузетно велике, што може довести до тога да модел припише превелики значај одређеним тежинама. То може довести до нестабилних модела са лошим перформансама.

С друге стране, градијенти који нестају настају када вредности градијената постану изузетно мале, што моделу може отежати учење. Ово може бити проблем јер моделу може бити потребно много времена да конвергира, или уопште неће конвергирати.

Како експлодирајући тако и нестајући градијенти су узроковани поновљеним множењем градијената док модел обрађује сваки временски корак у низу. Ово може довести до градијената који временом постају веома велики или веома мали, у зависности од почетних вредности тежина и градијената.

Рекурентне неуронске мреже имају бројне предности, укључујући њихову способност да обрађују секвенцијалне податке и поседовање интерне меморије, што им омогућава да уграде информације из прошлих временских корака. Међутим, као и сваки алгоритам за машинско учење, РНН такође имају неке недостатке.

Нека од главних ограничења RNN-а укључују:

- Комплексност тренирања: RNN-ове може бити тешко истренирати, посебно када се ради са великим или сложеним скуповима података. Ово је делимично због чињенице да градијенти у RNN-овима могу лако да експлодирају или нестану, што може представљати изазов за оптимизацију модела.

- **Захтеви за меморијом:** RNN-ови захтевају значајну количину меморије за складиштење унутрашњег стања и тежина мреже. Ово може бити проблем када се ради са великим скуповима података или када се тренирају веома дубоке мреже.
- **Ограничени контекст:** RNN-ови могу запамтити само ограничену количину контекста због проблема нестајања и експлозије градијента. Ово може бити ограничење када се ради са веома дугим секвенцама или када се покушава моделовање сложених зависности.
- **Сложеност закључивања:** Закључивање (или предвиђање) са RNN-овима може бити рачунарски интензивно, посебно када се ради са веома дубоким мрежама. Ово може бити проблем при коришћењу RNN-ова у применама у реалном времену које захтевају брзо време предвиђања.
- **Ограничена применљивост:** Иако су RNN-ови добро прилагођени за обраду секвенцијалних података, нису толико ефикасни у обради података који нису секвенцијални, као што су слике или аудио. За ове врсте података, друге архитектуре неуронских мрежа, као што су конволуционе неуронске мреже, могу бити прикладније.

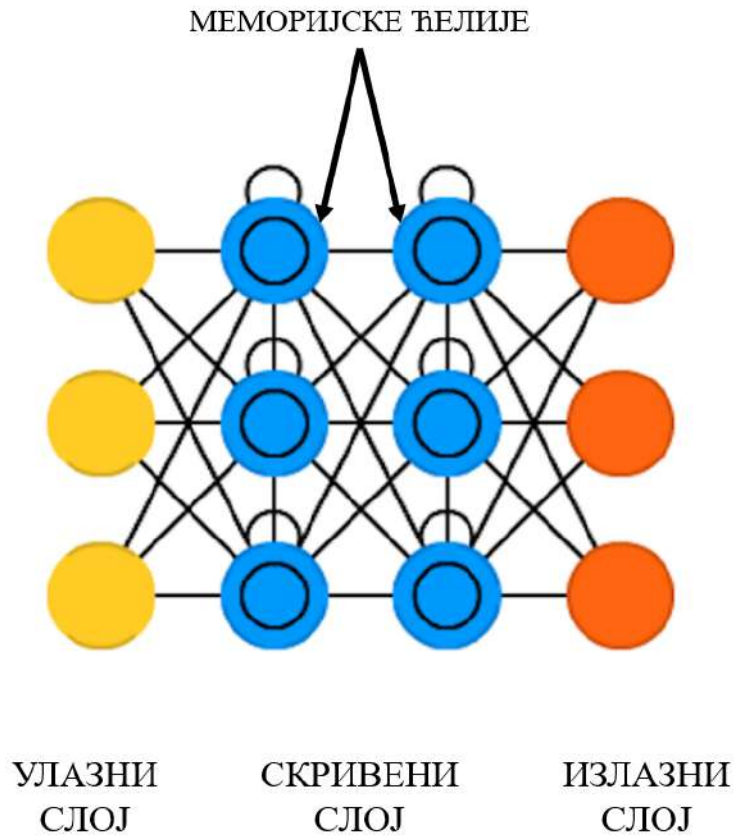
4.3 Мреже са дугим и кратким памћењем

Мреже са дугим и кратким памћењем (енгл. Long short term memory networks - LSTM) су врста рекурентних неуронских мрежа које су у стању да уче и памте дугорочне зависности у секвенцијалним подацима. Ове мреже су посебно дизајниране да превазиђу ограничења традиционалних рекурентних неуронских мрежа, које се често муче да задрже информације током дужег временског периода.

LSTM мреже се састоје од низа међусобно повезаних „меморијских ћелија“ које могу да читају, пишу и бришу информације из своје интерне меморије. Свака ћелија има три „капије“ које контролишу ток информација: улазну капију, капију заборава и излазну капију. Проблеми нестајања и експлодирања градијента могу се ублажити употребом капија. Улазна капија омогућава да се нове информације уписују у меморијску ћелију, капија заборава омогућава брисање информација из меморијске ћелије, а излазна капија омогућава коришћење информација из меморијске ћелије у тренутном временском кораку. Наведене капије се имплементирају помоћу сигмоидалних функција, које су ограничене између 0 и 1 и могу спречити да градијенти постану превелики или мали.

LSTM мреже такође користе механизам који се назива „везе из шпијунке“ који омогућава капијама да директно приступе претходном стању меморијске ћелије, додатно побољшавајући способност модела да учи и задржава информације. На Слици 21. имамо поједностављен приказ архитектуре LSTM неуронске мреже.

LSTM мреже се широко користе у различитим применама, укључујући обраду природног језика, препознавање говора и предвиђање временских серија. LSTM мреже су се показале као успешно решење у бројним задацима за које се раније сматрало да су изван могућности рекурентних неуронских мрежа и сматрају се једним од најсавременијих алгоритама за секвенцијалне податке.



Слика 21: LSTM неуронска мрежа[71]

Постоји неколико потенцијалних недостатака коришћења LSTM мрежа у дубоком учењу:

- Сложеност: LSTM мреже су релативно сложене у поређењу са другим типовима неуронских мрежа, што може отежати њихово разумевање и имплементацију.
- Време тренирања: LSTM мрежама може бити потребно дуже да се истренирају од других типова неуронских мрежа због њихове сложене структуре.
- Претерано прилагођавање: LSTM мреже могу бити склоне претераном прилагођавању, посебно ако нису правилно регуларизоване.
- Ограничена интерпретабилност: Услед комплексности LSTM мрежа може бити изазовно да се протумачи њихово интерно функционисање, што може отежати разумевање како LSTM мреже праве предвиђања.
- Ограничена доступност висококвалитетних података: Да би се истренирале ефикасне LSTM мреже, важно је имати велике количине података високог квалитета, који можда нису увек доступни.

- Ограничена генерализација: LSTM мреже можда неће бити тако ефикасне у генерализацији нових података као друге врсте неуронских мрежа, посебно ако се нови подаци значајно разликују од података коришћених за тренирање мреже.

4.4 Трансформери

Трансформери су револуционарна класа архитектура вештачких неуронских мрежа које су значајно унапредиле области обраде природног језика (енгл. Natural Language Processing - NLP), рачунарског вида и шире. Они су представљени у раду "Пажња је све што вам треба"[72] написаног од стране Васванија и сарадника 2017. године и од тада су представљали базу за многе водеће моделе као што су BERT, GPT и још много тога.

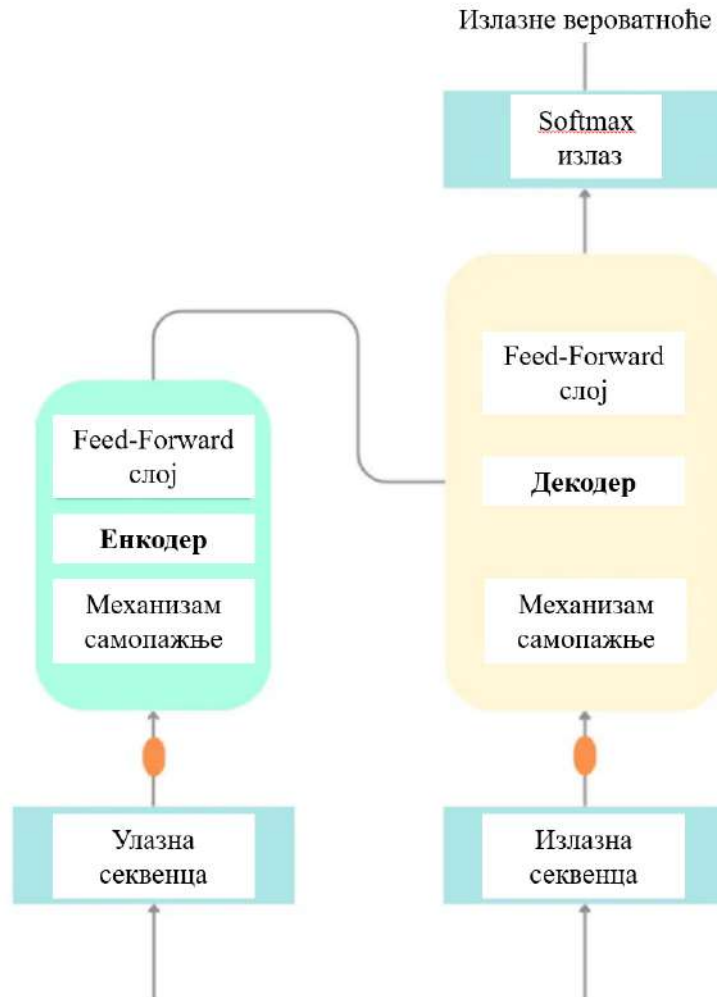
За разлику од претходних архитектура, трансформатори се у потпуности ослањају на механизам пажње, избегавајући традиционалне рекурентне слојеве. Ово им омогућава да паралелно обрађују улазне секвенце, значајно повећавајући ефикасност и способност да ухвате сложене зависности података. Језгро трансформатора је механизам самопажње, који израчунава одговор на позицији у низу пратећи све позиције и извлачећи глобалне зависности, без обзира на њихову удаљеност у улазној секвенци.

Типичан модел трансформера састоји се од енкодера и декодера: енкодер мапира улазну секвенцу у низ континуираних репрезентација, а декодер затим генерише излазну секвенцу из ових репрезентација. Хиперпараметри модела, укључујући број слојева, величину глава пажње и димензионалност мрежа за пропагацију унапред, кључни су за његове перформансе и захтевају пажљиво подешавање. Приказ трансформер архитектуре је дат на Слици 22.

Трансформери су широко прихваћени за различите примене, проширујући свој домет изван NLP-а на области као што су обрада слика и геномика. Њихова способност да рукују секвенцијалним подацима и хватају дугорочне зависности била је посебно корисна у превођењу језика, резимирању текста и системима за одговоре на питања.

Упркос својим предностима, трансформери долазе са низом изазова:

- Компутациони интензитет: Модели трансформера за тренирање обично захтевају значајне рачунарске ресурсе, укључујући напредне GPU или TPU јединице, због њихове сложене архитектуре и великог броја параметара.
- Потрошња меморије: Трансформери могу бити меморијски интензивни, посебно за дуге секвенце, пошто механизам самопажње захтева прорачуне за сваки елемент у низу.
- Хиперпараметарска осетљивост: Перформансе модела трансформера могу бити веома осетљиве на избор хиперпараметара. Проналажење правог скупа параметара често захтева опсежно експериментисање и фино подешавање.
- Захтевају велике количине података: Модели трансформера генерално захтевају велике количине података за тренирање да би ефикасно функционисали. Ово може бити ограничење у сценаријима где су подаци оскудни или скупи за добијање.



Слика 22: Трансформер архитектура[72]

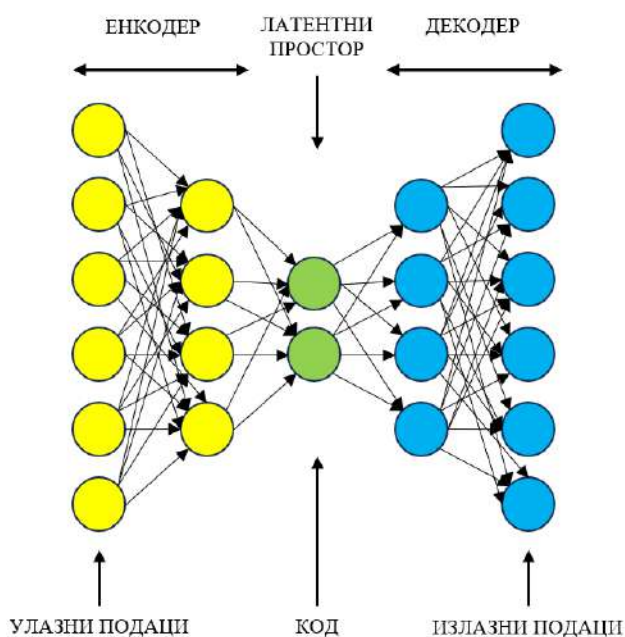
- Ризик од превеликог прилагођавања: С обзиром на њихову дубину и сложеност, трансформатори су склони превеликом прилагођавању, посебно када нема довољно података за тренирање. Технике регуларизације и пажљиво одређивање скупова података су неопходни да би се ово решило.
- Ограничена интерпретабилност: Као и многим моделима дубоког учења, трансформерима недостаје интерпретабилност, што отежава разумевање или објашњење њихових процеса доношења одлука.

Упркос овим изазовима, архитектура трансформера остаје кључни развој у машинском учењу. Његова способност да ефикасно обрађује и анализира секвенцијалне податке утрла је пут за значајан напредак у различитим доменама.

4.5 Аутоенкодери

Аутоенкодери (енгл. Autoencoders) су врста вештачких неуронских мрежа дизајнираних за задатке учења без надзора. Први пут су предложени од стране Крамера 1991 године [73] под називом ауто-асоцијативне неуронске мреже (енгл. Autoassociative Neural Networks). Састоје се од три слоја: улазног слоја, скривеног слоја и излазног слоја, где су улазни и излазни слој исте величине. Сврха аутоенкодера је да компресује улазне податке у нижедимензионалну репрезентацију која се још назива репрезентација латентног простора, а затим реконструише оригиналне улазне податке из репрезентације.

Процес компресије и реконструкције обављају две компоненте аутоенкодера: енкодер и декодер. Енкодер узима улазне податке и компресује их у репрезентацију латентног простора, познату као код. Декодер затим узима код и реконструише оригиналне улазне податке. Циљ тренирања аутоенкодера је да се минимизира грешка реконструкције, која представља разлику између оригиналних улазних података и реконструисаних података. На Слици 23. је приказана архитектура аутоенкодер неуронске мреже.



Слика 23: Аутоенкодер неуронска мрежа

Аутоенкодери имају неколико хиперпараметара који се могу подесити пре тренирања, укључујући величину кода, број слојева, број чворова по слоју и функцију губитка. Величина кода се односи на број чворова у скривеном слоју, тј. у репрезентацији латентног простора. Мања величина кода доводи до веће компресије, али може довести и до губитка информација. Број слојева и чворова по слоју се може подесити да би се контролисала сложеност аутоенкодера. Функција губитка мери грешку реконструкције и може се изабрати на основу опсега улазних података, са опцијама као што су средња квадратна грешка или бинарна кросентропија.

Аутоенкодери су корисни за задатке као што су смањење димензионалности, компресију података и откривање аномалија. Међутим, они су специфични за податке и нису способни за компресију без губитака. Пронашли су широк спектар примена у различитим индустријама и пословним секторима, користећи своју способност да ефикасно кодирају и тумаче сложене скупове података. У финансијској индустрији, аутоенкодери се користе за откривање аномалија, што је кључно за идентификацију лажних трансакција. У области здравствене заштите и биомедицине, аутоенкодери играју виталну улогу у задацима обраде слике. Користе се за смањење шума у медицинским снимцима, побољшавајући квалитет и јасноћу снимака магнетне резонанце, рендгенских снимака и других техника медицинског снимања. Овај побољшани квалитет слике помаже у прецизнијој дијагнози и анализи. У производњи и контроли квалитета, аутоенкодери се користе за откривање аномалија или недостатака у производима на монтажним линијама. Учењем стандардног изгледа и карактеристика производа, неуронска мрежа може да идентификује јединице које показују абнормалности, чиме се одржавају стандарди квалитета.

Такође су склони претераном прилагођавању ако је мрежа дубока и има велики број чворова по слоју, тако да могу бити неопходне технике регуларизације како би се избегао наведени проблем.

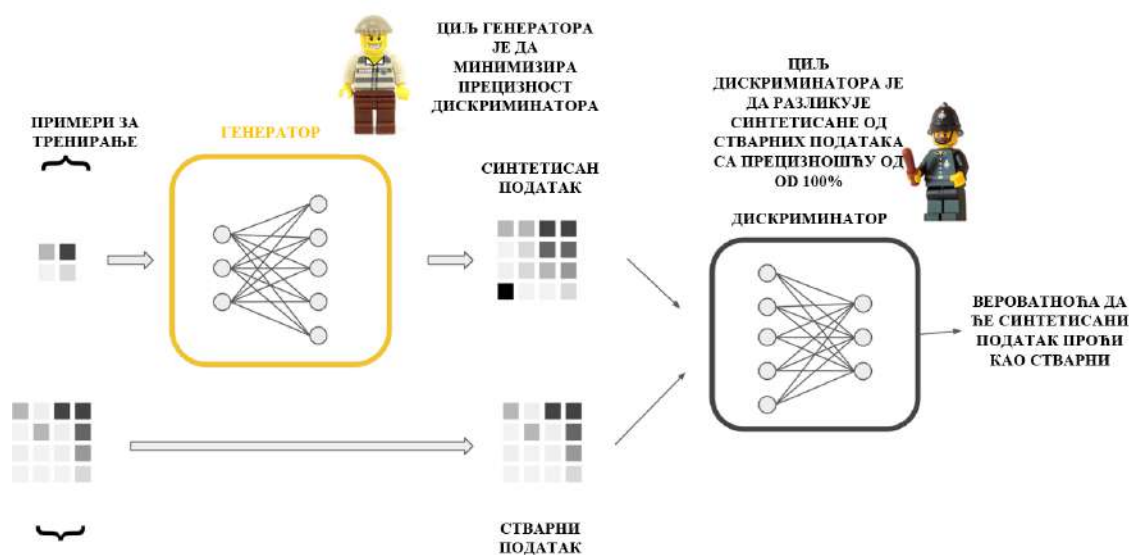
Постоји неколико потенцијалних недостатака коришћења аутоенкодера:

- Ограничена применљивост: Аутоенкодери се првенствено користе за смањење димензионалности и учење карактеристика података и нису прикладни за све врсте проблема.
- Лоше перформансе на великим скуповима података: Аутоенкодери могу бити спори у тренирању на великим скуповима података и можда неће моћи у потпуности да ухвате сложеност података.
- Осетљивост на хиперпараметре: Перформансе аутоенкодера могу бити осетљиве на избор хиперпараметара, као што су број слојева и величина скривеног слоја.
- Ограничена способност снимања сложених односа: Аутоенкодери су ограничени у својој способности да ухвате сложене односе између карактеристика података и можда неће радити добро на веома нелинеарним подацима.
- Компресија са губицима: Аутоенкодери су алгоритми компресије са губицима, што значи да реконструисани излаз можда неће бити идентичан оригиналном улазу. Ово може бити проблем за задатке где је важна висока сличност са оригиналним подацима.
- Захтевају велике количине података: Аутоенкодери генерално захтевају велике количине података да би научили смислене репрезентације и можда неће добро радити на малим скуповима података.

4.6 Генеративне супростављене мреже

Генеративне супростављене мреже су класа модела дубоког учења који су дизајнирани за учење без надзора, као унапређена верзија аутоенкодера. GAN генеришу синтетичке податке који су слични датом скупу података за тренирање. Састоје се од две главне компоненте: генератора и дискриминатора. Генератор је истрениран да производи синтетичке податке који су слични подацима за тренирање, док је дискриминатор истрениран да прави разлику између синтетичких података и стварних података.

Током обуке, генератор и дискриминатор се обучавају истовремено на супростављен начин. Генератор покушава да произведе синтетичке податке који се не разликују од стварних података, док дискриминатор покушава да прецизно класификује синтетичке податке као лажне, а праве податке као стварне. Како генератор напредује у генерисању синтетичких података који су слични подацима за тренирање, дискриминатор постаје бољи у разликовању синтетичких и стварних података. Овај процес се наставља све док генератор не буде у стању да произведе синтетичке податке који се не разликују од стварних података, у ком тренутку се сматра да је GAN истрениран. На Слици 24. је приказана генеративно супростављена мрежа.



Слика 24: Генеративно супростављена мрежа[74]

Једном истрениран, генератор се може користити за производњу синтетичких опсервација података који су слични подацима за тренирање. GAN-ови имају широк спектар примена, укључујући генерисање слика и видео записа, откривање аномалија, па чак и дизајнирање видео игара.

Једна од главних предности GAN-a је та што су у стању да генеришу синтетичке податке који су веома реалистични, што их чини моћним алатом за повећање података и генерисање синтетичких података за тренирање других модела машинског учења. У индустрији забаве и медија, GAN-ови су променили приступ приликом креирања садржаја. Користе се за генерисање реалистичних слика, видео записа и музике, нудећи нове алате за уметнике и креаторе садржаја. На пример, у филмској продукцији, GAN-

ови могу створити реалистичне специјалне ефекте или чак генерисати потпуно нове сцене или ликове. У индустрији игара, GAN-ови се користе за генерисање реалистичних текстура и пејзажа, побољшавајући визуелно искуство играња видео игара. Такође могу да креирају различите и реалистичне моделе ликова који се не могу контролисати али доприносе богатству света игара. У аутомобилској индустрији, посебно у развоју аутономних возила, GAN-ови се користе за генерисање разноврсних и реалистичних саобраћајних сценарија у сврхе симулације. Ово је кључно за обуку и тестирање система аутономне вожње у контролисаном, али разноликом окружењу.

Међутим, GAN-ове може бити тешко обучити због контрадикторне природе процеса обуке, а такође могу бити осетљиви на изборе хиперпараметара и иницијализацију.

Постоји неколико ограничења или недостатака у коришћењу генеративних супростављених мрежа:

- Потешкоће у тренирању: GAN-ове може бити тешко истренирати и могу захтевати доста финог подешавања да би се натерали да ефикасно раде. То је због супротстављених циљева мреже генератора и дискриминатора, што може довести до нестабилности у процесу обуке.
- Недостатак интерпретабилности: GAN-ови нису баш интерпретативни, што значи да може бити тешко разумети како дају своја предвиђања или генеришу синтетичке податке. Због тога може бити изазовно разумети зашто можда не успевају или на који начин да се побољша њихов учинак.
- Осетљивост на хиперпараметре: GAN-ови су осетљиви на хиперпараметре који се користе, као што су стопа учења или величина подскупа евалуације градијента. Проналажење оптималних вредности за ове хиперпараметре може бити изазовно и може захтевати опсежна експериментисања.
- Ограничена применљивост: Иако су се GAN-ови показали као успешни у генерисању висококвалитетних синтетичких слика и других типова података, они нису добро прилагођени за све врсте задатака. На пример, они можда неће бити ефикасни у моделовању сложених, високодимензионалних података или за задатке који захтевају прецизну контролу над излазом.
- Правни и етички проблеми: GAN-ови имају потенцијал да генеришу синтетичке податке који се могу користити у злонамерне сврхе, као што је прављење лажних видео снимака или слика за пропаганду или превару. Такође постоји забринутост у вези са потенцијалном злоупотребом GAN-ова за стварање синтетичких података који би могли да се користе за дискриминацију одређених група људи.

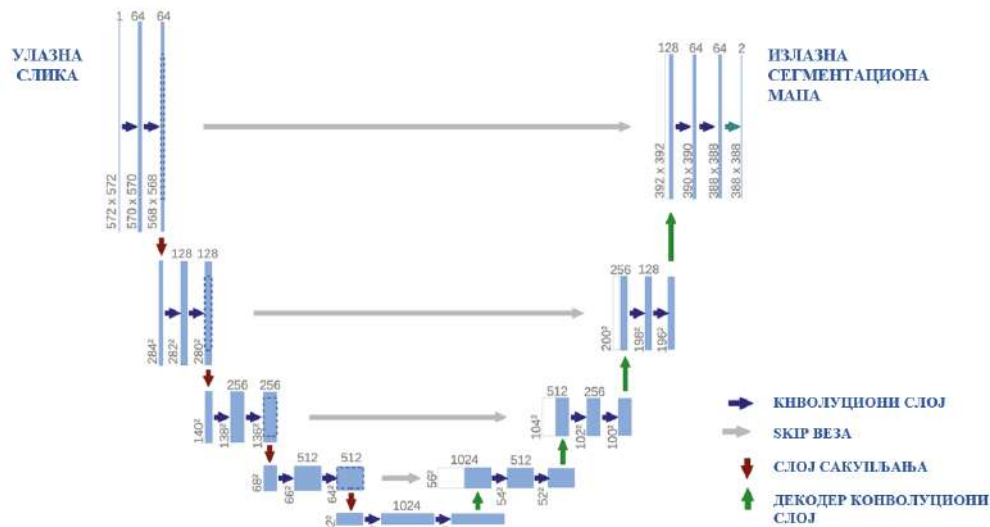
4.7 U-Net сегментациони модел

Сегментациони модели у машинском учењу су напредни алати за дубоко учење дизајнирани да интелигентно сецирају и интерпретирају дигиталне слике тако што их деле на смислене сегменте или групе пиксела. Ови модели су кључни у трансформацији сложених визуелних информација у разумљивији формат који се може анализирати.

Процес укључује класификацију на нивоу пиксела, где се сваки пиксел додељује одређеној категорији, на основу његових карактеристика као што су боја, интензитет и контекстуални однос са суседним пикселима. Наведена сегментација резултира детаљном мапом где је сваки појединачни сегмент корелиран са објектом или класом унутар слике. Ови модели имају широк спектар примена, од побољшања медицинске дијагностике преко прецизних анатомских слика до побољшања навигационих система у аутономним возилима.

У мноштву доступних сегментационих модела, U-Net се истиче, посебно по својим изванредним перформансама у сегментацији медицинске слике. U-Net архитектура, првобитно развијена за биомедицинске сврхе, је скројена да пружи и високу тачност и ефикасност, чак и са ограниченом доступношћу података што представља уобичајен изазов приликом анализе медицинских снимака.

На Слици 25. приказана је архитектура U-Net модела.



Слика 25: Архитектура U-Net сегментационог модела[75]

U-Net модел поседује јединствену структуру у облику слова У, која је у суштини низ два примарна пута: контракције и експанзивне путање. Пут контракције, такође познат као енкодер, састоји се од низа конволуционих слојева праћених слојевима скупљања. Овај део мреже је одговоран за хватање контекста унутар слике. Како подаци пролазе кроз ове слојеве, они се прогресивно дестилују у приказе обележја, истовремено смањујући димензионалност.

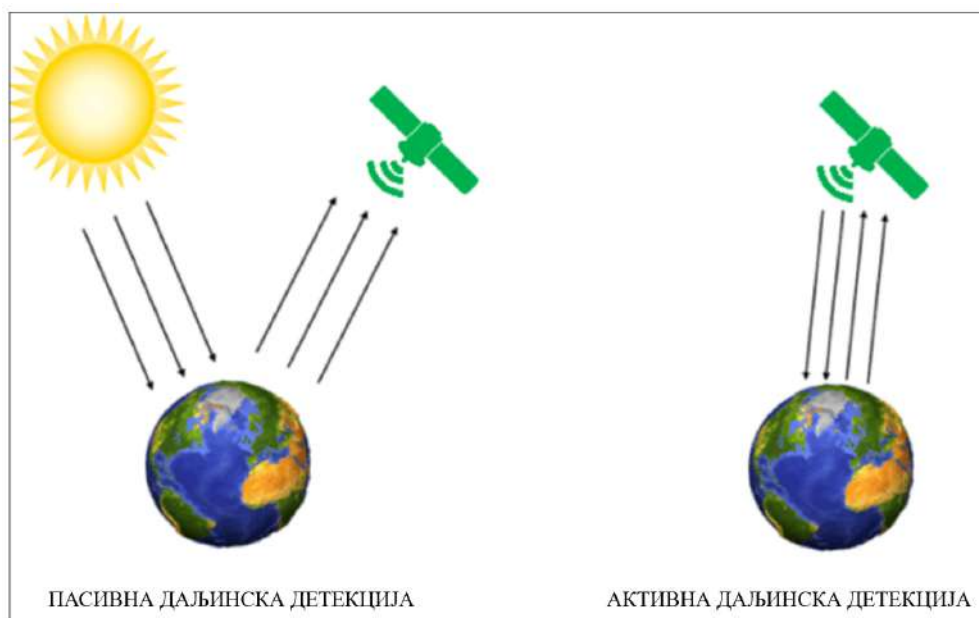
Супротно томе, експанзивна путања, или декодер, реконструише сегментирану слику из компресованих обележја добијених путем контракције. Ова реконструкција је олакшана транспонованим конволуцијским слојевима и повећањем узорковања, прогресивно враћајући просторне димензије излаза. Критична компонента архитектуре U-Net је инкорпорација такозваних прескачућих (енгл. skip) веза, које премашују одговарајуће слојеве у контракцијским и експанзивним путањама. Ове везе су инструменталне у прецизној локализацији сегмената слике, спајајући контекстуалне информације високог нивоа из дубљих слојева са детаљним, просторним информацијама из плићих слојева.

Способност U-Net модела да буде ефикасно трениран на релативно малим скуповима података а да и даље производи прецизне сегментације је једна од његових најзначајнијих предности. Наведена способност је посебно корисна у медицинским окружењима, где су велики скупови означених (лабелираних) података оскудни. Поред тога, док је U-Net првобитно био осмишљен за сегментацију биомедицинске слике, његова флексибилност и ефикасност довели су до успешних адаптација у разним другим областима које захтевају детаљну анализу слике.

5 ДАЉИНСКА ДЕТЕКЦИЈА

Даљинска детекција представља процес прикупљања информација о објекту или области са удаљености, коришћењем сензора који могу да открију различите врсте зрачења. Ови сензори се могу монтирати на сателите, авионе, беспилотне летелице (дронове) или друге типове платформи и могу детектовати различите врсте зрачења, укључујући видљиву светлост, инфрацрвено зрачење, као и микроталасно зрачење. Даљинска детекција нам омогућава да прикупљамо податке о Земљиној површини и атмосфери без да смо физички присутни на осматраној локацији.

Постоје два главна типа даљинске детекције: пасивна даљинска детекција и активна даљинска детекција. Пасивна даљинска детекција укључује детекцију природног зрачења, као што је видљиво светло или инфрацрвено зрачење. Ово се може урадити помоћу сензора на сателитима или авионима који детектују и мере интензитет зрачења које рефлектује или емитује Земљина површина. Активна даљинска детекција укључује емитовање сигнала и детекцију повратног сигнала након што се одбије од површине Земље. Примери активне даљинске детекције укључују радар и лидар.



Слика 26: Пасивна и активна даљинска детекција[76]

Даљинска детекција има дугу историју која сеже до раних дана фотографисања из ваздуха, која су се користила за снимање Земљине површине из ваздуха. Како је технологија напредовала и развијали се нови сензори који би могли да детектују зрачење изван видљивог спектра, поље даљинске детекције се проширило и укључило наведене нове методе прикупљања података. База развоја даљинске детекције такође укључује развој техника и софтвера за анализу слике, који нам омогућавају да извучемо корисне информације из података које прикупљају сензори даљинске детекције.

Сам термин даљинска детекција је скован да опише овај шири спектар метода прикупљања података. Евелин Пруит, научница која је радила у Канцеларији за поморска

истраживања америчке морнарице (енгл. U.S. Navy's Office of Naval Research), била је заслужна за ковање термина даљинска детекција 1960-их [48].

Евелин Пруит је препознала да термин фотоинтерпретација више није био довољан да опише посао који су она и њене колеге радиле са нефотографским сензорима. Ови сензори су били у стању да открију зрачење изван видљивог спектра, тако да термин фотоинтерпретација више није био тачан. Као резултат тога, Пруит је предложила употребу термина даљинска детекција за опис ове породице инструмената и техника које се користе за анализу података које су прикупили. Овај термин је постао широко прихваћен у овој области и користи се за описивање широког спектра сензора и метода прикупљања података који укључују прикупљање информација о објектима или областима са удаљености.

Лансирање сателита Landsat 1 1972. године била је велика прекретница у области даљинске детекције. Landsat 1 је био први сателит који је посебно дизајниран за посматрање копнених подручја Земље. Опремљен је разним сензорима који су могли да детектују зрачење у различитим деловима електромагнетног спектра, укључујући видљиву светлост, инфрацрвену и микроталасну. Ови сензори су омогућили Landsat 1 сателиту да прикупи детаљне податке о површини Земље, укључујући информације о коришћењу земљишта, земљишном покривачу и другим карактеристикама. Систематско, понављајуће посматрање Земљине површине које пружа Landsat 1 било је од непроцењиве вредности за широк спектар примена, укључујући мапирање, управљање ресурсима и праћење животне средине. У оквиру програма Landsat лансирано је још 8 сателита, закључно са Landsat 9 сателитом који је лансиран 2021 године.



Слика 27: Landsat 1 сателит

Програм Landsat дао је низ значајних доприноса у области даљинске детекције. Један од његових најважнијих доприноса био је да мултиспектрални сателитски подаци постану доступнији и приступачнији аналитичарима. Пре лансирања Landsat 1

сателита, мултиспектрални подаци су углавном били ограничени на специјализоване истраживачке лабораторије, али редовна доступност ових података кроз Landsat програм омогућила је много широј популацији аналитичара да има приступ прикупљеним подацима.

Још један важан допринос програма Landsat био је развој техника дигиталне анализе за даљинску детекцију. Доступност дигиталних података у стандардном формату створила је подстицај за развој софтвера за анализу слике, који се сада широко користе у применама даљинске детекције. Коначно, програм Landsat је послужио као модел за развој других сателита за посматрање земљишта којима управљају организације широм света.

Међу најпознатијима су NOAA сателити којима управља Национална управа за океане и атмосферу Сједињених Америчких Држава (енгл. National Oceanic and Atmospheric Administration - NOAA). NOAA сателити су лансирани 1978. године и представљају низ сателита у поларној орбити који су пружили критичне податке о времену и клими. Ови сателити су увелико побољшали нашу способност да предвидимо тешке временске прилике, као што су урагани и торнада.

Затим, SPOT (франц. Satellite Pour l'Observation de la Terre - SPOT) лансиран 1986. године. SPOT је француски сателит за даљинску детекцију који је први понудио мултиспектралне слике високе резолуције. Био је посебно користан за праћење промена у коришћењу земљишта и урбанизацији.

Сателити Terra и Aqua (шпан. Terra i Aqua) лансирани 1999. и 2002. године. Ови сателити су део NASA (енгл. National Aeronautics and Space Administration) система за посматрање Земље (енгл. Earth Observation System - EOS) и опремљени су скупом инструмената који прате различите аспекте Земљине површине и атмосфере. Посебно су корисни за праћење промена глобалне температуре, нивоа мора и вегетационог покривача.

Сателити из Sentinel серије Европске свемирске агенције (енгл. European Space Agency -ESA) који су део програма Коперник започетим лансирањем Sentinel-1A сателита 2014. године. Sentinel-1A, као и Sentinel-1B (лансиран 2016. године) обезбеђују радарско снимање Земљине површине, по свим временским приликама, дању и ноћу, за примене као што су поморски надзор, праћење деформација земљишта, мапирање морског леда и детекцију и праћење изливања нафте.

Sentinel-2A и Sentinel-2B (лансирани 2015. и 2017. године) су мултиспектрални сателити високе резолуције који обезбеђују снимке за примене као што су мапирање земљишног покривача, праћење усева и управљање катастрофама. Sentinel-3A и Sentinel-3B (лансирани 2016. и 2018. године) су сателити који носе више инструмената за мерење температуре површине мора, боје океана, нивоа мора и температуре површине копна. Sentinel-5P (лансиран 2017.) је сателит који мери низ атмосферских гасова и загађивача како би подржао праћење квалитета ваздуха и истраживање климатских промена.

Даљинска детекција има широк спектар примена, укључујући мапирање, праћење коришћења земљишта и земљишног покривача, проучавање климе на Земљи и проучавање природних ресурса Земље. То је мултидисциплинарно поље које укључује комбинацију физике, информатике и географије, између осталих дисциплина.

5.1 Спектар електромагнетног зрачења

Сви објекти емитују електромагнетно зрачење, али објекти такође рефлектују зрачење које су емитовали други објекти. Даљинска детекција укључује откривање и мерење овог зрачења, које може пружити вредне информације о објектима или областима које се проучавају. Разумевањем како се електромагнетно зрачење понаша док пролази кроз Земљину атмосферу и ступа у интеракцију са објектима на површини, истраживачи из области даљинске детекције су у стању да извуку корисне информације о карактеристикама ових објеката. Истраживачи из области даљинске детекције развијају знања о карактеристикама Земље као што су вегетација, структуре, земљиште, стене или водене површине.

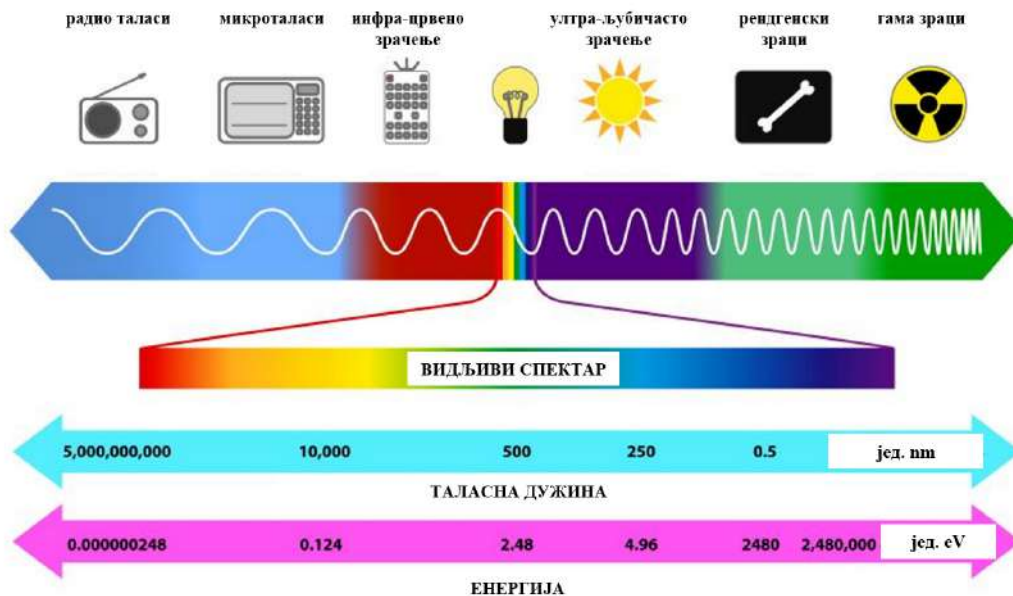
Електромагнетно зрачење се производи кроз различите механизме, укључујући промене у енергетским нивоима електрона, убрзање наелектрисаних честица, распад радиоактивних изотопа и топлотно кретање атома и молекула. Сунце, кроз нуклеарне реакције унутар свог језгра, емитује читав спектар електромагнетног зрачења које путује кроз вакуум свемира уз минималне промене. Када ово зрачење стигне до Земље, оно пролази кроз атмосферу и може се или рефлектовати од површине Земље или апсорбовати и поново зрачити као топлотна енергија. Даљински детектоване слике могу се добити како из рефлектованог тако и из термалног зрачења, као и из извора које генерише човек, као што су радари за снимање.

Електромагнетно зрачење има низ својстава која су важна у области даљинске детекције. Неке од најважнијих својстава су:

- Таласна дужина: Таласна дужина електромагнетног зрачења представља растојање између узастопних врхова електромагнетног таласа. Обично се мери у метрима, нанометрима или микрометрима. Што је таласна дужина краћа, то је већа енергија зрачења.
- Фреквенција: Фреквенција електромагнетног зрачења представља број врхова електромагнетног таласа који прођу фиксiranу тачку у одређеном времену. Обично се мери у херцима (Hz). Што је фреквенција већа, то је већа енергија зрачења.
- Амплитуда: Амплитуда је еквивалентна висини сваког врха електромагнетног таласа и представља меру његове снаге или интензитета. Обично се мери у ватима по квадратном метру.

Електромагнетни спектар се односи на цео опсег електромагнетног зрачења, који укључује радио таласе, микроталасе, инфрацрвено зрачење, видљиву светлост, ултраљубичасто зрачење, рендгенске зраке и гама зраке.

Радио таласи имају најдуже таласне дужине и најмању енергију, док гама зраци имају најкраћу таласну дужину и највећу енергију. Део електромагнетног спектра који можемо да видимо очима назива се видљива светлост, у коју спадају дугине боје: црвена, наранџаста, жута, зелена, плава, индиго и љубичаста. Инфрацрвено зрачење има таласне дужине дуже од таласних дужина видљиве светлости и емитују га топли објекти, као што су на пример Сунце или камин. Микроталаси имају таласне дужине између радио таласа и инфрацрвеног зрачења и користе се у области комуникација и радара.



Слика 28: Електромагнетни спектар

Ултраљубичасто зрачење има таласне дужине краће од таласне дужине видљиве светлости и одговорно је за опекотине од сунца и рак коже. X зраци имају краће таласне дужине од ултраљубичастиг зрачења и користе се за прављење слика унутрашњости тела (рендгенски снимак). Гама зраци имају најкраћу таласну дужину и највећу енергију од свих врста електромагнетног зрачења и користе се у лечењу рака и за прављење слика унутрашњости чврстих објеката.

Као што је већ наведено, видљива светлост представља део електромагнетног спектра који можемо да видимо и чини само мали део пуног опсега електромагнетног зрачења. Даљински сензори могу бити дизајнирани да детектују зрачење у различитим деловима електромагнетног спектра, што нам омогућава да прикупимо информације о објектима и областима које нису видљиве људском оку. Проучавајући ове различите делове електромагнетног спектра, можемо сазнати о карактеристикама и својствима објеката и области на начине који не би били могући, ако бисмо користили сазнања добијена од сензора који детектују само видљиву светлост.

5.2 Интеракција са Земљином атмосфером (расипање, преламање, апсорпција)

Електромагнетно зрачење које се користи за даљинску детекцију мора да прође кроз Земљину атмосферу, што може утицати на квалитет слика и података које прикупљају сензори. Када су сензори за даљинску детекцију постављени на авионе који лете ниско, атмосферски ефекти могу бити минимални, али када су сензори постављени на сателите, зрачење мора проћи кроз целу дубину атмосфере, што може имати већи утицај на квалитет слике. Разумевање интеракције електромагнетне енергије са атмосфером је стога важан аспект праксе даљинске детекције. Како у урбаним тако и у руралним

подручјима, атмосфера може да модификује сунчево зрачење расипањем, преламањем и апсорпцијом. Наведени процеси могу утицати на осветљеност и боју зрачења које посматрамо, а такође могу утицати на квалитет слика и података добијених даљинском детекцијом.

Расипање, у области даљинске детекције, се односи на процес којим се електромагнетно зрачење одбија од честица или молекула у атмосфери. Расипање је селективан процес, што значи да се различите таласне дужине зрачења расипају у различитим степенима. Количина расипања која се јавља зависи од величине честица или молекула, њихове концентрације, таласне дужине зрачења и дубине атмосфере кроз коју енергија пролази. Расипање узрокује да се део долазног сунчевог зрачења рефлектује назад у свемир, а део да се усмери ка површини Земље.

Неки примери расипања укључују:

- Плаво небо: Када се сунчева светлост расипа од стране Земљине атмосфере, плава светлост се расипа више од других боја, због чега небо изгледа плаво.
- Црвени заласци сунца: Када је Сунце ниско на небу, његова светлост мора да прође кроз већи део Земљине атмосфере, што узрокује да се краће таласне дужине плаве светлости расипају више од дужих таласних дужина црвене светлости. Због тога сунце изгледа црвено или наранџасто при изласку и заласку сунца.
- Измаглица: Измаглицу изазивају ситне честице у атмосфери које расипају светлост и смањују видљивост. Измаглицу могу проузроковати различити извори, укључујући загађење, прашину и дим.
- Магла: Магла представља временски феномен који настаје када се влажност ваздуха у близини тла повећа до тачке у којој ваздух постаје засићен воденом паром. Капљице воде у магли расипају светлост, због чега је тешко видети кроз густу маглу.
- Дуга: Дуга је узрокована расипањем сунчеве светлости од стране кишних капи. Различите боје дуге су узроковане различитим таласним дужинама светлости која се распршује под различитим угловима.

Расипање може имати бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Смањени контраст: Расипање може да смањи контраст на сликама добијеним даљинском детекцијом тако што узрокује да се део зрачења рефлектује назад у свемир уместо да га апсорбује површина. Ово може отежати разликовање различитих типова карактеристика или материјала.
- Варијације у осветљености: Расипање може да изазове варијације у осветљености на сликама добијеним даљинском детекцијом, посебно у областима са високом концентрацијом честица или молекула. Ове варијације могу отежати прецизно тумачење слика.

- Промене спектралних карактеристика: Расипање може да промени спектралне карактеристике података добијених даљинском детекцијом, што може утицати на тачност информација које се добијају из података.
- Повећана атмосферска бука: Расипање може повећати ниво буке у подацима добијених даљинском детекцијом, што може смањити укупан квалитет података.

Преламање у области даљинске детекције, односи се на савијање електромагнетног зрачења док оно пролази кроз медијум са различитим индексом преламања. Индекс преламања медијума представља меру колико се брзина светлости успорава док пролази кроз одређени медијум. Када електромагнетно зрачење прелази из једног медија у други са различитим индексом преламања, оно ће се савијати или преламати.

Неки примери преламања укључују:

- Фатаморгана: фатаморгана је оптичка илузија узрокована преламањем светлости док она пролази кроз слојеве ваздуха са различитим индексима преламања. На пример, фатаморгана воде на врелом путу је узрокована преламањем светлости док она пролази кроз топлији ваздух изнад пута и хладнији ваздух изнад тла.
- Сноп светионика: Сноп светлости из светионика се прелама док пролази кроз Земљину атмосферу, због чега изгледа да се „савија“ док путује кроз ваздух. Због тога се сноп светионика може видети из даљине чак и када је усмерен право нагоре.
- Изобличење слике у телескопима: Преламање може проузроковати да слике које се виде кроз телескопе изгледају изобличене, посебно ако телескоп није дизајниран да исправи овај ефекат.
- Изобличење слике у камерама: Преламање такође може проузроковати да слике које се виде кроз камере изгледају изобличене, посебно ако камера није дизајнирана да исправи овај ефекат. Због тога неке камере имају сочива са ниским индексом преламања, што смањује количину преламања која се јавља.
- Преламање светлости кроз воду: Када светлост прође кроз воду, она се прелама јер је индекс преламања воде другачији од индекса преламања ваздуха. Због тога се чини да су објекти изобличени када се гледају кроз воду.

Преламање може имати бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Изобличење слике: Преламање може проузроковати да слике изгледају изобличене или искривљене, посебно када је сензор на великој висини и зрачење мора да прође кроз велику дубину атмосфере. Ово може отежати прецизно тумачење слика.
- Нетачна мерења: Преламање може утицати на тачност мерења направљених коришћењем података добијених даљинском детекцијом, посебно ако се преламање не узме у обзир. На пример, ако преламање узрокује да се зрачење савије даље од површине, мерења карактеристика површине могу бити потцењена.

- Промене спектралних карактеристика: Преламање може да промени спектралне карактеристике података добијеним даљинском детекцијом, што може утицати на тачност информација које се издвајају из података.

Апсорпција, у области даљинске детекције, односи се на процес којим материјал апсорбује електромагнетно зрачење док пролази кроз материјал. Апсорпција се дешава када енергију зрачења апсорбују атоми или молекули материјала, узрокујући да они вибрирају или ротирају. Апсорпција је селективан процес, што значи да различити материјали апсорбују различите таласне дужине зрачења у различитим степенима.

Неки од примера апсорпције укључују:

- Крема за сунчање: Крема за сунчање апсорбује одређене таласне дужине ултраљубичастиг зрачења како би заштитила кожу од штетног дејства сунчевих зрака.
- Филтери у камерама: Филтери у камерама апсорбују одређене таласне дужине светлости да би побољшали контраст или боју слика.
- Апсорпција топлоте тамним површинама: Тамне површине апсорбују више топлоте од светлих јер апсорбују више енергије од сунчевих зрака. Због тога је често хладније у сенци него на директној сунчевој светлости.
- Апсорпција светлости пигментима: Пигменти у биљкама апсорбују одређене таласне дужине светлости да би фотосинтетизовале храну из угљен-диоксида и воде. Различити пигменти апсорбују различите таласне дужине светлости, због чега листови имају различите боје.

Апсорпција може имати бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Промене у осветљености: Апсорпција може проузроковати да слике изгледају тамније или светлије, у зависности од количине апсорпције која се јавља. Ово може отежати прецизно тумачење слика.
- Нетачна мерења: апсорпција може утицати на тачност мерења направљених коришћењем података добијеним даљинском детекцијом, посебно ако се апсорпција не узме у обзир. На пример, ако је апсорпција већа на одређеним таласним дужинама, мерења површинских карактеристика на тим таласним дужинама могу бити потцењена.
- Промене спектралних карактеристика: Апсорпција може да промени спектралне карактеристике на подацима добијеним даљинском детекцијом, што може утицати на тачност информација које се издвајају из података.

Како би кориговали ефекте расипања, преламања и апсорпције, истраживачи из области даљинске детекције користе различите технике, као што су на пример алгоритми за корекцију атмосфере или алгоритми геометријске корекције. Разумевање механизма расипања, преламања и апсорпције, као и њихових ефеката на податке добијене даљинском детекцијом представљају важан аспект праксе даљинске детекције.



Слика 29: Графички приказ расипања, преламања и апсорпције

5.3 Атмосферски прозори

Атмосферски прозори су области електромагнетног спектра где је атмосфера релативно транспарентна, што значи да не апсорбује или расипа значајну количину зрачења у тим таласним дужинама. Атмосферски прозори омогућавају зрачењу у тим таласним дужинама да релативно неометано пролази кроз атмосферу, што омогућава даљинску детекцију површине Земље у тим таласним дужинама. Постоји неколико атмосферских прозора у електромагнетном спектру, укључујући видљиве, блиске инфрацрвене и термалне инфрацрвене прозоре.

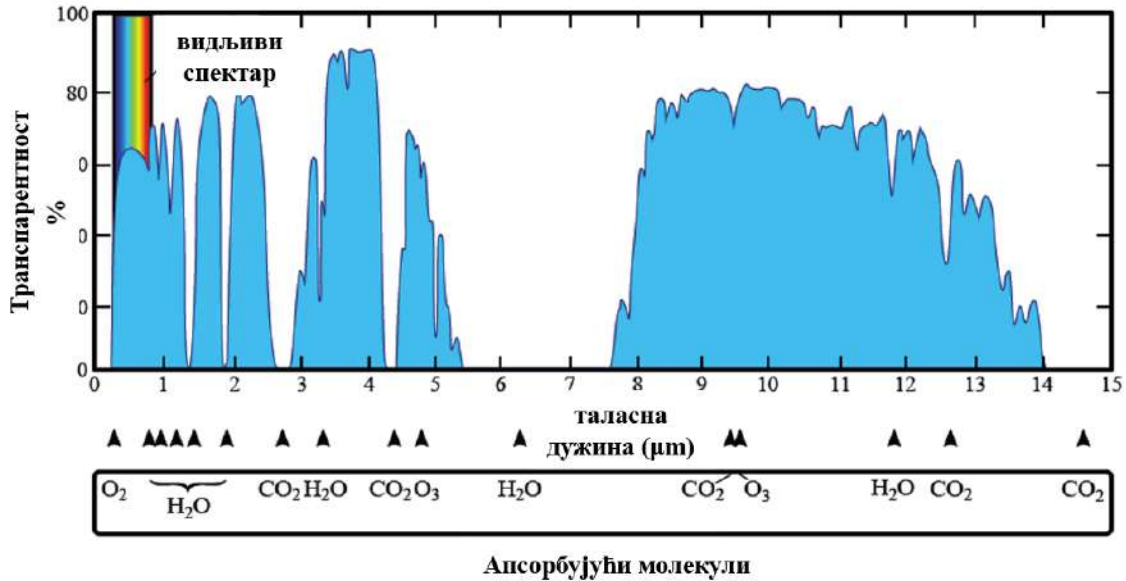
Видљиви прозор је део електромагнетног спектра који је видљив људском оку и креће се од око 400 до 700 нанометара таласне дужине. Атмосфера је релативно транспарентна у овом прозору, што омогућава даљинску детекцију површине Земље помоћу видљиве светлости. Овај прозор се користи за визуелну интерпретацију слика, као и за идентификацију специфичних типова површина и материјала на основу њихове боје и својстава рефлексије.

Блиско инфрацрвени прозор је део електромагнетног спектра одмах иза видљивог прозора и креће се од око 700 до 2500 нанометара таласне дужине. Атмосфера је релативно транспарентна у овом прозору, што омогућава даљинску детекцију површине Земље коришћењем блиско инфрацрвеног зрачења. Овај прозор се користи за проучавање вегетације, јер биљке снажно рефлектују блиско инфрацрвено зрачење, а такође се користи за идентификацију специфичних типова површина и материјала на основу њихових рефлектујућих својстава.

Термални инфрацрвени прозор је део електромагнетног спектра који објекти емитују као топлотно зрачење и креће се од око 8 до 14 микрометара таласне дужине. Атмосфера је релативно транспарентна у овом прозору, што омогућава даљинску детекцију температуре површине Земље коришћењем топлотног инфрацрвеног зрачења. Овај прозор се користи за мерење температуре, као и за идентификацију специфичних типова површина и материјала на основу њихових својстава топлотне емисије.

Атмосферски прозори имају бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Побољшан квалитет слике: Пошто је атмосфера релативно транспарентна у атмосферским прозорима, на зрачење у тим таласним дужинама мање утичу атмо-



Слика 30: Атмосферски прозори[48]

сферско расипање и апсорпција, што резултира побољшаним квалитетом слике.

- Побољшана спектрална резолуција: Атмосферски прозори омогућавају детекцију специфичних таласних дужина зрачења, које се могу користити за побољшање спектралне резолуције података добијеним даљинском детекцијом.
- Повећана осетљивост: Транспарентност атмосфере у атмосферским прозорима омогућава детекцију слабијих сигнала, што резултира повећаном осетљивошћу података добијеним даљинском детекцијом.
- Побољшана прецизност: Недостатак атмосферских сметњи у атмосферским прозорима омогућава прецизнија мерења површинских карактеристика које се детектују.

Сваки атмосферски прозор има различите карактеристике и користи се за различите врсте примена у домену даљинске детекције. На пример, видљиви прозор се користи за визуелну интерпретацију слика, блиско инфрацрвени прозор се користи за проучавање вегетације, а термални инфрацрвени прозор се користи за детекцију температуре. Разумевање карактеристика и примене атмосферских прозора је важан аспект праксе даљинске детекције.

5.4 Интеракција са Земљином површином (рефлексија, трансмисија, дифракција)

Када електромагнетна енергија стигне до површине Земље, може се рефлектовати, апсорбовати или трансмитовати. Пропорције које се узимају у обзир за сваки процес зависе од природе површине, таласне дужине енергије и угла осветљења.

Рефлексија електромагнетног зрачења у даљинској детекцији односи се на процес којим се електромагнетна енергија рефлектује назад у атмосферу након што стигне до површине Земље. Количина енергије која се рефлектује зависи од рефлективних карактеристика површине, таласне дужине енергије и угла осветљења. Површине које имају високу рефлексију, као што су снег или вода, рефлектоваће већи удео долазне енергије, док ће површине које су мање рефлективне, као што су вегетација или земљиште, рефлектовати мањи део. Рефлективна својства површине одређују састав површине, храпавост и структура и могу значајно да варирају у различитим типовима површина. Рефлексија електромагнетног зрачења представља важан фактор у даљинској детекцији, јер одређује количину енергије која се враћа сензору и може се користити за извођење закључака о својствима површине која се детектује.

Рефлексија електромагнетног зрачења има бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Осветљеност слике: Рефлективна својства површине имају директан утицај на осветљеност слике добијене даљинском детекцијом. Површине које имају високу рефлексију, као што су снег или вода, изгледаће светлије на слици, док ће површине које имају мање рефлексије, као што су вегетација или земља, изгледати тамније.
- Контраст слике: Рефлективна својства површине такође утичу на контраст слике добијене даљинском детекцијом. На пример, ако је рефлексија једне површине много већа од рефлексије друге површине, контраст између две површине ће бити већи.
- Спектралне информације: Рефлективна својства површине у различитим таласним дужинама могу се користити за извођење закључака о својствима површине која се детектује. На пример, рефлексија вегетације у видљивој и блиској инфрацрвеној таласној дужини се често користи за процену здравља и биомасе биљака. Слично томе, рефлексија земљишта у видљивим и блиско инфрацрвеним таласним дужинама се често користи за процену садржаја воде у земљишту и плодности.
- Топографија површине: На рефлексију електромагнетног зрачења такође може утицати топографија површине, као што су нагиб и оријентација површине. Ово се може користити за извођење закључака о топографским карактеристикама површине, као што су постојање брда и долина.

Трансмисија електромагнетног зрачења у даљинској детекцији се односи на процес којим електромагнетна енергија пролази кроз површину Земље. Количина енергије која се преноси зависи од трансмисионих својстава површине, таласне дужине енергије и угла осветљења. Површине које су прозирне, као што су стакло или вода, ће преносити већи део долазне енергије, док ће површине које су непрозирне, као што су чврсти камен или метал, преносити мањи део. Трансмисиона својства површине су одређена саставом површине, дебљином и структуром и могу значајно да варирају у зависности од различитих типова површина. Трансмисија електромагнетног зрачења је важан

фактор у даљинској детекцији, јер одређује количину енергије која може да прође кроз површину и стигне до сензора на другој страни.

Трансмисија електромагнетног зрачења има бројне ефекте на слике и податке добијене даљинском детекцијом. Неки од најзначајнијих ефеката укључују:

- Осветљеност слике: Трансмисиона својства површине имају директан утицај на осветљеност слике добијене даљинском детекцијом. Површине које су прозирне, као што су вода или танка вегетација, омогућиће да прође више енергије, што ће резултирати светлијом сликом, док ће површине које су непрозирне, као што су чврсти камен или метал, преносити мање енергије и резултирати тамнијом сликом.
- Контраст слике: Трансмисиона својства површине такође утичу на контраст слике добијене даљинском детекцијом. На пример, ако је површина на једној страни објекта провидна, а површина на другој страни непрозирна, постојаће велики контраст између две површине на слици добијеном даљинском детекцијом.
- Спектралне информације: Трансмисиона својства површине у различитим таласним дужинама могу се користити за извођење закључака о својствима и саставу површине која се детектује. На пример, пренос електромагнетног зрачења у видљивој и блиској-инфрацрвеној таласној дужини може се користити за процену дебљине ледених покривача, снежног покривача и вегетацијских крошњи.
- Информације о дубини: Пренос електромагнетног зрачења се такође може користити за процену дубине објеката на Земљиној површини, као што су водене површине или вегетацијске крошње. Ово се може урадити мерењем преноса енергије на различитим таласним дужинама или угловима, и коришћењем ових информација за процену дубине објеката.

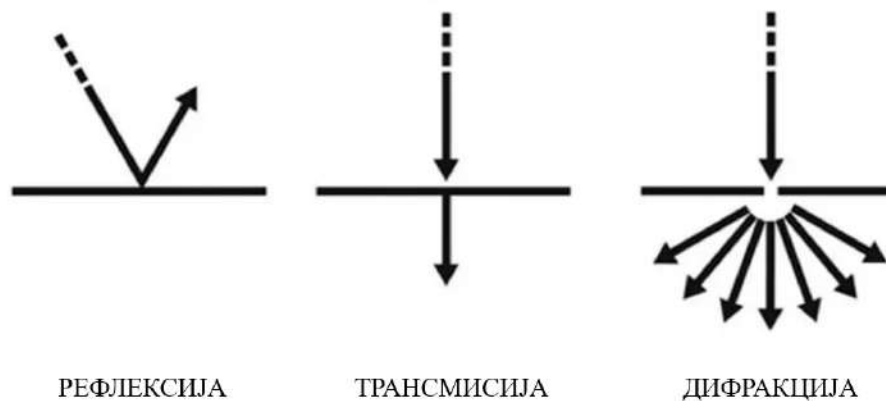
Дифракција електромагнетног зрачења се односи на фундаментални феномен где се електромагнетна енергија, као што су светлост, радио таласи или рендгенски зраци, шири док наилази на препреке или пролази кроз отворе упоредиве по величини са таласном дужином зрачења. Овај процес доводи до стварања карактеристичних образаца ширења и интерференције.

Када електромагнетни таласи наиђу на препреку као што је ивица објекта или уски прорез, они се не шире само право напред; уместо тога, савијају се око ивица и шире се, формирајући сложене шаре. Ови обрасци зависе од величине препреке или отвора у односу на таласну дужину електромагнетног зрачења. Дифракција је посебно изражена када је препрека или отвор приближно исте величине као таласна дужина електромагнетних таласа.

Дифракција електромагнетног зрачења има неколико ефеката на слике и податке добијене даљинском детекцијом:

- Резолуција слике: Дифракција може утицати на оштрину и резолуцију слике. На пример, у оптичким телескопима, дифракција ограничава максималну резолуцију коју постиже телескоп.

- Осветљеност слике: Дифракција електромагнетних таласа утиче на осветљеност слика добијених даљинском детекцијом. Дифракцијски таласи могу изазвати варијације у интензитету светлости коју прима сензор, што доводи до разлика у осветљености слике. На пример, области у којима се таласи мање дифрактирају могу изгледати светлије, док области са већом дифракцијом могу изгледати тамније.
- Контраст слике: На контраст у сликама такође утиче дифракција. Ово је посебно евидентно када се дифракцијски таласи са различитих површина различито интерферишу. На пример, површине које изазивају већу дифракцију могу изгледати мање јасне за разлику од оних које изазивају мању дифракцију, што утиче на укупан контраст и јасноћу слике.
- Спектралне информације: Својства дифракције површина на различитим таласним дужинама могу пружити вредне спектралне информације. У даљинској детекцији, анализа начина на који различите површине дифрагују електромагнетне таласе на различитим таласним дужинама може дати увид у својства тих површина. На пример, обрасци дифракције у различитим спектралним опсезима могу се анализирати да би се проценило здравље вегетације или својства земљишта.
- Топографија терена: Топографија терена може значајно утицати на дифракцију електромагнетних таласа. Карактеристике као што су брда, долине и оријентација површина могу променити начин на који се таласи дифрактују, што се може користити у даљинској детекцији да би се извукли закључци о топографским карактеристикама терена, као што је присуство нагиба или таласа.



Слика 31: Графички приказ рефлексије, трансмисије и дифракције

Разумевање ефеката рефлексије, трансмисије и дифракције на податке добијене даљинском детекцијом представљају важан аспект праксе даљинске детекције, јер омогућава истраживачима да прецизно тумаче својства и процесе материјала који се детектују.

5.5 Платформе за даљинску детекцију

Платформе за даљинску детекцију су возила која носе сензоре који се користе за прикупљање електромагнетног зрачења у циљу генерисања слика Земљине површине. Платформе за даљинску детекцију пружају увид у Земљину површину из перспективе одгоре и продукују слике налик на мапе. Избор платформе представља важну одлуку у даљинској детекцији јер утиче на избор сензора и може утицати на целокупан изглед стратегије за одређени пројекат даљинске детекције. Платформе се разликују по величини, домету, управљивости, цени и висинском домету, а свака има своје предности и недостатке у пракси даљинске детекције. Платформе за даљинску детекцију могу бити са земље, из ваздуха или из свемира, у зависности од специфичних захтева примене даљинске детекције.

Платформе за даљинску детекцију са земље су стационарне или мобилне платформе које се користе за прикупљање електромагнетног зрачења са тла или са мале надморске висине. Ове платформе укључују сензоре монтиране на возила, торњеве или друге структуре, као и ручне или преносиве инструменте.



Слика 32: Слика фиксног земаљског сензора за даљинску детекцију

Тип даљинске детекције са земље има неколико предности, укључујући ниску цену, флексибилност и могућност прикупљања података високе резолуције. Посебно је погодан за прикупљање података на сложеном терену или у областима где је приступ ограничен. Међутим, платформе за даљинску детекцију са земље су такође ограничене у величини области које могу да покрију и на њих може утицати присуство дрвећа, зграда или других препрека.

Примери платформи за даљинску детекцију са земље укључују:

- Сензори постављени на возила која се крећу по тлу: Ово укључује сензоре монтиране на аутомобиле, камионе или друга возила који се користе за прикупљање података док се возило креће.
- Фиксни земаљски сензори: Ово укључује сензоре монтиране на торњеве или друге структуре које се користе за прикупљање података са фиксне локације.
- Ручни или преносиви инструменти: Ово укључује инструменте које носи оператер и који се користе за прикупљање података на терену. Примери укључују спектрометре, лидарске скенере и мултиспектралне уређаје за снимање.

Платформе за даљинску детекцију из ваздуха су возила која се користе за прикупљање електромагнетног зрачења са површине Земље у сврху даљинске детекције. Платформе за даљинску детекцију из ваздуха раде у Земљиној атмосфери и укључују авионе са фиксираним крилима, хеликоптере и беспилотне ваздушне системе (енгл. Unmanned Aerial Vehicles - UAV). Платформе за даљинску детекцију из ваздуха обезбеђују средства за контролу надморске висине, оријентације, путање и времена неопходних за добијање корисних слика снимљених на даљину. Обично се користе за мапирање великих површина, прикупљање података на сложеном терену и прикупљање података у тешким или опасним окружењима.

Примери платформи за даљинску детекцију из ваздуха укључују:

- Ваздухоплови са фиксираним крилима: То су летелице које су опремљене сензорима и које се користе у сврху даљинске детекције. Ваздухоплови са фиксираним крилима се обично користе за мапирање великих површина и способни су да покрију велике удаљености за кратко време. Могу да лете на великим висинама, пружајући широко видно поље, али имају ограничену могућност прилагођавања путање лета и прикупљања података на малим висинама.
- Хеликоптери: То су летелице са ротационим крилима које су опремљене сензорима и које се користе у сврху даљинске детекције. Хеликоптери су лакши за маневрисање од ваздухоплова са фиксним крилима и способни су да лете на малим висинама, што их чини корисним за прикупљање података на сложеним теренима или урбаним подручјима. Међутим, они су ограничени у удаљености коју могу да пређу и скупљи су за рад од ваздухоплова са фиксним крилима.
- Беспилотне ваздушне системе (дронов): То су летелице којима се управља даљински или аутономно и опремљене су сензорима за потребе даљинске детекције. То су мале, лагане летелице које нуде могућност летења на малим висинама и имају велики маневарски капацитет, што их чини корисним за прикупљање података у тешким или опасним окружењима. Међутим, они имају ограничен домет и издржљивост у поређењу са другим платформама за даљинску детекцију из ваздуха.

Платформе за даљинску детекцију из свемира односе се на летелице које се лансирају у свемир и носе сензоре који се користе за прикупљање електромагнетног зрачења



Слика 33: Слика беспилотног ваздушног система за даљинску детекцију[48]

како би се формирале слике Земљине површине. Платформе за даљинску детекцију из свемира укључују сателите, који су у стању да круже око Земље и снимају слике са велике висине. Имају способност пружања глобалне перспективе, јер могу да покрију велике области Земљине површине у једној слици. Такође имају могућност снимања слика из широког опсега таласних дужина и могућност редовног снимања слика, што омогућава праћење промена током времена. Скупљају податке континуирано, без обзира на временске услове или доба дана. Такође, могу да покрију области које су тешке или немогуће да се снимају путем осталих платформи за даљинску детекцију. Међутим, ове платформе имају нека ограничења, укључујући нижу просторну резолуцију у поређењу са неким ваздушним платформама и потребу за специјализованом опремом и инфраструктуром за пријем и обраду података.

Примери платформи за даљинску детекцију из свемира укључују:

- Landsat: Ова серија сателита, којом управља Геолошки завод САД (енгл. US Geological Survey), прикупља мултиспектралне слике Земљине површине од 1972. године. Landsat сателити имају просторну резолуцију од 30 метара и прикупљају податке у неколико таласних дужина, укључујући видљиву и инфрацрвену светлост.
- Sentinel: Ова серија сателита, којима управља Европска свемирска агенција, прикупља слике високе резолуције за различите примене, укључујући мапирање земљишног покривача и праћење природних катастрофа. Sentinel сателити имају просторну резолуцију од 10 метара или више.
- MODIS: Спектрорадиометар за снимање умерене резолуције (енгл. Moderate Resolution Imaging Spectroradiometer - MODIS) је сензор који је постављен на два NASA сателита, Terra и Aqua. MODIS прикупља податке у 36 спектралних опсега, укључујући видљиву и инфрацрвену светлост, са просторном резолуцијом од 250 метара.

MODIS подаци се користе за разне примене, укључујући праћење вегетације и температуре површине земљишта



Слика 34: Слика Sentinel 2 сателита за даљинску детекцију[77]

5.6 Sentinel 2 мисија

Сателитске мисије Sentinel, које чине саставни део програма Коперник, представљају свеобухватну иницијативу за посматрање Земље коју је развила Европска свемирска агенција у сарадњи са Европском комисијом. Свака Sentinel мисија је заснована на констелацији од два сателита, нудећи робусне скупове података за праћење животне средине и за студије климатских промена.

Међу њима, мисија Sentinel-2 је посебно прилагођена за праћење земљишта. Пружа оптичке слике високе резолуције неопходне за праћење вегетације, земљишног и воденог покривача, унутрашњих пловних путева и обалних подручја. Мисија Sentinel-2, која је кључна за управљање земљиштем и праћење пољопривреде, састоји се од два идентична сателита, Sentinel-2A и Sentinel-2B. Лансирани у јуну 2015. и марту 2017. године, они раде у тандему, покривајући копнене површине Земље, велика острва, унутрашње и приобалне воде сваких пет дана на екватору и чешће на средњим географским ширинама. Ово обезбеђује високу учесталост поновних посета и континуитет података, што је кључно за праћење животне средине.

Сваки сателит Sentinel-2 опремљен је најсавременијим мултиспектралним инструментом (енгл. Multi Spectral Instrument - MSI), способним да снима слике у 13 спектралних опсега. Ови опсеги се крећу у видљивом, блиском инфрацрвеном и краткоталасном инфрацрвеном делу спектра, са просторним резолуцијама од 10м, 20м и 60м. Ова спектрална разноликост је кључна за детаљно мапирање земљишног покривача, праћење вегетације и надзор шума. Опсеги укључују четири појаса на 10м (видљиви и блиски инфрацрвени), шест опсега на 20м (црвена ивица, краткоталасни инфрацрвени) и три опсега на 60м (аеросол, водена пара, цирус).

Примене Sentinel-2 су огромне и разноврсне. Оне обухватају праћење промена земљишта, процену пољопривредних пракси, праћење стања шума, управљање природним катастрофама и студије климатских промена. Висока резолуција и мултиспектрални подаци мисије су инструментални у одрживом управљању земљишним ресурсима.

Подаци прикупљени од стране Sentinel-2 сателита су слободно доступни и служе истраживачима, комерцијалним субјектима и креаторима политика. Доступни су преко платформи као што је Copernicus Open Access Hub. Штавише, подаци Sentinel-2 често допуњују податке из других Sentinel мисија, пружајући свеобухватну перспективу праћења животне средине.

Технолошки, Sentinel-2 означава значајан скок у посматрању Земље. Висока резолуција, мултиспектралне могућности мисије, заједно са њеним честим временом поновних посета, пружају детаље без преседана у посматрању Земље, што је неопходно за прецизно праћење животне средине и управљање природним ресурсима. Мисија Sentinel-2 је остварила дубок утицај не само на научну заједницу, већ и на креирање политика и на практичне примене у пољопривреди, реаговању у случају катастрофе и очувању животне средине.

5.7 Индекси даљинске детекције

Индекси даљинске детекције, изведени из сателитских снимака, кључни су за праћење животне средине. Они омогућавају префињен и ефикасан приступ анализи Земљине површине. Индекси даљинске детекције, израчунати кроз специфичне комбинације спектралних опсега, омогућавају екстракцију циљаних информација из сложених података које су ухватили сателити за посматрање Земље као што је Sentinel-2. Сваки индекс је дизајниран да истакне различите карактеристике површине, што их чини вредним ресурсом у различитим областима укључујући пољопривреду, урбано планирање и управљање животном средином.

Индекс нормализоване разлике вегетације (енгл. Normalized Difference Vegetation Index - NDVI) је основни индекс у даљинској детекцији, који се првенствено користи за процену здравља и густине вегетације.

Израчунава се помоћу формуле:

$$NDVI = \frac{NIR - \text{црвени}}{NIR + \text{црвени}}$$

где NIR и црвени представљају блиски инфрацрвени и црвени опсег (опсег 8 и опсег 4, респективно, код Sentinel-2 сателита). Вредности NDVI се крећу од - 1 до 1, са вишим вредностима које указују на гушћу зелену вегетацију. Овај индекс је критичан у процени SOC-а јер помаже у идентификацији подручја са снажном вегетацијом, која су често повезана са вишим нивоима SOC-а због акумулације биљне биомасе и њеног разлагања у органску материју.

Индекс нормализоване разлике изграђености (енгл. Normalized Difference Built-up Index - NDBI) користи краткоталасне инфрацрвене и блиске инфрацрвене опсега (опсег 11 и опсег 8, код Sentinel-2 сателита) за разликовање урбаних или изграђених подручја.

Израчунава се помоћу формуле:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$$

NDBI користи већу рефлексију изграђених подручја у SWIR опсегу. NDBI је посебно користан у SOC студијама за идентификацију и искључивање урбаних подручја, која обично показују ниже нивое SOC због затварања тла и смањене вегетације.

Још један значајан индекс је индекс нормализоване разлике воде (енгл. Normalized Difference Water Index - NDWI), који наглашава садржај воде у вегетацији и површинским водним телима. Израчунава се помоћу формуле:

$$NDWI = \frac{\text{зелени} - NIR}{\text{зелени} + NIR},$$

при чему зелени и NIR одговарају опсегу 3 и опсегу 8 код Sentinel-2 сателита. NDWI је значајан у процени SOC-а јер помаже у идентификацији мочвара и поплавлених региона. Ове области играју јединствену улогу у глобалном циклусу угљеника, које често карактерише различита динамика SOC због услова богатих влагом.

Вегетацијски индекс прилагођен земљишту (енгл. Soil Adjusted Vegetation Index - SAVI) је дизајниран да минимизира утицај осветљености тла, посебно у областима са ретком вегетацијом.

Израчунава се помоћу формуле:

$$SAVI = (1 + L) * \frac{NIR - \text{црвени}}{NIR + \text{црвени} + L},$$

где је L фактор подешавања позадине крошњи, често постављен на вредност 0.5. Укључивањем и NIR и црвеног опсега (опсег 8 и опсег 4 код Sentinel-2 сателита), SAVI обезбеђује тачнију репрезентацију здравља вегетације у таквим регионима. Ова тачност је кључна за процену SOC, јер обезбеђује да се утицај вегетације на органски угљеник у земљишту на одговарајући начин узме у обзир, посебно у областима са ограниченим биљним покривачем.

На крају, индекс голог земљишта (енгл. Bare Soil Index - BSI) је кључан за идентификацију подручја са изложеним земљиштем. Користи комбинацију SWIR, црвеног, NIR и плавог опсега (опсежи 11, 4 8 и 2 код Sentinel-2 сателита).

Израчунава се помоћу формуле:

$$BSI = \frac{(SWIR + \text{црвени}) - (NIR + \text{плави})}{(SWIR + \text{црвени}) + (NIR + \text{плави})}$$

BSI је посебно релевантан у SOC студијама за разликовање површина голог тла од вегетативних региона, пошто ове области показују различите карактеристике SOC због одсуства вегетационог покривача.

5.8 Атрибути терена

Атрибути терена, су једна од суштински најважнијих ставки у домену даљинске детекције. Атрибути терена су критични за тумачење и разумевање карактеристика Земљине површине. У даљинској детекцији, атрибути терена су изведени из дигиталних

модела елевације (енгл. Digital Elevation Models - DEM), који представљају репрезентације надморске висине Земљине површине снимљене са сателитских или снимака из ваздуха. Анализа атрибута терена је кључна у различитим еколошким, геолошким и студијама заштите животне средине, јер нуде увид у топографске карактеристике које утичу на природне процесе као што су хидрологија, обрасци вегетације, ерозија и стабилност рељефа.

У даљинској детекцији, атрибути терена као што су закривљеност (енгл. curvature), нагиб (енгл. slope), аспект (енгл. aspect) и надморска висина (енгл. elevation) се издвајају из DEM-а помоћу специфичних математичких формула, од којих свака пружа јединствен увид у пејзаж.

Закривљеност мери савијање површине, што указује на процесе ерозије и таложења:

- Профилна закривљеност утиче на убрзање протока воде дуж нагиба.
- Планска закривљеност утиче на бочно кретање воде и транспорт седимента.

Закривљеност у тачки се израчунава као збир других извода надморске висине у правцу x (исток-запад) и у правцу y (север-југ):

$$\text{закривљеност} = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}$$

где су $\frac{\partial^2 z}{\partial x^2}$ и $\frac{\partial^2 z}{\partial y^2}$ други изводи надморске висине (z) у односу на координате x и y .

Нагиб је стрмина или инклинација терена, кључан за разумевање потенцијала отицања и ерозије:

$$\text{нагиб} = \tan^{-1} \left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \right)$$

где $\frac{\partial z}{\partial x}$ и $\frac{\partial z}{\partial y}$ представљају брзину промене надморске висине (z) у правцима x и y .

Аспект одређује правац нагиба, утичући на микроклиматске услове:

$$\text{аспект} = \tan^{-1} \left(\frac{\partial z / \partial y}{\partial z / \partial x} \right)$$

Овде, $\frac{\partial z}{\partial x}$ и $\frac{\partial z}{\partial y}$ су градијенти надморске висине у правцу исток-запад и север-југ.

Надморска висина представља висину изнад референтне тачке, обично нивоа мора. Подаци о надморској висини се директно добијају из DEM-а и не захтевају прорачун преко формуле. Утиче на еколошке и климатске услове у различитим пределима.

6 ОРГАНСКИ УГЉЕНИК У ЗЕМЉИШТУ

Угљеник је кључан за опстанак свих живих организама, јер представља темељни елемент из ког су изграђени сви познати облици живота на Земљи. Угљеник представља есенцијални елемент за све живе организме и креће се кроз различите резервоаре угљеника током различитих процеса. Резервоари угљеника се односе на различита складишта у којима се угљеник налази у животној средини. Угљеник стално кружи између резервоара кроз различите процесе као што су фотосинтеза, дисање и разлагање у процесу који се назива циклус угљеника.

Главни резервоари угљеника укључују[78]:

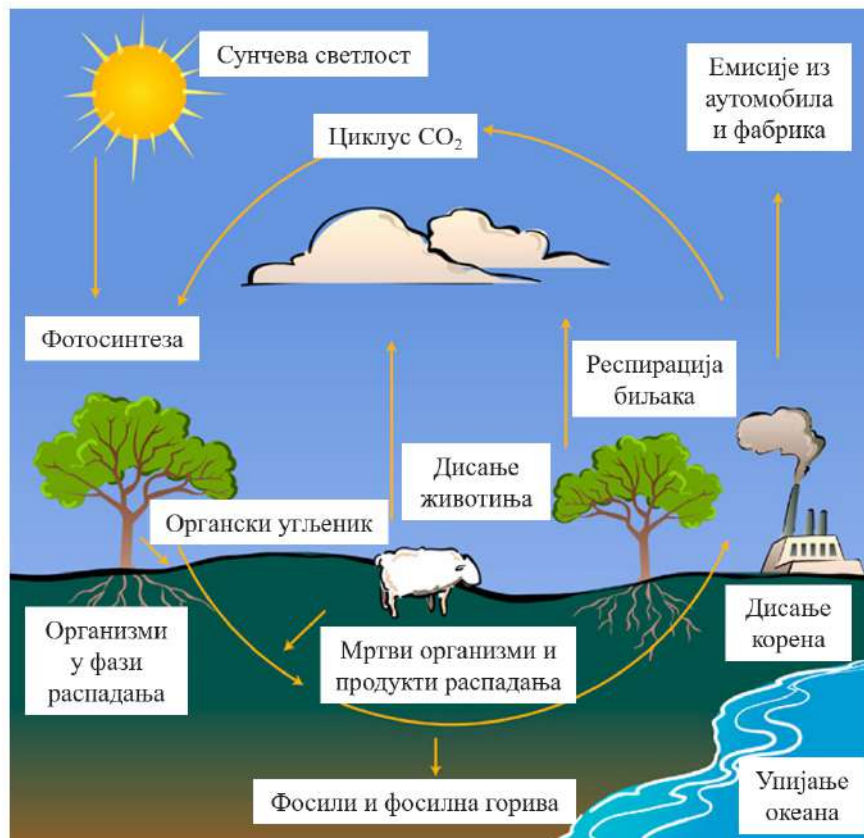
- Атмосферу: Угљеник је присутан у атмосфери у облику угљен-диоксида (CO_2) и других гасова. Атмосфера игра кључну улогу у регулисању климе на Земљи.
- Биосферу: Биосфера укључује све живе организме на Земљи, као што су биљке, животиње и микроорганизми. Угљеник се складишти у биосфери у облику биомасе, која укључује све, од листова и корена до стабљика и дебла.
- Хидросферу: Хидросфера обухвата сву воду на Земљи, укључујући океане, језера, реке и подземне воде. Угљеник се налази у хидросфери у облику раствореног CO_2 , који је важан за хемијски састав океана и биолошку продуктивност морских екосистема.
- Геосферу: Геосфера укључује све стене, седименте и тло на Земљи. Угљеник се налази у геосфери у облику фосилних горива (као што су угаљ, нафта и природни гас) и у облику неорганских једињења у стенама и минералима.

Процеси кретања угљеника између резервоара се могу груписати у две главне категорије: био-гео-хемијски циклуси и антропогене активности [78].

Био-гео-хемијски циклуси подразумевају природне процесе којима се преноси угљеник између различитих резервоара.

Главни био-гео-хемијски циклуси који померају угљеник кроз животну средину укључују:

- Фотосинтезу – Процес који се одвија у биљкама, алгама и неким бактеријама, где се CO_2 апсорбује из атмосфере и претвара у органска једињења коришћењем енергије сунчеве светлости.
- Дисање: Процес који се одвија у свим живим организмима, укључујући биљке, животиње и микроорганизме. Органска једињења се разграђују да би се ослободила енергија и вратио CO_2 у атмосферу.
- Распадање: Процес који се одвија када организми умру или одбаце своје лишће, микроорганизми разграђују њихове остатке, ослобађајући једињења угљеника назад у животну средину.



Слика 35: Циклус кружења угљеника[79]

- Размена између океана и атмосфере: CO₂ се раствара у океану и може се разменити између атмосфере и океана путем дифузије. Океани такође апсорбују велику количину CO₂ кроз процес фотосинтезе од стране морских организама и могу да складиште овај угљеник као седимент и у шкољкама морских створења.

Антропогене активности подразумевају људске активности које представљају извор емисије угљеника и изазивају промене у природној равнотежи циклуса угљеника.

Антропогене активности укључују:

- Сагоревање фосилних горива као што су угаљ, нафта и природни гас за производњу енергије, транспорт и индустријске процесе.
- Крчење шума и промене у коришћењу земљишта које смањују капацитет биосфере да складишти угљеник.
- Пољопривредна и сточарска производња, које такође утичу на кружење угљеника.

Процењује се да укупна количина угљеника који је ускладиштен у геосфери износи 2500 гигатона (Gt) и укључује око 1550 Gt органског угљеника у земљишту и 950 Gt

неорганског угљеника у земљишту (енгл. Soil Inorganic Carbon - SIC). Резервоар геосфере је 3.3 пута већи од атмосферског резервоара који садржи 760 Gt и 4.5 пута већи од биосферског резервоара који складишти 560 Gt угљеника и представља највећи копнени резервоар угљеника [80]. Имајући у виду разлику у величини складишта угљеника у геосфери у поређењу са другим резервоарима угљеника (посебно атмосфером), мала промена у количини органског угљеника у земљишту може да произведе велику промену у нивоу атмосферског угљен-диоксида CO₂. На пример, глобално смањење количине органског угљеника у земљишту од 5 процената на дубини тла до 3 метра, довело би до ослобађања 117 Gt угљеника у атмосферу, узрокујући повећање атмосферског резервоара угљеника од 14 процената [80]. Насупрот томе, мало повећање складиштења угљеника у земљишту из атмосфере огледало би се у значајном смањењу угљеника који је ускладиштен у атмосфери (у виду CO₂ и представљало би глобално значајно смањење.

Земљиште је сложена мешавина органских и неорганских састојака са различитим физичким и хемијским својствима, која показује велику варијабилност од локације до локације или чак унутар истог поља [81]. Земљиште представља кључну компоненту копнених екосистема, јер олакшава циркулацију енергије и материјала између атмосфере и биосфере [82]. Како би се обезбедило ефикасно праћење и омогућила адекватна процена стања земљишта, неопходан је одабир одговарајућих индикатора о стању земљишта.

Иако постоје бројни индикатори који су подједнако важни, изабрани индикатори треба да испуњавају одређене критеријуме. Они би требало да:

- буду прихваћени од стране стручњака као валидни индикатори,
- да се њихово мерење спроводи рутински и да је примењено у великој мери,
- да буду схваћени и прихваћени од стране опште популације како би се постигло глобално прихватање и утицај [83].

Избор одговарајућих индикатора омогућава чврсту процену промена стања земљишта и може побољшати управљање животном средином, усвајање одговарајућих политика и доношење одлука. Мерење садржаја органског угљеника у земљишту је широко прихваћен показатељ квалитета земљишта, јер SOC игра централну улогу у различитим функцијама земљишта [84]. Поред тога, мерење SOC је уобичајена анализа својстава земљишта. Штавише, угљеник као елемент је добро познат и признат од стране глобалне популације [85]. Због ових фактора, SOC се сматра вредним индикатором за процену и праћење промена у плодности земљишта.

Плодност земљишта се дефинише као способност земљишта да ефикасно функционише као компонента у напредујућем екосистему [86]. Количина и квалитет органског угљеника у земљишту су уско повезани са кључним функцијама земљишта, укључујући минерализацију хранљивих материја, стабилност агрегата, пропустљивост за ваздух и воду, задржавање воде и способност инфилтрације [84]. Наведене функције земљишта су повезане са широким низом атрибута екосистема. На пример, високи нивои SOC у минералним земљиштима имају тенденцију да буду у корелацији са високом продуктивношћу биљака што заузврат има позитивне последице на станиште, дистрибуцију и величину популације дивљих животиња [87].

SOC има потенцијал да веже значајан значајан део CO₂ из атмосфере и да на тај начин смањи негативне ефекте повећања концентрације CO₂ у атмосфери, које значајно доприносе глобалном загревању и климатским променама [88], као и са потребом за обезбеђивањем сигурности хране на светском нивоу [89].

Ова двострука улога SOC-а у на глобалном нивоу сугерише да се наведене користи могу постићи кроз стратегије које подразумевају очување постојећих залиха SOC-а (избегавање губитака) и обнављање залиха у земљиштима осиромашеним угљеником [90]. Кроз заштиту и повећање ускладиштеног SOC-а може се:

- заштитити или повећати плодност земљишта,
- одржати или повећати отпорност на климатске промене,
- смањити ерозија земљишта,
- смањити конверзије станишта (где се примењује кроз очување природних екосистема) [91],

Наведене погодности су у складу са циљевима одрживог развоја Уједињених нација (енгл. United Nations Sustainable Development Goals - UN SDG), циљевима Оквирне конвенције Уједињених нација о климатским променама (енгл. United Nations Framework Convention on Climate Change - UNFCCC) и Конвенције Уједињених нација о борби против дезертификације (енгл. United Nations Convention on Combating Desertification - UNCCD)) [92].

Органски угљеник у земљишту игра кључну улогу у ублажавању климатских промена делујући као резервоар угљеника, али претходне студије су истакле историјски губитак угљеника из овог базена [93] и потенцијал за будући убрзани губитак у сценаријима загревања [94] [95]. Као природно климатско решење (енгл. Natural Climate Solution), SOC има потенцијал да обнови резервоар угљеника у земљишту и да заштити од даљих емисија CO₂ које су резултат предвиђеног коришћења земљишта и климатских промена. Ова двострука улога земљишта у глобалном буџету угљеника сугерише да се климатске користи могу постићи кроз стратегије које и чувају постојеће залихе SOC (избегавају губитак) и обнављају залихе у земљишту осиромашеном угљеником [90].

Упркос научном консензусу око потенцијала и вишеструких погодности које доноси развој и примена ускладиштења и секвестрације органског угљеника у земљишту за ублажавање климатских промена оно остаје ограничено у пракси.

Током протеклих деценија објављене су многе процене глобалних залиха органског угљеника у земљишту, а већина студија извештава о глобалној процени SOC-а од отприлике 1500 петаграма угљеника - Pg C (1 Pg представља метричку јединицу еквивалентну једном квадрилиону (10¹⁵) грама), али постоје значајне варијације међу проценама (у распону од 504 до 3000 Pg C) [96]. Велика варијација у проценама глобалних залиха SOC-а настаје услед разлика у периоду узорковања, интензитету и просторној резолуцији база података профила земљишта, као и услед разлика у приступима израчунавању самих процена [1]. Неравномерна дистрибуција геореференцираних профила тла широм света је још један разлог за варијације у проценама [2].

Поред тога, неке студије су укључивале неоргански угљеник, различите нивое садржаја камена [3], природне или антропогене појаве које утичу на нивое угљеника у земљишту, попут поплава, ерозије, пожара, ђубрење земљишта, орање [4] у њиховим проценама залиха угљеника. Такође, студије су показале да разлике у глобалним проценама залиха SOC такође могу настати због запреминске густине (енгл. bulk density) органског земљишта [97]. Наведена недоследност отежава поређење резултата у различитим студијама. Штавише, ове варијације наглашавају потребу за опрезом у обради података и тумачењу разлика у проценама, као и за континуираним побољшањима у прикупљању и обради података како би се добиле боље глобалне процене залиха SOC.

Област од посебног интереса представља намена земљишта и промене земљишног покривача (енгл. Land Use and Land Cover Change - LULCC) која се односи на природне и антропогене модификације начина на који људи употребљавају земљиште (коришћење земљишта) и физичке и биолошке објекте који се налазе на земљиној површини (земљишни покривач). Овај општи појам обухвата све веће људске утицаје на животну средину, који утичу на већа подручја и интензивнији су него у прошлости. Ове промене утичу на еколошке процесе и својства од локалног до глобалног, укључујући климатске промене, биодиверзитет и загађење [1].

LULCC заузима велики удео у глобалној емисији угљеника и налази се на другом месту по количини антропогенних емисија угљеника, одмах иза емисија насталих из сагоревања фосилних горива [98]. Међутим, квантификација емисија LULCC-а је изазовна због сложених интеракција између залиха угљеника, просторне дистрибуције и праксе управљања земљиштем. Количина угљеника ускладиштеног у биомаси и земљишту, просторна дистрибуција залиха угљеника и утицаји праксе управљања земљиштем на биомасу и залихе угљеника у тлу утичу на емисије угљеника из LULCC-а. Услед комплексности и међузависних односа наведених фактора, LULCC представља најнеизвеснији део глобалног кружења угљеника [99] [100].

Док су глобалне залихе угљеника у биомаси и њихова дистрибуција релативно добро истражене и квантификоване са разумном сигурношћу [101] [102] [103] [104], у поређењу са угљеником у земљишту, постоји значајна неизвесност у погледу количине угљеника који се емитује кроз измене намене земљишта и промену земљишног покривача [105] [106]. Треба напоменути да је утицај залиха угљеника у биомаси и земљишту различит. Садржај угљеника у биомаси је неизвесан и креће се од 47-59 процента суве биљне материје [107].

Насупрот томе, садржај угљеника у земљишту је стабилнији и добро објашњен у научним круговима [1]. Пошто динамика угљеника у екосистему увелико варира у зависности од равнотеже између залиха угљеника у земљишту и биомаси, важно је разумети дистрибуцију угљеника у земљишту у односу на угљеник из биомасе [108]. Док је угљеник биомасе мапиран у неколико студија, мало њих је покушало да мапира дистрибуцију SOC-а на глобалном нивоу, где и даље постоје велике неизвесности о величини глобалних SOC залиха, њиховој просторној дистрибуцији и емисијама угљеника из земљишта.

Упркос томе, већина политика ублажавања климатских промена се фокусира на угљеник ускладиштен у биомаси [108]. Међутим, да би се ефикасно решило питање емисија угљеника, неопходно је узети у обзир залихе угљеника како у биомаси тако и у земљишту, јер оба играју важне и различите улоге у глобалној количини угљеника.

Намена земљишта и промена земљишног покривача имају значајан утицај на залихе органског угљеника у земљишту. Конверзија природне вегетације у обрадиво земљиште доводи до значајних губитака SOC, у горњем слоју земљишта на дубини од 1 метра, у распону од 25-50% [109] [110], док конверзија на пашњаке обично доводи до мањих губитака у количини SOC [99]. Утицај LULCC-а и управљања на SOC у великој мери зависи од типа земљишта [111], при чему минерална земљишта доживљавају губитак од 20-40% првобитних залиха угљеника у земљишту када се природни травњаци и шуме претворе у обрадиве површине [112] [113].

Забринутост због великих промена у залихама угљеника услед LULCC-а је подстакла рад и истраживање у пољу процењивања емисија угљеника. У дугорочној студији која је обухватила период од 1990 до 2007 године процењено је да је глобалан ниво SOC у шумама био релативно стабилан и износио је 380 Pg C, али су и идентификована значајна одступања између региона [114]. Залихе SOC-а у тропским шумама су се смањиле за 7,7% (са 164,0 на 151,3 Pg C), углавном због губитка SOC узрокованог крчењем шума унутар нетакнутих тропских шума, док су залихе SOC-а у бореалним и умереним шумама порасле за 4,5 и 7,6%, у истом периоду, због ширења шума након напуштања пољопривреде и смањене сече.

Такође, треба обратити пажњу на чињеницу да SOC није равномерно распоређен у самом земљишту, нити је једнако дистрибуиран на различитим дубинама услед утицаја типа вегетације и климе на вертикалну дистрибуцију SOC у различитим типовима земљишта [115]. Вертикална дистрибуција SOC-а значајно варира у зависности од врсте вегетационог покривача, при чему су области покривене ниским растињем, тј. жбуњем имале најдубљу дистрибуцију, а области под шумама најплићу [115].

На расподелу SOC је такође утицала клима, при чему је количина падавина и садржај глине у земљишту био у позитивној корелацији са укупним садржајем SOC, док је температура била у негативној корелацији [115]. Разлике у дистрибуцији SOC са типом вегетације настале су због разлика у обрасцима алокације биљака и дистрибуцији корена, што утиче на количину уноса угљеника у земљиште.

Земљишта у Северној Америци складиште велике количине угљеника, са количином од приближно 366 Pg C, који се налази у горњем 1 метру земље [116] [117]. У Сједињеним Америчким Државама, већина залиха органског угљеника у земљишту налази се у земљиштима којима се не управља интензивно, као што су шуме, мочваре и пашњаци [118].

Међутим, без обзира на коришћење земљишта, већина SOC се налази у горњем слоју земљишта, који може бити подложен губитку. Конкретно, 74.5% SOC у Северној Америци налази се у горњих 30 cm минералног земљишта [119]. Ове информације сугеришу да је одржавање и заштита залиха SOC-а на земљиштима којима се не управља интензивно, посебно у горњем слоју тла, важна за ублажавање губитака угљеника и обезбеђивање дугорочног здравља земљишта.

Тачна процена и континуирано праћење залиха SOC представљају кључни први корак ка разумевању и дефинисању одговарајућих активности како би се могло утицати на очување и обнављање залиха SOC.

У Сједињеним Америчким Државама, током претходних деценија, било је неколико координисаних напора да се процени количина органског угљеника ускладиштеног у земљишту широм Сједињених Држава на статистички робустан и поуздан начин [84].

Програм за инвентар и анализу шума спроведен од стране Шумарске службе Министарства за пољопривреду САД (енгл. United States Department of Agriculture Forest Service - FIA program), брза процена угљеника (енгл. Rapid Carbon Assessment - RaCA), национална кооперативна анкета о земљишту (енгл. National Cooperative Soil Survey - NCSS) и Међународна мрежа угљеника у земљишту (енгл. International Soil Carbon Network - ISCN) су четири различите процене органског угљеника у земљишту спроведене у Сједињеним Америчким Државама. ФИА прикупља податке о шумском земљишту и дрвећу и укључује податке о тлу на подзорку парцела да би проценила количину SOC-а у шумском тлу и стељи (први растресити слој шумског земљишта) на националном нивоу. RaCA представља подухват на националном нивоу, који предводи Служба за очување природних ресурса (енгл. Natural Resources Conservation Service - NRCS), да се систематски узоркује и мери SOC широм САД користећи доследну и поновљиву методологију. NCSS представља национално партнерство савезних, регионалних, државних и локалних агенција, приватних субјеката и институција које прикупљају информације о земљишту користећи заједничке процедуре. База података (енгл. Gridded Soil Survey Geographic - gSSURGO), која садржи просторне податке о различитим својствима земљишта укључујући SOC, производ је NCSS. ISCN је мрежа за координацију истраживања која олакшава размену података и просторно експлицитне процене SOC-а у контексту пејзажа, климе, коришћења земљишта користећи независне истраживачке пројекте.

Кључне разлике између ових процена леже у њиховом обиму, периодици прикупљања података, методама прикупљања података и географској покривености. FIA и RaCA се фокусирају на националне процене SOC-а, при чему FIA прикупља податке првенствено у шумским подручјима, а RaCA узоркује са различитих типова земљишних покривача као и са земљишта где се примењују разноврсне пољопривредне праксе. NCSS обезбеђује истраживања земљишта на нивоу округа и просторне податке о својствима земљишта, укључујући SOC, док ISCN прикупља податке из независних истраживачких пројеката са варијацима у географској покривености и у методама које се користе. Друга разлика је у томе што ФИА и RaCA прикупљају примарне податке, док NCSS и ISCN прикупљају и компилирају постојеће податке.

Упркос наведеним разликама, све наведене процене деле заједнички циљ, а то је да се процени ниво SOC широм Сједињених Америчких Држава (САД). FIA и RaCA програми су дизајнирани да обезбеде национално конзистентне процене SOC-а на различитим просторним скалама, док NCSS и ISCN пружају вредне информације за специфичне регионе и истраживачка питања. Све четири процене доприносе разумевању SOC-а у САД и могу помоћи у доношењу информисанијих политика и пракси управљања у вези са здрављем земљишта и секвестрацијом угљеника.

Постојање различитих приступа процени SOC може бити проблематично и представља изазов за тачну процену залиха SOC и промена током времена. Сваки приступ има своје предности и ограничења, а резултати добијени коришћењем једног приступа можда неће бити директно упоредиви са онима добијеним коришћењем другог приступа. Ово може отежати идентификацију трендова и промена у SOC-у у различитим регионима и екосистемима, што заузврат може искомпликовати напоре за управљање и ублажавање утицаја коришћења земљишта и климатских промена на SOC. Да би се решио овај изазов, важно је развити стандардизоване методе за процену и извештавање

ње о SOC, као и координисати и интегрисати податке из различитих напора пројеката процене.

Замисао је била да добијене процене буду основа за откривање будућих промена нивоа SOC-а у целој земљи, међутим постоје одређена ограничења за добијање адекватних процена на толико великим просторима. Једно ограничење представља чињеница да дубина земљишта и залиха SOC може варирати широм земље. Такође, већина процена не узима у обзир тло дубље од 1 метра, иако дубље земљиште може бити важан резервоар SOC [120]. Такође, ограничење представља и чињеница да је тешко открити промене SOC мање од око 25% услед природне хетерогености земљишта, као и грешака узорковања и мерења до којих може доћи [121]. Нетачности у проценама могу настати услед различитих фактора, као што су на пример укључивање стена већих од 2 мм у узорке земљишта, промене у текстури земљишта, као и осетљивост залиха SOC на запреминску густину [122].

Напредне технологије мерења и статистичке методе које могу да објасне различите изворе несигурности имају потенцијал да одговоре на горе поменуте изазове и омогуће детекцију нијансираних промена у органском угљенику у земљишту. Непотпуно разумевање о томе како на промене нивоа SOC утичу клима, коришћење земљишта, управљање земљиштем и едафски фактори доприносе сложености у дизајнирању одговарајућег метода/алата за праћење, извештавање и верификацију стања SOC [123].

6.1 Директно мерење промена у залихама SOC

Директна мерења залиха SOC ослањају се на физичко узорковање и мерење садржаја угљеника у земљишту. Типично, фракције fine земље (мање од 2 mm) и грубих минерала (веће од 2 mm) у земљишту се квантификују, заједно са концентрацијом органског угљеника (OC) (%) и запреминском густином земљишта [124]. У неким случајевима може бити важно квантификовати и узети у обзир грубу фракцију подземних OC, као на пашњацима или шумским земљиштима [124]. Прецизна процена садржаја стена у узоркованом земљишту представља изазов, јер може значајно утицати на запреминску густину земљишта, а самим тим и на количину земљишта која се узоркује унутар дате дубине узорковања [125] [126].

Директна мерења залиха SOC такође захтевају одговарајући дизајн студија и протоколе узорковања како би се узела у обзир велика просторна варијабилност залиха SOC [88]. Да би се минимизирали потенцијални извори грешке и минимална разлика која се може детектовати (тј. најмања разлика у залихама SOC-а која се може детектовати као статистички значајна између два периода узорковања), често је потребан велики број узорака земљишта [127].

Међувладин панел о климатским променама (енгл. Intergovernmental Panel on Climate Change - IPCC) препоручује минималну дубину узорковања од 30 cm да би се правилно процениле промене у садржају угљеника у земљишту. Међутим, неколико дугорочних агрономских истраживања наишло је на проблем услед повећања дубине механичке обраде земљишта током година, услед све веће снаге пољопривредних машина, што може додати несигурност у резултате [123]. Повећање садржаја угљеника у земљишту кроз методе као што су пољопривреда без механичке обраде и сађење усева са дубоким кореном захтевају узорковање земљишта са већих дубина како би се потврдио очекивани

ефекат [128] [129]. Препоручује се дубље узорковање земљишта на дубини од 100 см, али то често захтева посебне машине и велика новчана средства [124].



Слика 36: Узорковање земљишта за процену садржаја органског угљеника[130]

Поред трошкова везаних за прикупљање и обраду узорака земљишта и само мерење садржаја угљеника у земљишту може бити изузетно скупо. На пример, процењено је да би коришћење методе сувог сагоревања, која представља стандардну референтну методу за SOC анализу како би се откриле значајне промене у залихама SOC у шумским екосистемима у Финској, коштало око 4 милиона евра само за једну кампању узорковања. Како би се обезбедило континуирано праћење стања SOC морао би да се поново уради циклус узорковања за следећи интервал (после нпр. 10 година), што би опет подразумевало селектовање и узорковање са 3.000 парцела на националном нивоу [5].

Како би се решили високи трошкови повезани са директним мерењем залиха SOC, комбинација директних мерења и моделирања може се користити за процену промене SOC на већим просторним скалама. Овај приступ може бити исплативији и коришћен је за процену промене SOC у националним платформама инвентара гасова стаклене

баште. Процена исплативости различитих метода за мерење и секвестрацију угљеника при различитим наменама земљишта и другачијим социо-економским условима је кључна за дефинисање ефикасности различитих пракси управљања земљиштем у циљу повећања секвестрације угљеника у земљишту.

6.2 Процењивање промена у залихама SOC путем мерења токова угљеника

Једну од алтернатива директном мерењу SOC представља метода мерења токова угљеника, која се користи за процену промена у залихама органског угљеника у земљишту током времена. Подразумева мерење нето баланса токова угљеника које размењује екосистем, што укључује почетно усвајање угљеника фотосинтезом (брutto примарна производња), његове накнадне делимичне губитке респирацијом (земљиште, биљке) и даље уносе угљеника (као што су органско ђубрење) и испусте (као што је жетва) [131]. Наведени нето биланс се назива нето размена екосистема (енгл. Net Ecosystem Exchange - NEE) или нето производња екосистема.

Мерење токова се обично врши помоћу мерења у комори или путем методе вртложне коваријансе (енгл. Eddy Covariance - EC), која укључује процену размене угљендиоксида између екосистема и атмосфере [132].

Међутим, ова мерења су подложна бројним несигурностима, укључујући неопходне претпоставке повезане са обрадом података, чињеницу да је ова метода мерење тачке у простору и релативно мале нето промене у резервоарима угљеника у тлу у поређењу са угљеником ускладиштеним у биомаси када се мери у кратким временским периодима (тј. мањим од 5 година) [133]. Да би се превазишле ове несигурности дошло је до развоја инструментације (напредак у перформансама и подешавању анализатора) [134], у прикупљању и обради података (у регистраторима података, софтверима) што је у великој мери побољшало поузданост процена [135].

Штавише, хармонизоване мреже дугорочних места за посматрање, као што су интегрисани систем за посматрање угљеника (енгл. Integrated Carbon Observation System - ICOS) [136] у Европи и глобална мрежа FLUXNET, створене су да обезбеде приступ стандардизованим [137] подацима и да квантификују ефикасност секвестрације угљеника и/или емисије гасова са ефектом стаклене баште (енгл. Greenhouse gases - GHG).

6.3 Процењивање промена у залихама SOC коришћењем геостатистичких техника

Геостатистичке технике се обично користе за мапирање садржаја органског угљеника у земљишту на целим површинама од интереса. Традиционалне геостатистичке методе, попут кригинга, укључују узорковање и мерење SOC променљиве на одређеним локацијама, а затим процену вредности на неузоркованим локацијама кроз интерполацију, регресију или друге технике предвиђања. Тачност ових метода у великој мери зависи од величине узорка, конфигурације узорковања и степена просторне корелације варијабле SOC [139].

Насупрот традиционалних постоје и хибридне геостатистичке технике које узимају



Слика 37: Систем вртложне коваријансе[138]

у обзир корелацију животне средине. Хибридне методе омогућавају коришћење секундарних информација, као што су надморска висина, нагиб, аспект, тип тла и намењене земљишта, које су често доступне у финијој просторној резолуцији од узоркованих вредности примарне варијабле SOC. Када је корелација између примарних и секундарних варијабли значајна, хибридне технике генерално резултују прецизнијим локалним предвиђањима од обичног кригинга или других униваријантних предиктора [140] [141].

Геостатистичке технике које узимају у обзир корелацију животне средине укључују ко-кригинг, кригинг са спољним помаком и регресијски кригинг [142]. Хибридне методе се разликују по својим претпоставкама о односима између примарних и помоћних варијабли и начину на који се секундарне информације користе у процени примарне променљиве на неузоркованим локацијама.

Ко-кригинг је метода која укључује и примарне и секундарне варијабле за процену вредности примарне варијабле на неузоркованим локацијама. Претпоставља да су примарне и секундарне варијабле просторно повезане и процењује примарну варијаблу узимајући у обзир просторну корелацију између две варијабле [142]. Кригинг са спољним помаком је још један метод који користи секундарне информације за процену примарне променљиве на неузоркованим локацијама. У овој методи, претпоставља се да секундарна променљива има линеарну везу са примарном променљивом, а резидуали из ове везе се процењују коришћењем кригинга [142]. Регресиони кригинг је метода која комбинује предности регресионе анализе и кригинга. Користи однос између примарне променљиве и секундарних варијабли да би предвидео вредности примарне променљиве на неузоркованим локацијама, а затим користи кригинг да процени резидуале [142].

6.4 Процењивање промена у залихама SOC путем метода даљинске детекције са земље

Директне методе за мерење SOC укључују деструктивно узорковање, као и лабораторијску обраду и анализу, који су дуготрајни и скупи. Методе даљинске детекције са земље подразумевају технике које не нарушавају структуру земљишта и екосистема и омогућавају мерење и прикупљање података на недеструктиван начин.

Последњих година, спектроскопске технике, као што је блиска и средња инфрацрвена спектроскопија (енг. Near-Infrared Spectroscopy - NIRS и енгл. Mid-Infrared Spectroscopy - MIRS), су се појавиле као обећавајуће недеструктивне методе за мерење SOC. NIRS и MIRS функционишу тако што мере како земљиште интерагује са светлосним зрачењем различитих таласних дужина, што може дати информације о садржају SOC, као и другим хемијским и физичким својствима земљишта[143].

Једна од главних предности спектроскопских метода је што омогућавају много бржи проток узорака у поређењу са методама сувог сагоревања, што значајно смањује време и трошкове анализе. Поред тога, за спектроскопске методе се употребљава опрема која је физички мања и није толико захтевна за употребу и потенцијално се може применити у теренским лабораторијама уз нижу употребу ресурса [144].

Упркос предностима спектроскопских метода, тачност процена SOC добијених овим техникама мора бити пажљиво калибрисана за различите географске области и типове земљишта користећи методе сувог сагоревања као референцу.

Такође, спроводе се истраживања у домену побољшања искористивости спектроскопских метода, са идејом да се уређаји не носе ручно и да оператер не врши мерења, него да се развију "сензори у покрету". Сензори у покрету би били монтирани на тракторе или друга наменска возила за узорковање, којима би се прелазило преко површине од интереса и уређаји би континуирано мапирани концентрације SOC [145].

Поред тога, тестиране су и разне друге неконвенционалне спектроскопске технологије, као што су ласерски индукована декомпозициона спектроскопија распадања (енгл. laser-induced breakdown spectroscopy - LIBS), инфрацрвена спектроскопија Фуријеове трансформације дифузне рефлексije (енгл. diffuse reflectance Fourier transform infrared spectroscopy - DRIFTS) и нееластично неутронско расејање (енгл. inelastic neutron scattering - INS) [146], али ниједан метод се још није појавио као одржива замена за конвенционалне методе анализе.

6.5 Процењивање промена у залихама SOC путем метода даљинске детекције из ваздуха

Даљинска детекција из ваздуха коришћењем авиона

Хиперспектрално снимање из ваздуху представља драгоцено средство за процену услова земљишта и за мапирање варијабилности унутар области од интереса са високом тачношћу. Методама даљинске детекције из ваздуха могу да се покрију велике површине током једног лета, јер може да се утиче на трајање опсервационог лета [148].

Омогућавају прикупљање драгоцених информација за дигитално мапирање земљишта, као и податке који се могу искористити у циљу сегментирања области од интереса



Слика 38: Мобилни уређај за NIR спектроскопију[147]

на основу хетерогености тла [149]. Авиони имају капацитет да носе терете великих тежина што омогућава да се на њих инсталирају хиперспектрални сензори различитих спектралних опсега, као и да раде под условима високе покривености облацима [150]. Коришћењем хиперспектралног сензора за добијање спектралних података у различитим врстама тла откривено је да се рефлексја смањује у зависности од типа земљишта, не само због варијација у садржају SOC-а, него и услед хетерогености у минералогји и садржају влаге у земљишту [151].

Друга студија је процењивала перформансе хиперспектралног сензора, постављеног на авион, за моделе предвиђања SOC. Показано је да грубоћа тла има значајан утицај на тачност модела јер су најнеповољни услови резултирали нижим вредностима тачности модела [152].

Такође, употребом хиперспектралног сензора утврђено је да су процене SOC веома осетљиве на утицај вегетационог покривача, што доводи до прецењености или потцењености у процени у зависности од техника за прераду података које су коришћене [153].

Даљинска детекција из ваздуха коришћењем беспилотних ваздушних система
Током последњих неколико година, дошло је до све већег интересовања за коришћење беспилотних ваздушних система (БВС) као платформе за даљинску детекцију због њихове ниске цене и могућности коришћења напредних сензора [154]. БВС су опремљени сензорима који имају високу спектралну резолуцију и малу величину, што је у комбинацији са смањеном ценом камера и самих платформи довело до експоненцијалног



Слика 39: Специјално прерађени авион ДЦ-8 у власништву NASA[48]

повећања њихове примене. БВС комбинују карактеристике сателитског и снимања из авиона, чиме су у могућности да обезбеде високу просторну резолуцију, као и кратко време поновног снимања подручја од интереса, што је потребно за покривање разноликости подручја од интереса [154].

Међутим, упркос овим предностима, још увек постоји ограничена употреба БВС у истраживањима о процени својстава земљишта због проблема који се тичу поузданости платформе, ограничене носивости, спектралног опсега сензора, ограниченог трајања лета и обраде слике [155]. Упркос напретку који је постигнут у процени неколико еколошких и климатских варијабли заснованих на применама БВС, усвајање наведених платформи још увек не представља оптимално решење за праћење параметара земљишта. Показано је да употреба БВС опремљеним мултиспектралном камером за процену SOC има велики потенцијал. Међутим, уочена су и ограничења, као што је прецењивање вредности за ниске концентрације SOC и потцењивање вредности за високе концентрације SOC.



Слика 40: Беспилотни ваздушни систем за даљинску детекцију (NASA)[156]

6.6 Процењивање промена у залихама SOC путем метода даљинске детекције из свемира

Употреба слика добијених даљинском детекцијом из свемира има велики потенцијал за генерисање просторних мапа горњег хоризонта тла кроз повезаности између специфичних хемијских веза земљишта и електромагнетног зрачења [157].

Оптичка сателитска мултиспектрална слика ушла је у широку употребу за процењивање SOC од 1980-тих година [158], а хиперспектрални подаци постали су популарни када је HiRegion свемирски систем постао доступан од 2000. године [159].

Мултиспектрални сензори прикупљају податке у неколико специфичних таласних дужина, обично у видљивим и блиско-инфрацрвеним деловима електромагнетног спектра. Обично имају неколико уских спектралних опсега, крећу се од 3 до 10 и дају ограничен број спектралних потписа за различите типове покривача земљишта. Хиперспектрални сензори, са друге стране, прикупљају податке у бројним уским, непрекидним спектралним опсезима који покривају таласне дужине од видљивих, па до инфрацрвених региона електромагнетног спектра. Обично имају стотине спектралних опсега, пружајући врло детаљан и континуирани спектрални потпис за сваки пиксел.

Међутим, употреба наведених технологија за опсервирање земљишта је ограничена потребом за геометријским, атмосферским и радиометријским исправкама података. Поред тога, неопходно је било упаривање са истодобним подацима (опсервацијама) са тла, а јављали су се и проблеми са препрекама везаним за вегетациони покривач земљишта. Као резултат тога, спроведен је мањи број студија које користе сателитске сензоре за процену SOC [157].

У последње време, процена и мапирање SOC-а на основу података даљинске детекције из свемира доживљавају значајан помак због промена у постојећим праксама, попут измене у политици која сада омогућава дистрибуирање Landsat података без плаћања надокнада и бесплатно коришћење података са Sentinel серије сателита.

Такође, долази до развоја нових приступа попут развоја и коришћења великих флота сателита величине кутије за ципеле (компанија Planet из САД).

Даље имамо и лансирање нових програма за употребу хиперспектралних сензора

попут Програма мапирања и анализе животне средине (енгл. Environmental Mapping and Analysis Program - EnMAP). EnMAP програм ће пружити до сада невиђене токове података за преузимање и праћење SOC преко VNIR-SWIR спектралног распона.

Све наведено доводи до тога да употреба сателитских сензора и података добијених даљинском детекцијом за процену и мапирање садржаја органског угљеника у земљишту добија све више на значају.

Претходна истраживања су показала да подаци добијени путем даљинске детекције могу бити успешно коришћени као помоћне варијабле у мапирању својстава земљишта. Употребом геостатистичких метода у комбинацији са различитим параметрима добијеним путем даљинске детекције добијен је метод који је прецизнији од употребе само обичног кригинга, у предвиђању просторне варијабилности SOC и развоју висококвалитетних мапа [160].

Хиперспектрални подаци са сателитских снимака су такође корисни за процену садржаја SOC, имамо пример коришћења података из Hyperion сензора на EO-1 сателиту за процену садржаја SOC, где су постигнути обећавајући резултати [12]. Међутим на тачност процена утицали су фактори попут ниског односа сигнала и шума, као и просторна резолуција Hyperion-а од 30 м. Неопходно је напоменути да након деактивације Hyperion-а 2017. године не постоји више активан хиперспектрални сателит који снима слике у VNIR-SWIR региону. Услед тога, предности коришћења хиперспектралних података добијених из свемира за процену SOC су демонстриране путем студија симулације, као што су EnMAP, PRISMA, и HypIRI [17] [18].

Недавно су технике експлоатације података коришћене како би се прикупиле спектралне рефлексије земљишта са сателитских слика, омогућавајући развој модела предвиђања својстава тла, укључујући SOC [161].

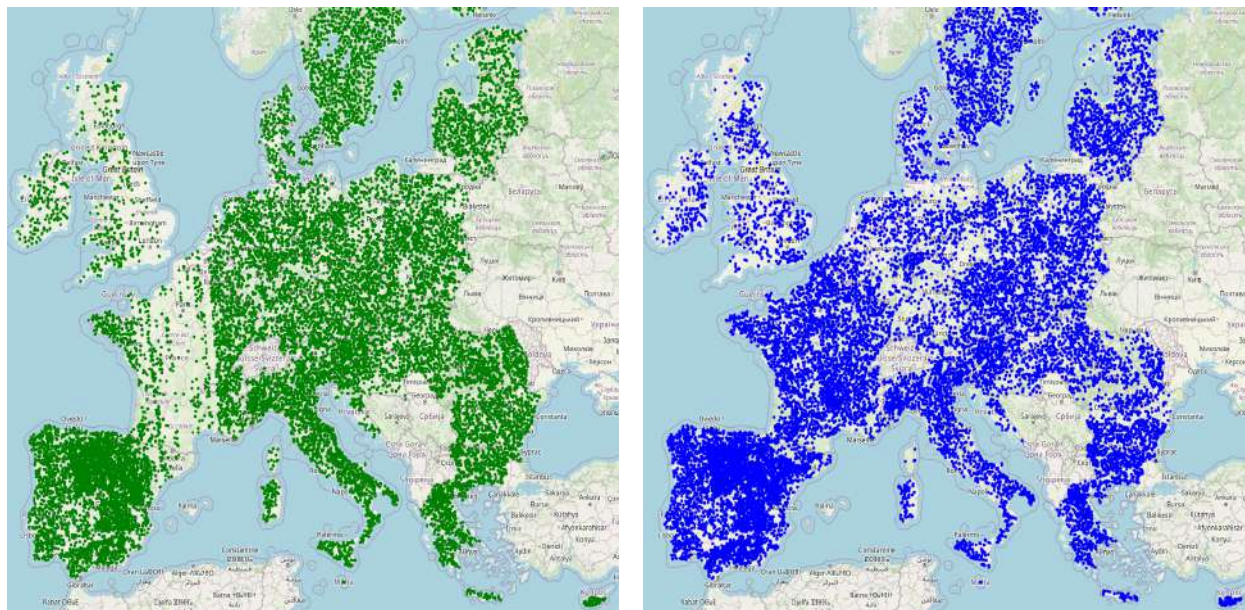
Такође се показало да је употреба мултиспектралних сателитских сензора попут Sentinel-2 повољна за добијање висококвалитетних информација о варијацијама у SOC [162].

6.7 Статистичко истраживање оквира површине употребе/покривача земљишта - LUCAS

Од 2006. године, статистичко истраживање оквира површине употребе/покривача земљишта (енгл. Land Use/Cover Area Frame Statistical Survey - LUCAS) се спроводи сваке 3 године у циљу прикупљања података о коришћењу земљишта и земљишном покривачу широм Европске уније. Истраживање је јединствено по томе што пружа информације in-situ, што значи да се подаци прикупљају директно са самог земљишта. LUCAS истраживања генеришу три врсте информација: (а) микроподатке који садрже статистичке информације прикупљене у свакој тачки где је вршено узорковање (б) фотографије тачака и пејзажа и (в) статистичке табеле са агрегираним резултатима по употреби земљишта и по коришћењу земљишта на географском нивоу. Прикупљени микроподаци, између осталог, служе за продукцију, верификацију и валидацију података CORINE LC скупа података.

Истраживање се састоји од двофазног узорка подручја. У првој фази, оквир од око 1 100 000 геореференцираних тачака (тзв. главни узорак или узорак прве фазе) се

систематски бира из мреже од 2 km^2 која покрива територију EU-28. Овај оквир се затим стратификује према класама покривача земљишта. Земљишни покривач за ове тачке је класификован на основу фото-интерпретације фотографија из ваздуха или сателитских снимака снимљених 2004. и 2005. године. Из главног узорка, затим се бира узорак друге фазе са циљем да се обезбеди статистички значајна покривеност за сваки регион, али узимајући у обзир доступност тачака за узорковање [163]. На пример, LUCAS не даје никакве податке за тачке изнад 1500м. Слика 41. приказује просторну расподелу података LUCAS 2015. и 2018. године.



Слика 41: Просторна расподела података LUCAS 2015. и 2018. године

6.8 Чилеанска база података о органском угљенику у земљишту - CHLSOC

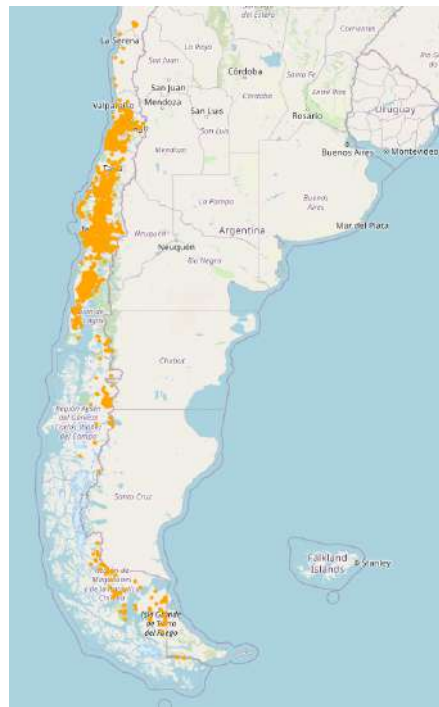
Чилеанска база података о органском угљенику у земљишту (енгл. Chilean Soil Organic Carbon Database - CHLSOC) је највеће и најсвеобухватније складиште података о органском угљенику у земљишту у Чилеу. CHLSOC репозиторијум је настао заједничким напором који је укључивао 39 јавних и приватних институција. CHLSOC представља националну иницијативу без преседана.

Направљен између маја 2018. и априла 2019., CHLSOC укључује различите изворе података, укључујући истраживања земљишта, публикације и необјављене истраживачке податке. База података садржи 13.612 тачака података, од којих је 89% раније било необјављено или недоступно научној заједници.

CHLSOC пружа вредан увид у временску дистрибуцију података о земљишту у Чилеу. Датум прикупљања узорка је доступан за више од 90% укључених података, омогућавајући истраживачима да истраже промене у SOC током времена и истраже временске обрасце и трендове. Већина тачака је узоркована између 2006. и 2018. године, а неки подаци датирају из 1959. године. У скупу података постоји 6.900 тачака

релевантних за ово истраживање, пошто су прикупљене у временском оквиру у којем је мисија Sentinel-2 била оперативна.

У Чилеу, дистрибуција података о земљишту и SOC је високо концентрисана у регионима који имају интензивне пољопривредне и шумарске активности, обухватајући приближно 25% територије земље [164]. Ова подручја, која се одликују висококвалитетним земљиштем и доступним водним ресурсима, доживела су значајну конверзију коришћења земљишта за пољопривреду, шумарство и урбани развој. Међутим, изван ових региона, постоји приметан недостатак података о земљишту, посебно у областима са ограниченим пољопривредним и шумарским активностима. Слика 42. приказује просторну дистрибуцију CHLSOC скупа података.



Слика 42: Просторна дистрибуција CHLSOC скупа података

Да би се олакшало истраживање, управљање земљиштем и настојања у вези са доношењем одговарајућих политика, CHLSOC је слободно доступан регистрованим корисницима за преузимање под међународном јавном лиценцом Creative Commons Attribution 4.0.

6.9 Координација информација о животној средини - CORINE

Координација информација о животној средини (енгл. Coordination of Information on the Environment - CORINE) је база података Европске агенције за животну средину (енгл. European Environment Agency - ЕЕА) и њених земаља чланица у оквиру Европске мреже за информисање и посматрање (енгл. European Network for Information and Observation - EIONET). Тренутно су CLC скупови података део Европског програма Коперник (енгл. European Copernicus programme) и ЕЕА координира услуге праћења

пейзажа за ПАН-европски регион. Укратко, CORINE је одређен да стандардизује прикупљање података о земљишту у Европи како би се подржао развој животне средине. Број земаља учесница се временом повећавао и тренутно укључује 33 земље чланице ЕЕА и шест земаља-сарадника ЕЕА [165].

Први развој CORINE-а почео је 1986. године и трајао је до 1998. године, а прва референтна година је 1990. (CLC1990). Нова верзија излази сваких шест година. До сада је имплементирано пет верзија, односно CLC1990, CLC2000, CLC2006, CLC2012 и CLC2018 [166] [165]. CORINE карте се традиционално добијају у векторском формату фотоинтерпретацијом сателитских снимка у размери 1:100000. Даље, правила CORINE мапирања остају иста: минимална јединица мапирања (енгл. minimal mapping unit - MMU) од 25 ха, минимална ширина мапирања (енгл. minimal mapping width - MMW) од 100 м (табела 1) [167].

| | CLC1990 | CLC2000 | CLC2006 | CLC2012 | CLC2018 |
|----------------------------------|-------------|---------------------|---------------------|---------------------|--------------------------|
| Сателит | Landsat 4/5 | Landsat 7 | SPOT 4/5, IRS P6 | IRS P6, RapidEye | Sentinel-2, Landsat 8 |
| Датум снимања сателитског снимка | 1986 - 1998 | 2000 (+/- 1 година) | 2006 (+/- 1 година) | 2011 - 2012 | 2017 - 2018 |
| Геометријска прецизност | 50 m | 25 m | 25 m | 25 m | 10 m (Sentinel-2) |
| CLC MMU | 25 ha | 25 ha | 25 ha | 25 ha | 25 ha |
| CLC MMW | 100 m | 100 m | 100 m | 100 m | 100 m |
| Локацијска прецизност | 100 m | >100 m | >100 m | >100 m | >100 m |
| Тематска прецизност | 85 % | 85 % | 85 % | 85 % | 85 % |
| Трајање (године) | 13 | 5 | 4 | 3 | 1.5 |
| Број земаља учесница | 22 | 32 | 38 | 39 | 39 |

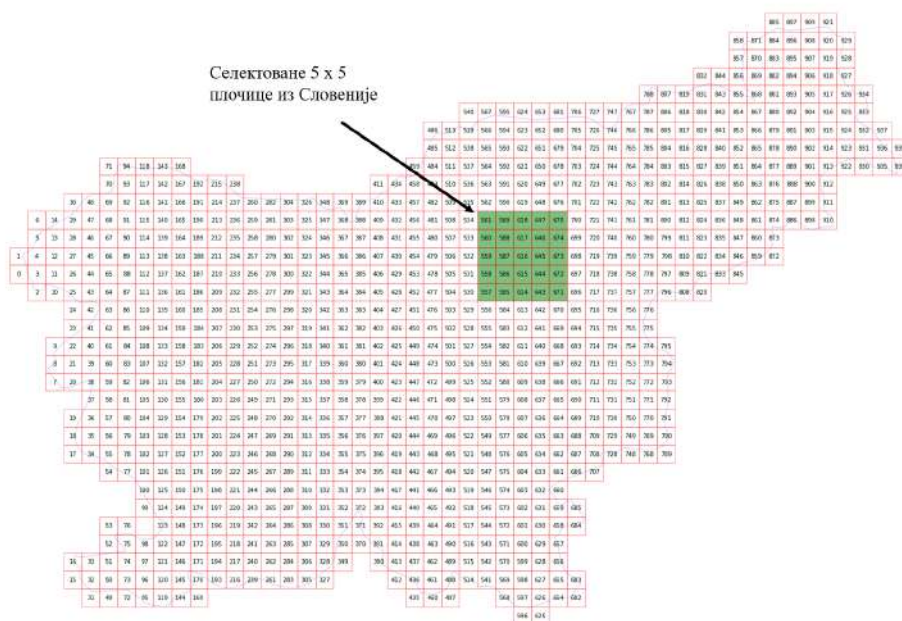
Табела 1: Еволуција CORINE базе података

CLC има детаљну хијерархијску номенклатуру на три нивоа. Најдетаљнији ниво је ниво 3, са максимално 44 категорије. Ниво 2 и ниво 1 имају највише 15, односно 5 категорија (видети табелу 2). Три нивоа су међусобно повезана пошто је номенклатура нивоа 1 хијерархијско поступно агрегирање категорија нивоа 2 и нивоа 3 [168] [165].

Нажалост, тематска прецизност CORINE (85%) није довољно висока да би омогућила његову употребу за тренирање алгорита машинског учења који би поуздано детектовао LC. Због тога је примењен мањи, али тачнији скуп LC података када је у питању тренирање прве фазе приступа представљеног у овом истраживању.

6.10 EO-learn земљишни покривач

За тренирање прве фазе предложеног приступа употребљени су "ground-truth" LC подаци који су доступни за територију Словеније и којима се може приступити преко сајта [169]. Цела територија Словеније подељена је на 940 плочица, свака величине 500 x 500 пиксела, при чему сваки пиксел одговара површини од 10 x 10 квадратних метара (Слика 43.).



Слика 43: Подела територије Словеније на плочице

Језгро скупа података састоји се од сателитских снимака прикупљених од стране Sentinel-2 сателита. Наведена је референца на нивоу земље за земљишни покривач. Референца је доступна у облику геопакета, који садржи полигоне и њихове одговарајуће ознаке. Ознаке представљају следећих 7 класа:

| LULCID | Назив | Репрезентација (%) |
|--------|----------------------|--------------------|
| 1 | Култивисано земљиште | 21.6 |
| 2 | Шума | 38.2 |
| 3 | Пашњак | 24.8 |
| 4 | Грмље | 1.9 |
| 5 | Вода | 0.6 |
| 6 | Мочвара | 0.01 |
| 7 | Вештачка површина | 12.29 |

Табела 2: Класификација земљишта по употреби и земљишном покривачу

Иако су класе које су овде представљене важне за тренирање LULC фазе модела, оне се заправо не користе као резултати у приступу у овом истраживању. Такође је важно

напоменути да немају све класе исту заступљеност у скупу података за тренирање, што је посебно видљиво ако се узме у обзир географија региона Словеније (тј. претежно планински регион, са мало мочвара). Стога се очекује да перформансе самог модела сегментације неће бити избалансиране у односу на наведене класе. Међутим, од интереса у овом истраживању је само „знање“ модела које се инкапсулира у векторе латентних обележја, и нагласак није превише на стварним класама.

За сваку плочицу доступни су општи, статички и динамички подаци. Општи подаци обухватају димензије плочица, географске координате и временске ознаке када су снимљени сателитски снимци. Статички подаци обухватају информације о земљишном покривачу, категорисане у 7 могућих класа (култивисано земљиште, шума, пашњак, грмље, вода, мочваре и вештачка површина).

Претпоставља се да земљишни покривач или остаје стабилан или се веома мало мења током једногодишњег интервала за који се прикупљају подаци. Динамички подаци се састоје од сателитских снимака. У зависности од плочице, постоји отприлике између 50 и 100 сателитских снимака прикупљених током 2019. У тренирању ће бити коришћена једна слику по одговарајућој области, при том ће бити изабране оне слике које одговарају временском оквиру постојеће LULC мапе истинитости терена. Овај временски оквир је цела 2019. година, међутим, изабран је месец јул, због најмање покривености облаком.

7 ПРЕДЛОЖЕНИ ПРИСТУП ПРОЦЕНИ SOC ПУТЕМ ПРИМЕНЕ ДУБОКИХ НЕУРОНСКИХ МРЕЖА И ПОДАТАКА ДОБИЈЕНИХ ДАЉИНСКОМ ДЕТЕКЦИЈОМ

У овом истраживању, циљ нам је да одредимо органски садржај земљишта, тј. садржај органског угљеника у земљишту коришћењем снимака добијених од Sentinel-2 сателита, који пружају просторно континуиране и трошковно исплативе податке о стању Земљине површине. Да бисмо ово постигли, користимо технике машинског учења, које захтевају скуп података састављен од сателитских снимака и одговарајућих SOC мерења прикупљених узорковањем на терену, на којима се обучавају и валидирају ML модели.

Док су сателитски снимци лако доступни (у случају Sentinel-2 од 2015.), отворени скупови података који садрже SOC информације су ретки. У овој студији фокусирамо се на два таква скупа података са највећом количином SOC информација у временском оквиру операције Sentinel-2: LUCAS скупу података прикупљеном од стране ЕУ и чилеанској бази података о органском угљенику у земљишту CHLSOC.

Наведени скупови података садрже хиљаде узорака, али је њихова употреба и даље ограничена када је у питању тренирање DNN-ова од почетка до краја. Како би се заобишао наведени проблем, предлог је да се користи DNN истрениран за визуелни задатак за који је доступно довољно података (класификација земљишног покривача) као екстрактор обележја за SOC процењивач. Сличан приступу који су Гиршик и сарадници користили у њиховом раду [21]. Екстрактор обележја, предложен у овом истраживању је, међутим, трениран за мултиспектралне Sentinel-2 снимке и заснован је на архитектури за сегментацију слике.

Да би се истренирао екстрактор обележја, потребан је скуп података који садржи мапе класа земљишног покривача. Одлучено је да се употреби јавно доступан скуп података о земљишном покривачу за централни регион Словеније [169]. Подаци о LC доступни су за 2019. годину и упарени су са одговарајућим подацима из Sentinel-2 сателита за исту годину. Подаци покривају релативно малу површину од 15 км x 15 км, али пошто је њихова резолуција 10 метара по пикселу, то одговара вредностима од 1500 x 1500 (2,25 милиона) које су опредељене за тренирање и валидацију предложеног екстрактора обележја. Ово је два реда величине више од броја узорака података који постоје за SOC у скуповима података који ће бити коришћени у истраживању, а LC подаци су у природној резолуцији Sentinel-2 сензора, док су доступни SOC подаци расути по огромним подручјима.

Пошто одређивање SOC нема смисла за неке врсте земљишног покривача (нпр. воде или вештачке површине), да би процес генерисања SOC мапа био ефикаснији и тачнији, потребан нам је и скуп података о земљишном покривачу који има много већу покривеност од оног који се користи за обуку екстрактора обележја, чак и ако резолуција није тако висока.

На срећу, Европа је континент са најширим спектром наднационалних LC мапа. Сада је доступно мноштво детаљних, висококвалитетних скупова података који пружају LC информације за европски континент, као што су: HILDA, CORINE, PELCOM, Urban Atlas итд.

Од свих европских скупова података LC, CORINE пружа земљишни покривач који је један од најпознатијих, најстаријих и најчешће коришћених [167]. Стога ће бити употребљен CLC (CORINE LC) за накнадну обраду мапа које генерише предложени приступ, водећи рачуна да су процене доступне само за класе земљишног покривача за које је модел обучен да обради.

7.1 Приступ процени органског угљеника у земљишту

Предложени приступ машинског учења за процену органског угљеника у земљишту састоји се од две фазе, од којих свака обавља одређену функцију како би произвела коначне процене SOC-а. Прва фаза укључује обуку стандардне сегментационе дубоке неуронске мреже са U-Net архитектуром. Ова DNN је тренирана да процени покривач земљишта посматраног региона на основу мултиспектралних слика са сателита Sentinel-2 и сегментационе мапе са стварним подацима. Истренирани DNN генерише LC процену за регион од интереса (енгл. region of interest - ROI). Ова процена се затим користи за издвајање вектора латентних обележја за сваки излазни пиксел враћањем са излазног слоја U-Net мреже и приступањем последњем доступном конволуционом слоју исте димензије као и излаз.

Друга фаза користи векторе обележја екстраховане на координатама за које постоје лабораторијски одређене вредности SOC-а. Парови вектора обележја и SOC мерења се користе за тренирање независног ML модела, који врши стварну SOC процену. У време закључивања, две фазе генеришу просторно непрекидан скуп података SOC процена за делове слике које обрађује U-Net. Ове процене за различите делове слике се затим спајају да би се направила мапа процена SOC-а за произвољни географски регион.

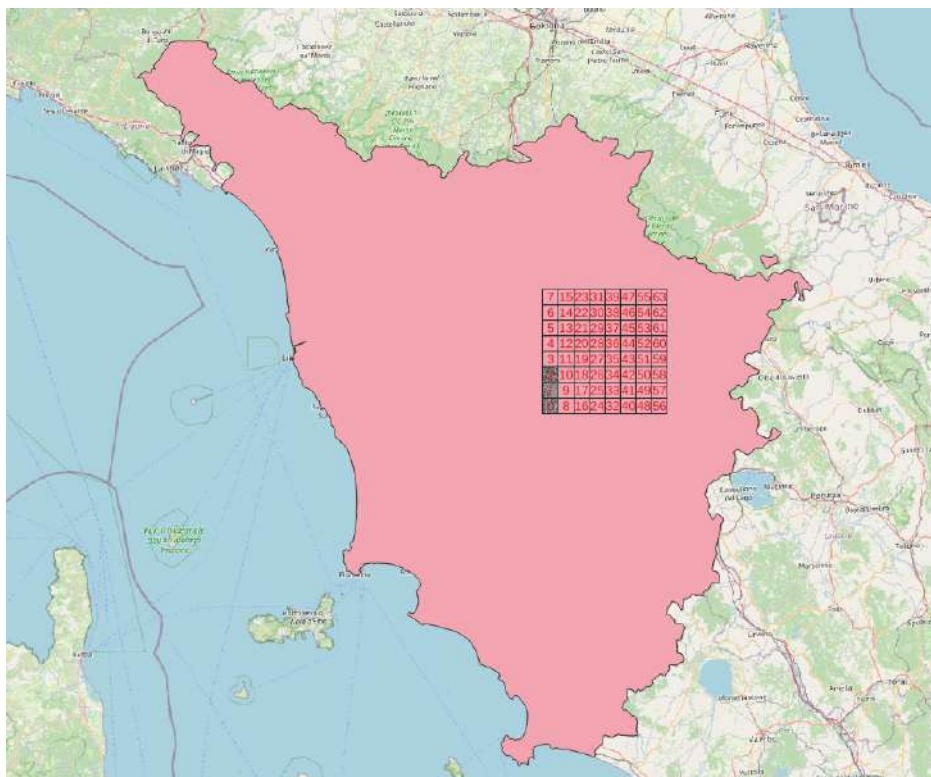
У оквиру истраживања је генерисан и приказан узорак мапе за регион Тоскане, Централна Италија, са резолуцијом од 10 метара, што је у складу са највећом доступном резолуцијом сателита Sentinel-2. Тоскана (видети слику 42.) покрива око 22.990 квадратних километара, обухватајући разноврстан спектар терена укључујући планине, брда и обалне равнице, што подразумева различите типове земљишта. Сваки тип земљишта показује различите садржаје органског угљеника, нудећи свеобухватно поље за SOC студије у различитим условима земљишта.

Тоскана је светски позната по пољопривредној производњи. Пољопривредни сектор Тоскане је глобално славан због свог квалитета и разноликости, од производње вина у Кјантију до маслињака који се простиру низ брда. Плодно земљиште у региону такође подржава широк спектар других усева, укључујући житарице, махунарке и поврће, што га чини каменом темељцем италијанске прехранбене и винске индустрије. С обзиром на виталну улогу органског угљеника у земљишту у одржавању плодности земљишта и промовисању одрживих пољопривредних пракси, Тоскана је изгледала као добар избор за тестирање предложеног модела и пружање додатне вредности из истраживања.

Додатак овој сложености је климатска разноликост Тоскане. Од благог медитеранског поветарца дуж њених обала до оштрије планинске климе, регион доживљава широку лепену временских услова. Климатска варијација директно утиче на разградњу органске материје и производњу биомасе, обе кључне детерминанте SOC. Стога, климатска варијабилност Тоскане омогућава дубинско испитивање SOC-а под различитим утицајима животне средине.

Саставни аспект подобности Тоскане за истраживање SOC-а је њена богата историја коришћења земљишта, детаљно документована током година. Ови историјски подаци пружају непроцењив увид у дугорочне SOC трендове и њихову корелацију са управљањем земљиштем и пољопривредним методологијама.

Штавише, разноврстан пејзаж Тоскане, заједно са различитим разграничењима коришћења земљишта, чини га примерним регионом за примене даљинске детекције у SOC студијама. Доступност сателитских података високе резолуције, као што су мисије Sentinel, омогућава детаљну и опсежну анализу SOC-а у целом региону.



Слика 44: Узорци улазних делова узетих за регион Тоскане

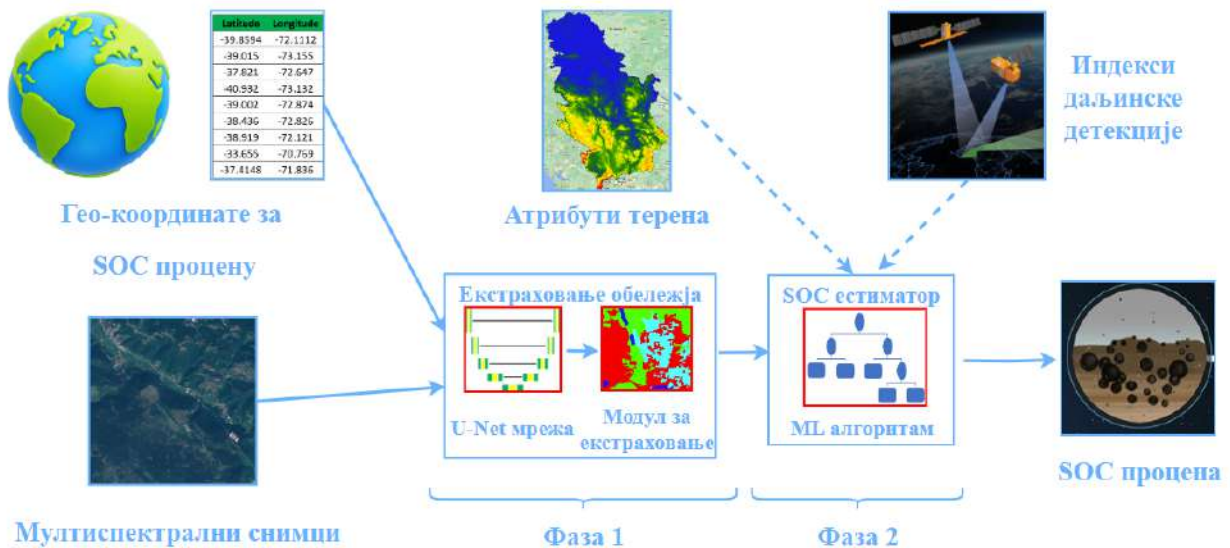
7.2 Фаза 1 – сегментациона дубока неуронска мрежа са U-Net архитектуром за процену LC

Фокус истраживања је стварање приступа који може да генерише мапу процена SOC-а исте резолуције као она коју пружају Sentinel-2 снимци. Међутим, количина података за тренирање који су доступни за овај задатак је ограничена, што може ометати перформансе најсавременијих модела машинског учења.

Поред тога, доступни извори података као што су LUCAS и CHLSOC скупови података су ретки, што онемогућава њихово коришћење за директан тренинг од почетка до краја за моделе машинског учења за задатке сегментације, који представљају основну технологију коју треба користити за ефикасно решавање проблема континуиране процене SOC.

Како би се превазишла наведена ограничења, у истраживању је предложена употреба трансферног учења. Прво се тренира мрежа за сегментацију, конкретно U-Net модел, да прецизно предвиди LC одређене географске локације на основу Sentinel-2 снимака. Слично приступу Гиршика и сарадника [21], затим се издвајају латентна обележја које је научила мрежа да би послужили као улазни подаци за ML модел друге фазе за процену SOC. Пошто сегментациона мрежа која се користи у првој фази чува просторне односе који постоје на снимку, то омогућава да се генерише просторно континуирано предвиђање нивоа SOC за дати регион од интереса, узимајући ове односе у обзир и побољшавајући тачност добијене процене.

На Слици 45. је приказан блок-дијаграм предложеног приступа.



Слика 45: Предложени приступ процени SOC

Како би се процениле перформансе целог приступа употребљена је комбинација скупова података CHLSOC, LUCAS 2015 и 2018, који су евалуирани употребом стандардних метрика за проблем регресије: средња апсолутна грешка, средња квадратна грешка, средња апсолутна процентуална грешка и коефицијент детерминације. Последња два пружају увид у то колико је велика грешка у односу на лабораторијски одређене вредности и колики је део варијансе података објашњен моделом.

7.3 Тренирање U-Net модела на јавно доступним подацима LULC за Словенију

У истраживању главни циљ је да се процени садржај органског угљеника у земљишту употребом сателитских снимака и техника машинског учења. Међутим, количина података за тренирање доступних за овај задатак је често ограничена, што може утицати на перформансе модела машинског учења.

Како би се превазишло наведено ограничење, идеја је да се употребе подаци о коришћењу земљишта и земљишном покривачу као фаза претходне обраде за модел ма-

шинског учења за процену SOC. Подаци LULC-a pružaју информације о типовима земљишног покривача у датом региону, што може бити корисно за предвиђање нивоа SOC.

Док су LULC подаци генерално доступнији од SOC података, још увек их није лако пронаћи као континуирану мапу у високим резолуцијама. Конкретно, скупови података као што су CORINE и други који обезбеђују LULC податке често су и сами резултати процедура машинског учења и не представљају добру основу за тренирање других алгоритама машинског учења. Подаци LULC-a доступни су за 2019. годину и упоређени су са одговарајућим подацима из Sentinel-2 за исту годину. Подаци LULC покривају област од 1500 x 1500 пиксела, што одговара 15 км x 15 км, са резолуцијом од 10 метара по пикселу, што је у складу са највећом доступном резолуцијом Sentinel-2 сателита.

Како би се креирао LC класификатор, у овом истраживању, коришћена је U-Net архитектура са улазом од 64 x 64 пиксела. Ово је обезбедило 1436 стварних LC вредности у једној димензији дела слике, тј. више од 2 милиона тренинг тачака по делу слике. Цео скуп података ео-learn који се користи за обуку и валидацију нашег модела садржи око 235 милиона узорака (тј. Sentinel-2 пиксела са стварном LC ознаком).

7.4 Прикупљање, обрада и чување сателитских снимака

Да би се направио високо технички и детаљан тока рада SOC процене, коришћење Sentinel-2 снимака кроз употребу програмског језика Python укључује сложене процесе, конфигурације и рачунарске технике. Овај ток рада је артикулисан кроз низ Python модула од којих сваки доприноси аквизицији, претходној обради и припреми сателитских снимака за процену SOC-a. Специфичности ових операција засноване су на техничкој реалности руковања мултиспектралним сателитским снимцима и припремама за анализу методама дубоког учења.

Python је свестран програмски језик високог нивоа познат по својој једноставности и читљивости, што га чини идеалним избором за почетнике и професионалце. Његова обимна стандардна библиотека и огроман екосистем пакета развијених од стране независних произвођача задовољавају широк спектар примена, од web развоја и анализе података до вештачке интелигенције и научног рачунарства.

Python програмски језик је успоставио доминанту улогу у области вештачке интелигенције због свог богатог екосистема библиотека и оквира прилагођених за машинско учење, дубоко учење и задатке обраде природног језика. Библиотеке као што су TensorFlow, PyTorch и scikit-learn пружају моћне алате за изградњу и обуку AI модела. У овом истраживању коришћене су библиотеке TensorFlow, scikit-learn, NumPy, као и низ самостално развијених функција неопходних за омогућавање имплементације предложеног приступа.

TensorFlow је библиотека отвореног кода за машинско учење коју је развио Google Brain за изградњу и тренирање различитих модела дубоког учења. Пружа флексибилан екосистем за имплементацију неуронских мрежа на низу платформи, од десктоп рачунара до мобилних уређаја и дистрибуираних система великих размера. Са својом обимном документацијом, снажном подршком заједнице и рачунарским могућностима високих перформанси, TensorFlow библиотека је постала незаобилазан алат у области вештачке интелигенције, омогућавајући истраживачима и програмерима да се позабаве

сложеним проблемима у областима дубоког учења.

Scikit-learn је разноврсна библиотека за машинско учење за Python програмски језик која је прилагођена кориснику, која пружа ефикасне алате за рударење и анализу података. Нуди широк спектар алгоритама за учење са надзором, као и за учење без надзора. Такође, садржи алате за избор модела, евалуацију модела као и претходну обраду података. Са својим интуитивним интерфејсом и обимном документацијом, scikit-learn нашироко користе истраживачи, научници и практичари за изградњу предиктивних модела и решавање различитих задатака машинског учења.

NumPy представља фундаменталну библиотеку за Python програмски језик намењену за нумеричко рачунање, која пружа подршку за велике, вишедимензионалне низове и матрице, заједно са колекцијом математичких функција за ефикасан рад на овим низовима. Служи као градивни блок за различите научне и инжењерске примене, нудећи функционалности за задатке као што су линеарна алгебра, Фуријеове трансформације, генерисање случајних бројева и још много тога. Са својом интуитивном синтаксом и могућностима високих перформанси, NumPy значајно убрзава нумеричка израчунавања и задатке манипулације подацима у Python-у.

Sentinel-2 сателити играју кључну улогу тако што пружају мултиспектралне слике у спектру од 13 опсега. Ови опсежи обухватају видљиве, блиске инфрацрвене и краткоталасне инфрацрвене делове спектра, са просторним резолуцијама од 10м, 20м и 60м по пикселу, задовољавајући широк спектар потреба за посматрањем.

У оквиру овог истраживања развијени су модули за прикупљање, обраду и чување сателитских снимака који су дизајнирани да олакшају аутоматизовано преузимање неопходних снимака путем коришћења услуге из платформе Sentinel Hub. Процес укључује неколико критичних корака и конфигурација како би се обезбедило прецизно прибављање података прилагођених захтевима истраживања.

Sentinel Hub представља платформу засновану на облаку која поједностављује приступ подацима посматрања Земље, укључујући огромне скупове података које генеришу сателити Sentinel-2. Нудећи интерфејс за програмирање апликације (енгл. Application Programming Interface - API) за испитивање и преузимање сателитских снимака, Sentinel Hub омогућава истраживачима и програмерима да интегришу податке у реалном времену и историјске податке у своје апликације и радне токове.

У Python модулима дефинисани су процеси одговорни за преузимање делова LULC снимака у Словенији који се у накнадним процесима у оквиру Фазе 1 користе за тренирање U-Net сегментационе мреже, као и за преузимање Sentinel-2 мултиспектралних слика.

Употребљава се Sentinel Hub библиотека за конфигурисање захтева за жељене Sentinel-2 снимке. Врши се аутентификација коришћењем ID-а клијента и тајног ID кода. Овај механизам за аутентификацију успоставља безбедну сесију са услугом Sentinel Hub, дајући модулима могућност да ефикасно шаљу упите у базу података сателитских снимака. У оквиру модула дефинише се листа делова снимака (тј. географских области од интереса) и спроводи се поступак итерације над дефинисаном листом како би се преузели одговарајући сателитски снимци. Преузимање снимака се обавља путем SentinelHubRequest-а, у којем се прецизирају жељене колекције сателитских података, у овом случају Sentinel-2 L1C, жељени опсежи (свих 13 опсега Sentinel-2 сателита), као и излазни формат.

Након конфигурисања захтева за снимке и њиховог преузимања обезбеђује се да се сателитски снимци ефикасно преузимају и чувају за даљу анализу. Ефикасност и ефективност Python модула у управљању проналажењем и складиштењем сателитских снимака су од суштинског значаја, с обзиром на велике количине података укључених у примене за даљинску детекцију и потребу за висококвалитетним подацима у моделима дубоког учења.

У оквиру модула коришћено је дефинисање ограничавајућих оквира (енгл. Bounding Boxes). Ограничавајући оквир је правоугаони оквир који представља границе објекта или области унутар слике. У географским применама, обично се дефинише помоћу два скупа координата: горњи леви (минимална географска дужина, максимална ширина) и доњи десни (максимална географска дужина, минимална географска ширина) углови. Ове координате се користе за оцртавање просторног обима области од интереса на површини Земље, ефективно ограничавајући регион за који се подаци извлаче или анализирају.

У овом истраживању ограничавајући оквири се користе да дефинишу географске опсеге за које се захтевају сателитски снимци Sentinel-2. Библиотека Sentinel Hub омогућава спецификацију ограничавајућих оквира као део захтева за преузимање снимака. Координате за ограничавајуће оквири усклађене су са глобалним координатним системом, у овом случају са светским геодетским системом 1984 (енгл. World Geodetic System 1984 – WGS 84), омогућавајући прецизан избор региона од интереса из огромних скупова података доступних преко Sentinel-2 сателита.

Такође, за потребе овог истраживања, вршена је конверзија географских координата из WGS 84 у универзални трансверзални Меркатор (енгл. Universal Transverse Mercator - UTM) координатни систем.

WGS 84 је глобални референтни систем за геодезију, навигацију и мапирање. Он обезбеђује стандардни координатни оквир за планету Земљу, одређује фиксну референтну тачку површине и успоставља глобални датум укључујући географску ширину, дужину и координате надморске висине. WGS 84 се широко користи у технологији глобалног система позиционирања (енгл. Global Positioning System – GPS) и стандардни је координатни систем који користи GPS.

Насупрот томе, UTM систем дели свет на низ лонгитудалних зона од 6 степени, од којих је свака пројектована коришћењем попречне Меркаторове пројекције. Ово резултира мрежом у којој су координате дате у метрима северно и источно од почетка зоне, што омогућава да мерења удаљености буду једноставнија и тачнија него у степенима географске ширине и дужине. Свака зона има своје параметре пројекције, чинећи UTM координате специфичним за зону.

UTM координате су у линеарним метрима, што поједностављује прорачуне удаљености и просторну анализу на малим површинама (пречника до неколико стотина километара). Ово је посебно корисно у прецизној пољопривреди, премеравању земљишта и пројектима праћења животне средине као што је SOC процена, где су прецизна мерења удаљености критична. Док су WGS 84 координате одличне за глобално позиционирање и мапирање, UTM координате су погодније за локалне и регионалне студије. Конверзија омогућава лакшу интерпретацију и манипулацију подацима приликом рада на одређеним локацијама или регионима.

Многи GIS алати и софтвери преферирају или захтевају податке у пројектованом

координатном систему као што је UTM за прецизну просторну анализу и мапирање. Конвертовање WGS 84 координата у UTM олакшава интеграцију сателитских снимака и других просторних података са GIS алатима, побољшавајући радни ток за SOC процену. Конкретно у овом истраживању је употребљаван QGIS алат за продукцију SOC континуиране мапе.

7.4.1 Учитавање, претходна обрада и припрема сателитских снимака, као и кореспондирајућих стварних података о терену (тј. маски стварних вредности)

Модул за учитавање, претходну обраду и припрему сателитских снимака, као и кореспондирајућих стварних података о терену (тј. маски стварних вредности), имплементира процесе који директно утичу на перформансе и ефикасност модела дубоког учења који се користе у каснијим фазама истраживања. Почетак процеса подразумева учитавање претходно сачуваних Sentinel-2 снимака и повезаних стварних података о терену, који обично укључују мапе класификације земљишног покривача.

Врши се пажљиво повезивање сваке слике са одговарајућим истинитим подацима о самом терену (маскама). Ово повезивање је кључно за тренирање модела да препознају различите типове земљишног покривача, што је кључни корак у идентификацији и конструкцији латентних обележја релевантних за процену SOC.

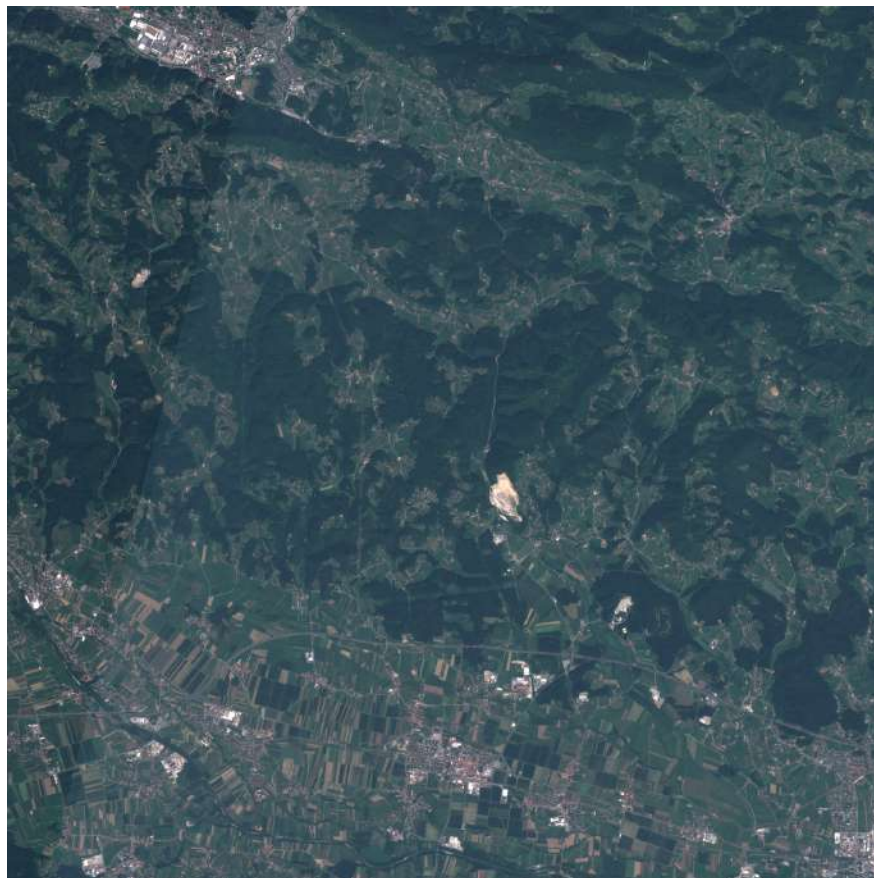
Улазна слика величине 1500 x 1500 пиксела (приказана на Слици 46.) се дели на мање делове (енгл. patches) величине 64 x 64 пиксела над којима се врши сама обрада.

Ова специфична величина мањих делова је изабрана како би се ускладила са изворним блоковима резолуције Sentinel-2 сателита, омогућавајући да се приликом анализе искористе инхерентне могућности просторне резолуције сателита. Поделом слике у мање делове осигурава се да сваки patch садржи довољно детаља за прецизну класификацију земљишног покривача, а да није толико велик да унесе непотребну сложеност рачунања.

Предност примењеног приступа се огледа у томе што омогућава да сваки patch садржи количину података којом се може управљати, осигуравајући да модел може да идентификује и учи из нијансираних варијација унутар пејзажа, што је кључно за разликовање различитих типова земљишног покривача. Такође, мање и конзистентне величине patch-ева помажу у управљању захтевима за коришћењем меморије и процесорском снагом, омогућавајући моделу да ради на доступном хардверу без угрожавања перформанси.

Значајан део функционалности примењених у модулима посвећен је претходној обради Sentinel-2 снимака. С обзиром на мултиспектралну природу ових снимака, модул имплементира специфичне кораке претходне обраде са циљем да стандардизује податке, чинећи их погоднијим за анализу помоћу модела дубоког учења.

Ово укључује процедуре нормализације које укључује примену специфичних средњих вредности и вредности стандардне девијације за сваки опсег, изведених из скупа података, да би се узеле у обзир варијације изазване атмосферским условима, разликама сензора и разликама у осветљењу. Ово се постиже одузимањем средње вредности скупа података и дељењем са стандардном девијацијом за сваки опсег, на основу претходне анализе скупа података Sentinel-2. На овај начин се обезбеђује да улазни подаци у модел дубоког учења одржавају конзистентност на различитим снимцима што олакшава



Слика 46: RGB улазна слика дела територије Републике Словеније величине 1500 x 1500 пиксела

прецизније издвајање и анализу обележја.

Нормализација и стандардизација мултиспектралних слика директно утичу на способност модела да разликује различите типове земљишног покривача, што је предуслов за тачну процену SOC. Класификација земљишног покривача, заузврат, даје информације о избору подручја за предвиђање SOC-а, наглашавајући важност прецизног и поузданог LC мапирања.

Да би се побољшала робусност модела и ублажили ефекти прекомерног прилагођавања, примењене су технике повећавања података (енгл. *data augmentation techniques*). Ове технике генеришу додатне примере за тренирање кроз геометријске трансформације као што су ротације, преокрети и скалирање оригиналних слика. То подразумева да се слике ротирају под одређеним угловима (45° , 90° , 180° , 270°) да се окрећу хоризонтално/вертикално да би се симулирали различити услови гледања, као и насумично исецање оригиналних слика.

Проширујући разноликост скупа података за тренирање на овај начин помаже се моделу да боље генерализује нове, тј. податке које никада није видео, што је критичан фактор за успех процене SOC-а у различитим географским регионима и условима.

Такође у овом кораку се врши подела учитаних и претходно обрађених података у скупове за тренирање и валидацију. Ова подела обезбеђује да је репрезентативни узорак

података резервисан за непристрасну процену способности модела да прецизно класификује LC и, сходно томе, ефективно процени вредности SOC. Ово даље омогућава фино подешавање параметара модела на основу повратних информација о валидацији.

Припрема и претходна обрада података су темељ у постизању функционалне SOC процене. Тачност SOC процена изведених из модела дубоког учења у великој мери зависи од квалитета, разноврсности и репрезентативности улазних података. Имплементацијом строгих поступака предобраде и увећавања улазног скупа података осигурава се да су снимци Sentinel-2 у оптималном стању за екстракцију и анализу обележја који ће бити извршени у Фази 2 предложеног приступа.

Повећавање скупа података и припрема уравнотежених скупова за тренирање и валидацију је кључна за развој робусних модела. Ови модели морају да раде доследно у различитим условима и локацијама да би били ефикасни у задацима процене SOC. Унапређивањем квалитета и разноврсности скупа података значајно се доприноси циљу стварања тачних, скалабилних и генерализујућих модела за процену SOC-а из података даљинске детекције.

7.4.2 U-Net сегментациони модел

У овом истраживању предложена је конструкција U-Net архитектуре која користи унапред обучени VGG16 модел као енкодер, прилагођен за задатак сегментације слике. Ова архитектура модела је посебно погодна за мапирање карактеристика високе резолуције, што је критично за задатке као што су класификација земљишног покривача и, што представља почетни корак ка прецизној процени органског угљеника у земљишту на основу снимака добијених путем даљинске детекције. Употребом VGG16, познате архитектуре конволуционе неуронске мреже, као основу за део енкодера U-Net, омогућава се коришћење VGG16 мреже за издвајања обележја, као и њена способност да разуме и класификује садржаје на сателитским снимцима. Енкодер се састоји од више конволуционих слојева и слојева максималног скупљања који прогресивно смањују узорковање улазне слике, омогућавајући мрежи да ухвати сложене карактеристике на различитим скалама.

У овом истраживању је коришћена претходно истренирана VGG16 мрежа на ImageNet Large Scale Visual Recognition Challenge (ILSVRC) скупу података која је увезена из TensorFlow библиотеке за DL.

ImageNet је визуелна база података великих размера дизајнирана за коришћење у истраживању препознавања визуелних објеката. Садржи преко 14 милиона слика, класификованих у више од 20.000 категорија. ImageNet је био кључан у унапређењу истраживања компјутерског вида, посебно кроз његову употребу у изазову ImageNet Large Scale Visual Recognition Challenge (ILSVRC), где се алгоритми такмиче да постигну најниже стопе грешака у класификацији.

Најшире познати подскуп ImageNet-а, је ILSVRC скуп података, који састоји се од 1,2 милиона слика за тренирање, 50.000 слика за валидацију и 100.000 слика за тестирање, које обухватају 1.000 категорија објеката.

Како се VGG16 мрежа традиционално користи са 3-каналним (RGB) сликама, било је неопходно да се изврши прилагођавање улазне слике са различитим бројем канала, како би се прилагодила мултиспектралној природи Sentinel-2 слика. Ова адаптација је

неопходна како би модел могао да обради јединствене информације које носе различити спектрални опсези.

Тежине из претходно истренираног VGG16 модела се поново користе за конволуционе слојеве енкодера, имплементирајући приступ трансфер учења. Овај метод значајно побољшава ефикасност учења и перформансе модела, посебно када су означени подаци ограничени или скупи за добијање, као што је често случај са скуповима података даљинске детекције.

За декодер део U-Net мреже имплементира се серија операција повећавања узорковања и конволуције. Прескачуће везе се користе за спајање обележја са слојева енкодера у одговарајуће слојеве декодера. Овај дизајн помаже моделу да поврати просторне информације изгубљене током смањења узорковања, омогућавајући прецизну локализацију и детаљне мапе сегментације које су од виталног значаја за тачну класификацију земљишног покривача.

Последњи слој модела је конволуцијски слој са софтмакс функцијом активације, дизајниран да класификује сваки пиксел у једну од неколико категорија. Број филтера у овом слоју одговара броју класа (типова земљишног покривача) које модел може разликовати (10 класа).

На Слици 47. приказана је детаљна архитектура предложеног модела за LC класификацију.

Енкодер део модела почиње са улазним слојем дизајнираним да прима мултиспектралне слике облика (64, 64, 13), где 13 одговара броју спектралних опсега у Sentinel-2 снимцима. Овај слој не садржи параметре за обуку.

Енкодер се састоји од неколико конволуционих блокова изведених из архитектуре VGG16, од којих сваки садржи Conv2D слојеве праћене MaxPooling2D слојем. Ови блокови прогресивно смањују узорковање улазне слике док повећавају дубину мапа обележја. Овај дизајн омогућава моделу да ухвати широк спектар обележја од основних текстура и ивица у почетним слојевима до сложених шара и делова објеката у дубљим слојевима. Конволуцијски слојеви користе 3 x 3 филтере, карактеристичну одлику VGG16 мреже, омогућавајући ефикасно издвајање обележја са релативно малим бројем параметара по слоју.

Последњи блок VGG16 енкодера блок block5_conv3 делује као мост између енкодера и декодера. Он даље обрађује мапе обележја, припремајући их за повећање узорковања у декодеру.

Декодер користи транспоновану конволуцију (Conv2DTranspose) за подизање узорковања, ефективно преокренувши просторно доње узорковање које изводи енкодер. Сваки корак подизања узорковања је праћен конкатенацијом (Concatenate) тј. спајањем са одговарајућом мапом обележја из енкодера (skip везе), чиме се поново уносе просторне информације изгубљене током узорковања на нижем нивоу. Овај процес је критичан за реконструкцију мапе сегментације са прецизним границама.

Након сваке конкатенације, модел примењује два Conv2D слоја са ReLU активацијом, након чега следи BatchNormalization. Ова комбинација омогућава моделу да прецизира обележја која су претходно узоркована, побољшавајући способност модела да тачно локализује и класификује сваки пиксел.

Модел се завршава са излазним слојем Conv2D са 10 филтера и софтмакс функцијом активације. Овај слој мапира мапу обележја 64 x 64 у 10 класа, које одговарају разли-

Модел: "LC класификатор"

| Слој (тип) | Излазни облик | # Параметара | Повезан са слојем |
|---------------------------------|----------------------|--------------|--|
| input_1 (InputLayer) | [(None, 64, 64, 13)] | 0 | |
| block1_conv1 (Conv2D) | (None, 64, 64, 64) | 7552 | input_1[0][0] |
| block1_conv2 (Conv2D) | (None, 64, 64, 64) | 36928 | block1_conv1[0][0] |
| block1_pool (MaxPooling2D) | (None, 32, 32, 64) | 0 | block1_conv2[0][0] |
| block2_conv1 (Conv2D) | (None, 32, 32, 128) | 73856 | block1_pool[0][0] |
| block2_conv2 (Conv2D) | (None, 32, 32, 128) | 147584 | block2_conv1[0][0] |
| block2_pool (MaxPooling2D) | (None, 16, 16, 128) | 0 | block2_conv2[0][0] |
| block3_conv1 (Conv2D) | (None, 16, 16, 256) | 295168 | block2_pool[0][0] |
| block3_conv2 (Conv2D) | (None, 16, 16, 256) | 590080 | block3_conv1[0][0] |
| block3_conv3 (Conv2D) | (None, 16, 16, 256) | 590080 | block3_conv2[0][0] |
| block3_pool (MaxPooling2D) | (None, 8, 8, 256) | 0 | block3_conv3[0][0] |
| block4_conv1 (Conv2D) | (None, 8, 8, 512) | 1180160 | block3_pool[0][0] |
| block4_conv2 (Conv2D) | (None, 8, 8, 512) | 2359808 | block4_conv1[0][0] |
| block4_conv3 (Conv2D) | (None, 8, 8, 512) | 2359808 | block4_conv2[0][0] |
| block4_pool (MaxPooling2D) | (None, 4, 4, 512) | 0 | block4_conv3[0][0] |
| block5_conv1 (Conv2D) | (None, 4, 4, 512) | 2359808 | block4_pool[0][0] |
| block5_conv2 (Conv2D) | (None, 4, 4, 512) | 2359808 | block5_conv1[0][0] |
| block5_conv3 (Conv2D) | (None, 4, 4, 512) | 2359808 | block5_conv2[0][0] |
| conv2d_transpose (Conv2DTranspo | (None, 8, 8, 512) | 1049088 | block5_conv3[0][0] |
| concatenate (Concatenate) | (None, 8, 8, 1024) | 0 | conv2d_transpose[0][0] block4_conv3[0][0] |
| conv2d (Conv2D) | (None, 8, 8, 512) | 4719104 | concatenate[0][0] |
| batch_normalization (BatchNorma | (None, 8, 8, 512) | 2048 | conv2d[0][0] |
| activation (Activation) | (None, 8, 8, 512) | 0 | batch_normalization[0][0] |
| conv2d_1 (Conv2D) | (None, 8, 8, 512) | 2359808 | activation[0][0] |
| batch_normalization_1 (BatchNor | (None, 8, 8, 512) | 2048 | conv2d_1[0][0] |
| activation_1 (Activation) | (None, 8, 8, 512) | 0 | batch_normalization_1[0][0] |

| | | | |
|---------------------------------|---------------------|---------|--|
| conv2d_transpose_1 (Conv2DTrans | (None, 16, 16, 256) | 524544 | activation_1[0][0] |
| concatenate_1 (Concatenate) | (None, 16, 16, 512) | 0 | conv2d_transpose_1[0][0] block3_conv3[0][0] |
| conv2d_2 (Conv2D) | (None, 16, 16, 256) | 1179904 | concatenate_1[0][0] |
| batch_normalization_2 (BatchNor | (None, 16, 16, 256) | 1024 | conv2d_2[0][0] |
| activation_2 (Activation) | (None, 16, 16, 256) | 0 | batch_normalization_2[0][0] |
| conv2d_3 (Conv2D) | (None, 16, 16, 256) | 590080 | activation_2[0][0] |
| batch_normalization_3 (BatchNor | (None, 16, 16, 256) | 1024 | conv2d_3[0][0] |
| activation_3 (Activation) | (None, 16, 16, 256) | 0 | batch_normalization_3[0][0] |
| conv2d_transpose_2 (Conv2DTrans | (None, 32, 32, 128) | 131200 | activation_3[0][0] |
| concatenate_2 (Concatenate) | (None, 32, 32, 256) | 0 | conv2d_transpose_2[0][0] block2_conv2[0][0] |
| conv2d_4 (Conv2D) | (None, 32, 32, 128) | 295040 | concatenate_2[0][0] |
| batch_normalization_4 (BatchNor | (None, 32, 32, 128) | 512 | conv2d_4[0][0] |
| activation_4 (Activation) | (None, 32, 32, 128) | 0 | batch_normalization_4[0][0] |
| conv2d_5 (Conv2D) | (None, 32, 32, 128) | 147584 | activation_4[0][0] |
| batch_normalization_5 (BatchNor | (None, 32, 32, 128) | 512 | conv2d_5[0][0] |
| activation_5 (Activation) | (None, 32, 32, 128) | 0 | batch_normalization_5[0][0] |
| conv2d_transpose_3 (Conv2DTrans | (None, 64, 64, 64) | 32832 | activation_5[0][0] |
| concatenate_3 (Concatenate) | (None, 64, 64, 128) | 0 | conv2d_transpose_3[0][0] block1_conv2[0][0] |
| conv2d_6 (Conv2D) | (None, 64, 64, 64) | 73792 | concatenate_3[0][0] |
| batch_normalization_6 (BatchNor | (None, 64, 64, 64) | 256 | conv2d_6[0][0] |
| activation_6 (Activation) | (None, 64, 64, 64) | 0 | batch_normalization_6[0][0] |
| conv2d_7 (Conv2D) | (None, 64, 64, 64) | 36928 | activation_6[0][0] |
| batch_normalization_7 (BatchNor | (None, 64, 64, 64) | 256 | conv2d_7[0][0] |
| activation_7 (Activation) | (None, 64, 64, 64) | 0 | batch_normalization_7[0][0] |
| conv2d_8 (Conv2D) | (None, 64, 64, 10) | 650 | activation_7[0][0] |

Укупан број параметара: 25.868.682
Параметри који се могу тренирати: 25.864.842
Параметри који се не могу тренирати: 3.840

Слика 47: Архитектура модела за LC класификацију

читим типовима земљишног покривача неопходним за процену SOC-а. Сваки пиксел у излазу 64 x 64 одговара предвиђеној ознаци класе.

Модел има укупно 25,868,682 параметра, значајан број који одражава сложеност модела и капацитет да научи детаљна обележја из улазних слика.

Број параметара који се могу тренирати је 25,864,842. Скоро сви параметри у моделу се могу тренирати. Ови параметри се прилагођавају током процеса тренирања како би се функција губитка свела на минимум.

Број параметара који се не могу тренирати је 3.840. Мали део параметара се не може тренирати, јер то су параметри који се налазе у слојевима BatchNormalization. Ови параметри се користе за нормализацију уноса сваког активационог слоја, смањујући унутрашње коваријантно померање и побољшавајући стабилност тренирања. Иако се не прилагођавају пропагацијом уназад, они играју кључну улогу у перформансама модела.

За тренирање модела коришћен је специјализовани Huawei сервер Atlas 800 (модел 3010) који представља моћну AI рачунарску платформу дизајнирану да убрза различите радне задатке, укључујући тренирање и закључивање AI модела. Интегрише 2 Ascend 910 NPU (енгл. Neural Processing Units - NPU), омогућавајући паралелну обраду високих перформанси за задатке дубоког учења. С друге стране, Huawei Ascend 910 NPU је самостални чип оптимизован за AI примене, који нуди импресивне могућности

обrade задатака као што су обрада природног језика, компјутерски вид и још много тога.

Модел је састављен са адам оптимизатором (енгл. Adaptive Moment Estimation - Adam), категоричком унакрсном ентропијом као функцијом губитка и као метрика за праћење прецизности модела изабрана је категоричка прецизност.

Адам оптимизатор представља алгоритам који израчунава стопе адаптивног учења за сваки параметар. За разлику од оптимизатора фиксне стопе учења, адам оптимизатор прилагођава стопу учења за сваки параметар, помажући да се ефикасно креће кроз крајолик оптимизације. Балансира предности експоненцијално опадајућег просека прошлих градијената (замах) и градијената на квадрат (скалирање), убрзавајући конвергенцију, посебно у контексту сложених или бучних градијента, што је уобичајено у задацима дубоког учења.

За LC класификацију, која укључује сложене, високодимензионалне улазе и потенцијално неконвексне пејзаже оптимизације, механизам адаптивне брзине учења је посебно користан. Помаже у проналажењу оптималних тежина ефикасније од оптимизатора са константном стопом учења, што потенцијално доводи до бољих перформанси модела са мање ручног подешавања брзине учења.

Категорична унакрсна ентропија је функција губитка која се често користи у проблемима класификације са више класа. Она мери несличност између праве дистрибуције ознака и предвиђања, ефективно квантификујући колико добро излазна дистрибуција вероватноће модела предвиђа праву дистрибуцију ознака. Посебно је погодна за задатке класификације са више класа, зато што додељује високу казну за предвиђања која су сигурна, али погрешна. Кажњавањем логаритма предвиђених вероватноћа, подстиче модел не само да правилно класификује, већ и да даје дистрибуције вероватноће које одражавају праву поузданост модела.

С'обзиром на циљ истраживања да се класификују типови земљишног покривача као корак ка процени SOC-а, коришћење категоричке унакрсне ентропије је прикладно и ефикасно. Осигурава се да предвиђања модела нису само тачна у смислу класе, већ и поуздана у предвиђеним вероватноћама, што је кључно за нијансирано разумевање и доношење одлука на основу излазних података модела.

Категоријска прецизност је метрика која израчунава проценат тачно предвиђених ознака од свих предвиђања. За проблеме класификације више класа, то је директна и интуитивна мера перформанси модела.

Нуди јасан и директан начин за процену перформанси класификације модела, пружајући увид у то колико често је модел тачан у предвиђању сваке класе. Пошто је лака за разумевање и тумачење, омогућава брзу процену и поређење перформанси модела у различитим итерацијама тренирања или архитектурама модела.

Коришћење категоричке прецизности као метрике у овом истраживању пружа директну меру тога колико добро предложени модел обавља задатак класификације земљишног покривача, што је критична компонента процене SOC-а. Помаже у праћењу напретка тренирања и процени ефикасности побољшања модела или подешавања параметара.

Приликом тренирања модела за LC класификацију кориштени су методи ModelCheckpoint, ReduceLROnPlateau и EarlyStopping из TensorFlow библиотеке.

ModelCheckpoint омогућава континуирано чување модела или тежина, у различи-

тим тачкама током тренирања. Може се подесити да сачува само најбољи модел на који се наиђе током тренинга, у складу са перформансама на одређеној метрици. Обезбеђује да се најбоља верзија модела не изгуби, посебно у дугим тренирањима где перформансе модела могу да варирају. Такође, омогућава да се тренирање настави из последњег сачуваног стања у случају да је процес тренирања прекинут, штедећи време и рачунарске ресурсе.

ReduceLROnPlateau прати одређену метрику, и ако се не види побољшање за унапред дефинисан број епоха, смањује стопу учења за вредност дефинисаног фактора. Ово је корисно из разлога што смањење стопе учења може помоћи моделу да изађе из платоа у достигнутим перформансама или да изврши фина прилагођавања како би минимизирао губитак.

С обзиром на комплексност задатка LC класификације, где модел може брзо да достигне плато због сложених образаца и односа у подацима, коришћење ReduceLROnPlateau може значајно утицати на перформансе модела. Динамичко прилагођавање брзине учења омогућава нијансираније учење како се модел приближава конвергенцији, што потенцијално доводи до боље тачности и стабилности модела.

EarlyStopping зауставља процес тренирања када праћени показатељ престане да се побољшава након одређеног броја епоха. То је директан начин да се спречи прекомерно прилагођавање тако што се не дозвољава моделу да настави са учењем када престане да прави значајна побољшања. Штеди рачунарске ресурсе прекидањем процеса тренирања када даље учење престане да доноси користи.

Као скуп за тренирање модела за LC класификацију употребљено је 70 процената скупа података ео-леарн, док је преосталих 30 процената коришћено за евалуацију перформанси модела.

За хиперпараметре модела LC класификације одређене су следеће вредности: величина серије (енгл. Batch Size) износи 32, број епоха (енгл. epochs) је 100, и почетна стопа учења (енгл. learning rate) од 0.001 .

Величина серије се односи на број узорака за тренирање који се користе у једној итерацији тренирања модела. Другим речима, то је величина подскупа података који модел обрађује пре ажурирања интерних параметара (тј. тежина).

Мања величина серије је ефикаснија за меморију, омогућавајући обуку на великим скуповима података без потребе за великим хардверским ресурсима. Међутим, премала величина серије може довести до нестабилног тренирања и проблема са конвергенцијом. Веће величине серија нуде тачнију процену градијента, али могу довести до јаза у генерализацији. Мање серије дају бучније процене градијента, које могу помоћи да се избегну локални минимуми, што потенцијално доводи до боље генерализације. Веће серије могу убрзати процес тренирања због мањег броја ажурирања потребних по епохи. Међутим, проналажење праве равнотеже је кључно, јер превелика величина серије може довести до лоших перформанси модела.

Величина серије од 32 је уобичајен избор, балансирајући потребу за ефикасношћу рачунара са предностима бучних процена градијента за избегавање локалних минимума. То је разноврсна величина која има тенденцију да добро функционише у различитим задацима и скуповима података, што је чини разумним избором за LC класификацију са снимака добијених даљинском детекцијом.

Епоха представља један комплетан пролаз кроз цео скуп података за тренирање. То

је јединица мере за колико пута ће алгоритам учења радити кроз цео сет за тренирање.

Више епоха значи да модел има више могућности да учи из читавог скупа података, што потенцијално доводи до бољих перформанси јер се тежине прилагођавају на основу кумулативне грешке током многих пролаза. Тренирање током превише епоха може довести до тога да се модел претерано прилагоди, где он превише добро учи податке за тренирање, укључујући и буку, што резултира лошом генерализацијом на нове, односно податке које види први пут. Број епоха директно утиче на време тренирања и коришћење ресурса. Више епоха захтева више рачунарског времена и снаге.

Постављање епоха на 100 је свестан избор како би се моделу дало довољно могућности за учење из података. С обзиром на сложеност LC класификације и варијабилност у подацима даљинске детекције, ово омогућава темељно учење. Међутим, важно је пратити знакове претераног прилагођавања и применити стратегије као што је рано заустављање да бисте зауставили тренирање када перформансе валидације достигну платое или деградирају.

Стопа учења је хиперпараметар који контролише колико треба променити тежине модела као одговор на процењену грешку сваки пут када се тежине модела ажурирају. То је кључни фактор у конвергенцији процеса тренирања на минимум функције губитка.

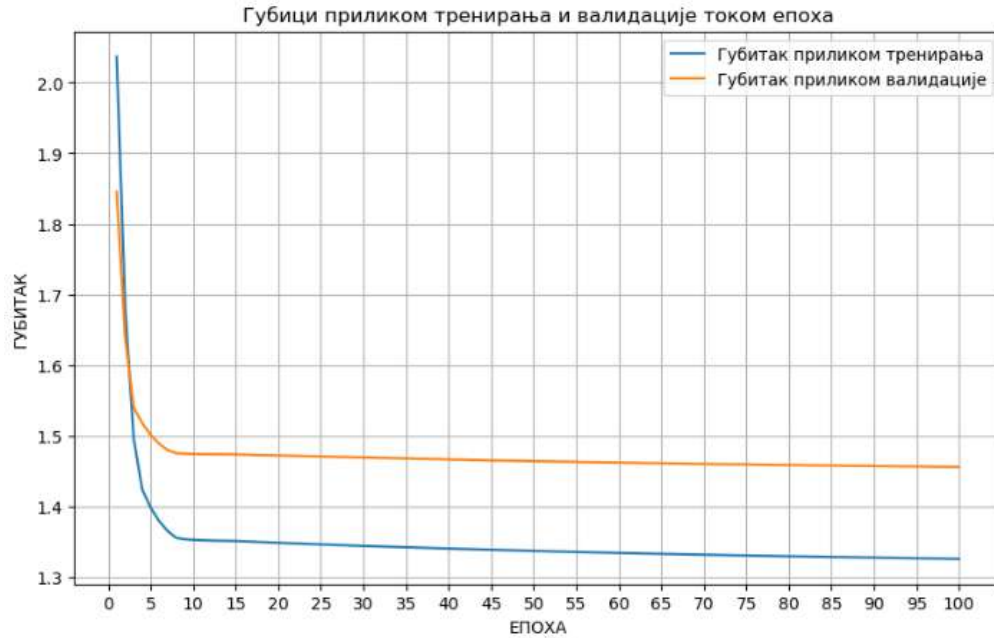
Већа стопа учења може убрзати конвергенцију, али може премашити минимум, док нижа стопа учења осигурава стабилнију конвергенцију, али по цену споријег тренирања. Одабир одговарајуће стопе учења је кључан за стабилност тренинга. Превисока стопа учења може довести до одступања у тренирању, док прениска стопа учења може довести до дугог времена конвергенције или заглављивања у локалним минимумима. Савремени алгоритми оптимизације, укључујући адам оптимизатор, могу да подесе брзину учења током тренирања, што помаже да се ублажи део осетљивости на почетну поставку брзине учења.

Стопа учења од 0.001 је стандардна за многе примене и подразумевана је за неколико алгоритама оптимизације, укључујући адам оптимизатор. Стопа учења од 0.001 успоставља равнотежу између брзине конвергенције и стабилности, што је чини сигурном полазном тачком. За тренирање LC класификације где ће модел можда морати да се креће кроз комплексне вредности губитака због разноврсне и високодимензионалне природе података даљинске детекције, ово представља разуман избор.

На Слици 48. приказани су губици приликом тренирања и валидације модела за LC класификацију током епоха.

Вредности губитака за тренирање, као и вредности губитака за валидацију се нагло смањују до 7. епохе. Код вредности губитака за тренирање приметне су осетније вредности смањења. Од 8. епохе вредности губитака за тренирање, као и вредности губитака за валидацију се смањују али многоструко мањим темпом током епоха, што указује да модел и даље учи и побољшава своја предвиђања током времена. Ово је позитиван знак да је процес тренирања ефикасан у смањењу грешке модела како на видљивим (скуп података за тренирање) тако и на подацима које види први пут (скуп података за валидацију).

Нижи темпо и постепено смањење и достигнут плато вредности губитака сугерише да се модел приближио свом капацитету да учи из датих података. У идеалном случају, пожељно је да се губитак остварен током тренирања и валидације смањи до тачке



Слика 48: Графички приказ губитака приликом тренирања и валидације модела за LC класификацију током епоха

стабилности са минималним размаком између то двоје.

Приметан је јаз између губитка тренинга и губитка валидације што може указивати на прекомерно прилагођавање, где модел ради добро на подацима за тренирање, али остварује лошије резултате на подацима које види први пут. Међутим, у овом случају, оба губитка се смањују током времена, што сугерише да се модел добро генерализује, иако постоји простор за побољшање у даљем минимизирању јаза.

7.5 Фаза 2 - SOC естиматор

Једном када је LC-класификатор истрениран, излаз из било ког од његових слојева може се употребити као скуп обележја за било који задатак трансфер учења који се односи на LC-класификацију. У овом истраживању одлучено је да се користи излаз последњег слоја мреже који има исте димензије као и излаз мреже, односно први деконволуциони слој мреже (уско грло), који се налази неких 16 слојева пре излаза модела. Његов излаз су латентна обележја високог нивоа за сваки пиксел, претпоставка је да су екстрахована обележја релевантна за процену SOC-а. Сваки пиксел је представљен вектором од 194 вредности. Стога за део слике имамо 794,624 (64 пиксела x 64 пиксела x 194) вредности.

На тај начин је конструисан U-Net само за енкодер, који, у тренутку инференција, пружа обележја за све пикселе у региону од интереса у једном пролазу.

Да бисмо креирали скупове података који се користе за тренирање и процену добијених SOC естиматора, Sentinel-2 снимци су преузети за географске локације за које је доступна лабораторијски одређена вредност SOC-а. Преузети снимци су снимљени што је могуће ближе датуму узорковања, уз услов да је облачност мања од 10%. Затим се

део слике од 64 x 64 пиксела издваја из ових снимака, центриран на свакој од локација са познатим вредностима SOC-а и уноси се у предложени екстрактор обележја, пружајући нам 194 вектора обележја (информације о географској ширини и дужини, као и 192 латентна обележја из LC-класификатора) за свако мерење SOC-а.

Потом је овај новодобијени скуп података коришћен за тренирање и процену бројних различитих модела машинског учења и приступа спроведених приликом експеримената извршених у оквиру овог истраживања.

Као саставни део овог истраживања за потребе генерисања индекса даљинске детекције и атрибута терена коришћени су Google Earth Engine и Google Earth Engine Code Editor. Добијени подаци су даље коришћени као додатна улазна обележја употребљена приликом различитих експеримената.

Google Earth Engine (GEE) је напредна технолошка платформа која користи рачунарство у облаку за обраду и анализу великих количина сателитских снимака и геопросторних података. Дизајниран је да прати промене животне средине током времена на глобалном нивоу, олакшавајући студије о климатским променама, крчењу шума, урбаној експанзији, природним катастрофама и коришћењу пољопривредног земљишта, поред бројних других примена. Основна снага Google Earth Engine-а лежи у његовом огромном каталогу података и његовој способности да брзо и ефикасно обавља сложене анализе на огромним географским подручјима и дугим временским периодима.

Каталог података укључује историјске колекције сателитских снимака које сежу уназад преко четрдесет година, као и скупове геопросторних података који покривају климу, време, топографију и још много тога. Ова свеобухватна архива омогућава корисницима да приступе петабајтима података без потребе за локалним складиштењем или процесорском снагом.

Google Earth Engine Code Editor је web базиран интерфејс који корисницима пружа алате за писање, отклањање грешака и покретање кодова који користе Earth Engine API. Code Editor је дизајниран да буде доступан корисницима са различитим нивоима стручности у програмирању, од почетника до напредних програмера.

Садржи уређивач кода где корисници пишу JavaScript код за интеракцију са Earth Engine API. Нуди функције као што су истицање синтаксе, довршавање кода и провера грешака за помоћ у развоју.

Кроз преглед мапа приказује геопросторне податке и резултате анализе. Корисници могу да визуелизују слојеве, да комуницирају са мапом да би прегледали различите регионе и да прилагођавају параметре визуелизације у реалном времену.

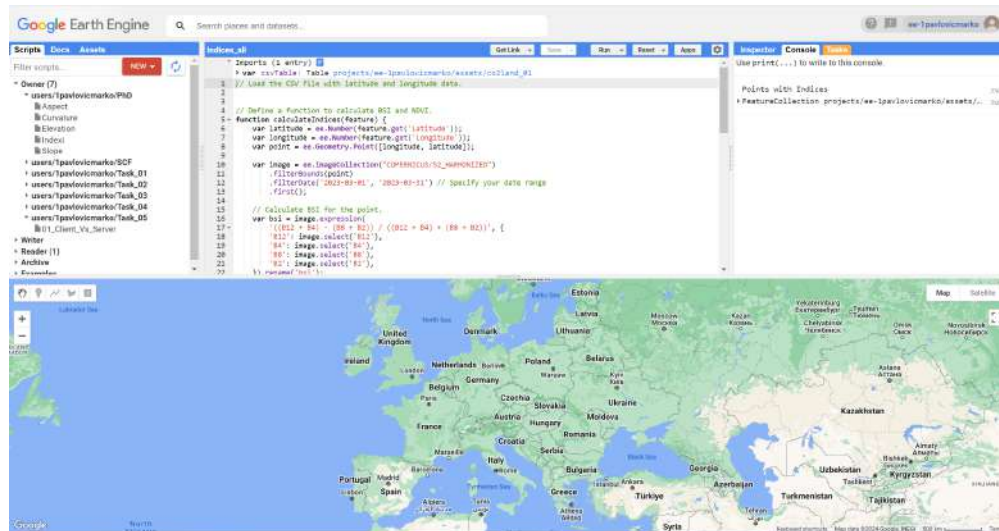
Преко функционалности конзола могу се видети излазни резултати прорачуна, изјаве за штампање и грешке, омогућавајући корисницима да отклоне грешке у својим кодовима и виде излазне текстуалне податке.

Функционалност инспектор омогућава корисницима да кликну на мапу и да прегледају вредности пиксела или карактеристика на одређеним локацијама.

Однос између Google Earth Engine и Google Earth Engine Code Editor је однос између платформе и алата. Earth Engine обезбеђује податке и рачунарску снагу, док је Code Editor кориснички интерфејс преко којег корисници комуницирају са Earth Engine-ом. У суштини, Code Editor је капија за приступ могућностима Earth Engine-а, омогућавајући корисницима да направе прилагођене кодове које користе функције обраде и анализе података Earth Engine-а. Корисници могу развити алгоритме за анализу трен-

дова у животној средини, креирање визуелизације и производњу модела података, све у контексту Code Editor-а, који се затим извршавају на платформи Earth Engine.

На Слици 49. видимо приказ корисничког интерфејса из Google Earth Engine Code Editor алата.



Слика 49: Google Earth Engine Code Editor кориснички интерфејс

За потребе овог истраживања развијен је модул на JavaScript програмском језику за употребу у Code Editor-у. Модул је дизајниран да обради податке о географским тачкама који се односе на мерења органског угљеника у земљишту. Обогађује ове тачке индексима даљинске детекције издвојеним из хармонизованих података Sentinel-2 сателита доступних преко GEE платформе. Израчунавају се индекси BSI, NDVI, NDWI, NDBI и SAVI.

Повезивањем тачака мерења SOC са скупом индекса даљинске детекције, овај модул значајно побољшава скуп улазних обележја за SOC моделе процене. Сваки индекс пружа јединствен увид у покривач земљишта, здравље вегетације, садржај воде и урбанизацију, што све може утицати на нивое SOC.

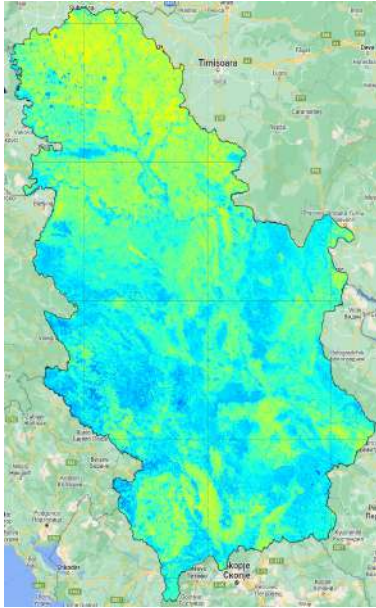
Избор слика из одређеног временског опсега осигурава да су екстраховани индекси временски релевантни и одражавају услове блиске временима мерења SOC, што је од виталног значаја за прецизно моделирање.

Употреба геометрије тачке за екстракцију индекса обезбеђује високу просторну прецизност, омогућавајући SOC процене које тачно одражавају локалне услове.

Комбинација индекса (BSI, NDVI, NDWI, NDBI, SAVI) нуди вишедимензионални приказ површине земљишта, омогућавајући нијансиранија предвиђања SOC-а која узимају у обзир различите еколошке и антропогене факторе.

На Слици 50(А). су приказане вредности BSI за територију Републике Србије, на Слици 50(Б). имамо приказ у виду композитног снимка, и на Слици 50(В). су приказане вредности NDVI за територију Републике Србије.

На Слици 50(А). плава и светло плава боја представљају области са ниским и умереним BSI вредностима. Жута и наранџаста наглашавају регионе са умереним до високим



(A) BSI



(Б) композитни снимак



(B) NDVI

Слика 50: Слика Републике Србије са вредностима BSI, RGB и NDVI

BSI вредностима, што указује на голу земљу или ретку вегетацију. Црвена је резервисана за подручја са највишим BSI вредностима, обично означавајући голу земљу или урбана подручја са минималном вегетацијом. Ова разноврсна палета визуелно разликује различите BSI вредности, помажући у тумачењу земљишног покривача.

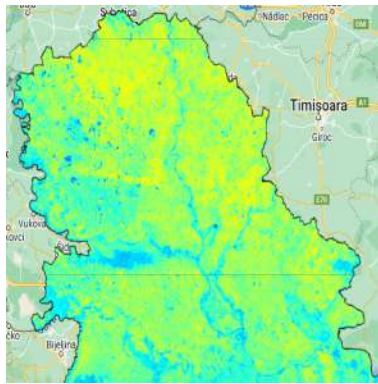
На Слици 50(Б). је територија Републике Србије представљена природним бојама пејзажа.

Док на Слици 50(В). тамно зелена означава подручја са највишим вредностима NDVI, што обично указује на здраву и густу вегетацију. Светлије нијансе зелене представљају умерене вредности NDVI, показујући мање густу или наглашену вегетацију. Жути и браон тонови се користе за истицање подручја без вегетације или са ретком вегетацијом, као што су пустиње или неплодно земљиште.

На Слици 51. су приказане вредности BSI, NDVI, као и композитни снимак територије АП Војводина.

Такође, развијен је додатни модул који има за циљ да обогати тачке мерења органског угљеника у земљишту са атрибутима терена који су изведени из података шатл радара топографске мисије (енгл. Shuttle Radar Topography Mission - SRTM). Овај процес укључује издвајање карактеристика терена као што су закривљеност, нагиб, надморска висина и аспект за сваку тачку.

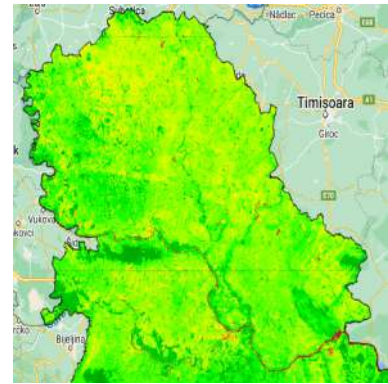
SRTM представља значајан напор у даљинској детекцији и посматрању Земље, лансиран у фебруару 2000. на свемирском шатлу Endeavour. Ова мисија је имала за циљ стварање најкомплетнијег дигиталног модела елевације високе резолуције Земљине копнене површине, који покрива преко 80% Земљине кугле од 60 степени северне до 56 степени јужне географске ширине. Користећи напредну радарску технологију, SRTM прикупља податке емитујући радарске сигнале према Земљи и мерећи време потребно



(A) BSI



(Б) композитни снимак



(B) NDVI

Слика 51: Слика АП Војводина са вредностима BSI, RGB и NDVI

да се врате, процес који је омогућио прецизно израчунавање топографских карактеристика планете. Добијени DEM из SRTM-а имају резолуцију до 30 метара, пружајући непроцењив ресурс за широк спектар примена укључујући геологију, хидрологију, планирање коришћења земљишта и студије животне средине.

За креирање маске за водена тела, како би се осигурало да се атрибути терена израчунавају само за копнене површине, искоришћен је скуп података промена глобалног шумског покривача (енгл. Global Forest Change) који су развили др Метју Хансен и његов тим са Универзитета у Мериленду. Наведени скуп података представља револуционарни напор у праћењу и анализи глобалног шумског покривача и промена током времена. Користећи податке сателитског програма Landsat, који пружа детаљне и свеобухватне слике Земљине површине, овај скуп података нуди неупоредив поглед на динамику шума од 2000. године надаље.

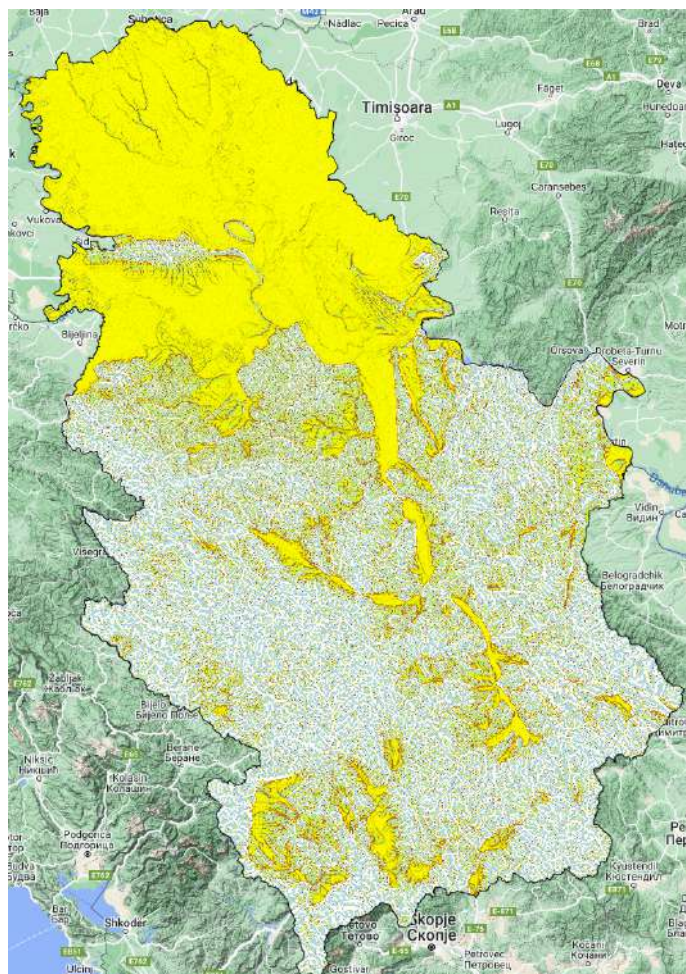
Да би се добили подаци о закривљености, нагибу, надморској висини и аспекту за сваку тачку употребљен је TAGEE (енгл. Terrain Analysis in Google Earth Engine - TAGEE) скуп функција и алата за анализу терена у оквиру GEE платформе које креирају и деле корисници за спровођење детаљне анализе атрибута терена.

На Слици 52. видимо приказ закривљености за територију Републике Србије.

Палета изабрана за визуелизацију закривљености представља градијент који прелази са конкавних на конвексне карактеристике терена. Почиње светлоплавом бојом, што указује на области негативне закривљености, типичне за конкавне површине као што су унутрашњост долина. Како се вредности закривљености приближавају нули, сугеришући равније или равне површине, палета прелази у светло зелену. Ово померање боје суптилно указује на промену од закривљеног према унутра ка више нивелисаном терену.

Жута боја означава области где је закривљеност приближно нула, означавајући равне или планарне регионе. Ова боја служи као граница између конкавних и конвексних терена. Како закривљеност постаје позитивна, што указује на конвексне површине где терен почиње да се извија према споља као спољашња површина брда или хумка, палета се помера у смеђу. Ова промена боје означава благи пораст терена.

Бела се користи за области које показују позитивне вредности закривљености, ти-



Слика 52: Приказ закривљености за територију Републике Србије

пичне за конвексне површине као што су врхови брда или гребена.

На Слици 53. видимо приказ нагиба за територију Републике Србије.

Палета која се користи за визуелизацију нагиба је дизајнирана да пренесе стрмину терена, у распону од благих нагиба до стрмих падина. Одабране боје напредују од светле до тамне, а свака представља различите степене стрмине падине.

Палета почиње светло цијан бојом, што указује на веома благе падине или скоро раван терен. Овај избор боја је идеалан за истицање подручја са минималним нагибом, где је нагиб готово неприметан. Како нагиб почиње да се повећава, боје прелазе кроз различите нијансе плаве. Светлоплава до нешто дубље плава се користе за ове релативно ниске, али све веће углове нагиба. Ове нијансе пружају визуелни знак за подручја у којима терен почиње да расте, али остаје лак за прелазак.

Како нагиб постаје све израженији, палета се помера у средње плаву. Ова боја означава области умерене стрмине, које би могле бити теже за навигацију или указују на значајне карактеристике пејзажа као што су обронци. За стрмије нагибе, користи се тамнија плава, наглашавајући области где је стрмина терена израженија и може утицати на факторе попут отицања воде или ерозије.



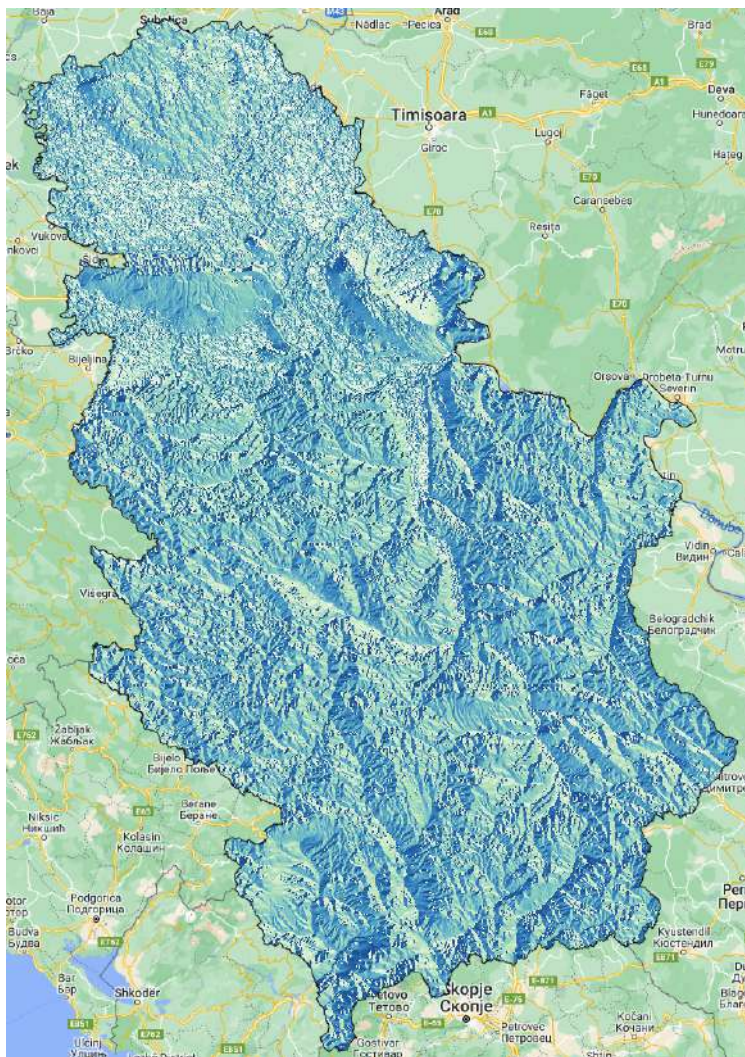
Слика 53: Приказ нагиба за територију Републике Србије

Коначно, веома тамно плава боја представља најстрмије падине на терену. Ове области, приказане у најмрачнијој нијанси, обично су најизазовнији делови пејзажа, које карактеришу оштри нагиби и потенцијално неравни терен.

На Слици 54. видимо приказ аспекта за територију Републике Србије.

Палета која се користи у приказу визуализације аспекта представља кардиналне правце (север, исток, југ, запад) и њихове међусмерове (североисток, југоисток, југозапад, северозапад). Помаже у визуелном тумачењу података везаних за аспекте, кроз брзо разумевање оријентације падина и њене потенцијалне импликације на различите факторе окружења.

Боја која се користи за падине окренуте ка северу је мека и хладна нијанса светло зелене. Символизује правац где су падине генерално засенчене и хладније, често фаворизујући вегетацију која успева у таквим условима. Крећући се у смеру казаљке на сату од севера, палета прелази у нешто топлију и светлију зелену боју да представља падине окренуте североистоку. Ове падине добијају мешавину сунчеве светлости и сенке током



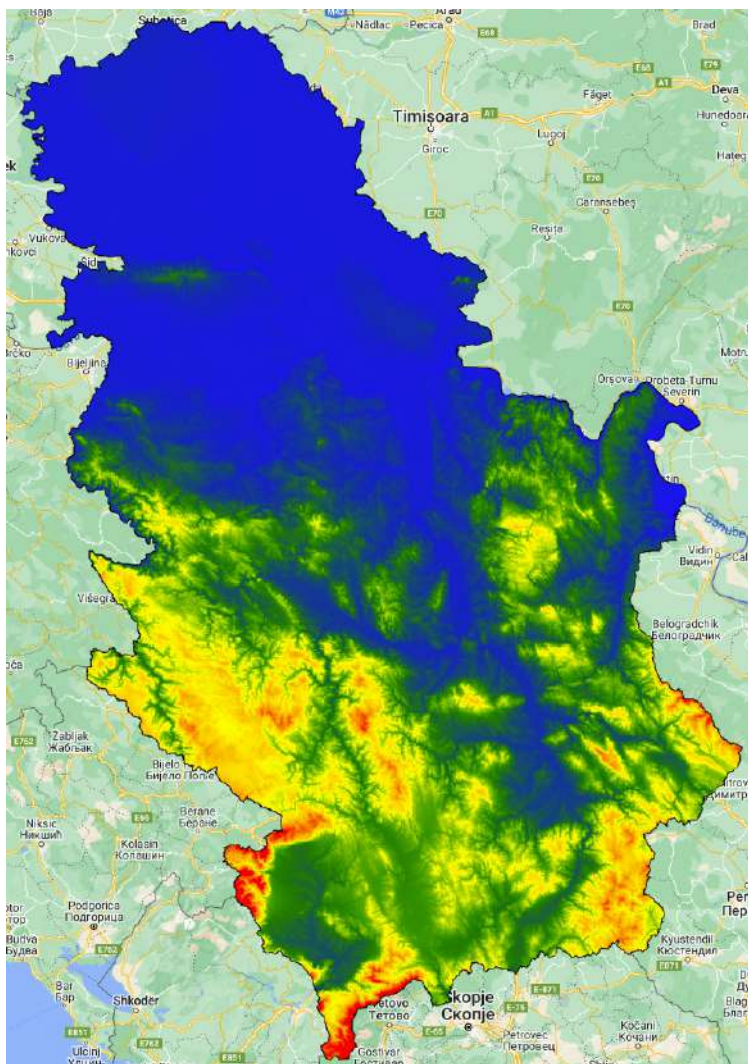
Слика 54: Приказ аспекта за територију Републике Србије

дана. Падине окренуте ка истоку илустроване су мирном цијан бојом . Ова боја означава падине које добијају јутарњу сунчеву светлост, чинећи их топлијим и погоднијим за одређене врсте вегетације.

Настављајући напредовање у смеру казаљке на сату, палета аспекта се помера у небеско плаву боју како би означила падине окренуте према југоистоку. Ове падине добијају довољно сунчеве светлости ујутру и често су топлије. Падине окренуте према југу су приказане подебљаном и богатом плавом бојом. Ова боја представља падине које су значајан део дана изложене директној сунчевој светлости, стварајући топлију микроклиму. Тамноплавом бојом се симболизују падине окренуте према југозападу. Ове падине добијају поподневну сунчеву светлост, пружајући јединствену микроклиму.

На Слици 55. видимо приказ надморске висине за територију Републике Србије.

Палета изабрана за визуелизацију надморске висине приказује варијације у нивоима надморске висине широм географске области. Плавом бојом су приказани делови са најнижим надморским висинама. Означава области које су близу нивоа мора или на



Слика 55: Приказ надморске висине за територију Републике Србије

нижим надморским висинама. Представља депресије, водена тела или низије. Зелена представља нешто више надморске висине од плаве. Одговара областима са умереним вредностима надморске висине, као што су равнице, брда или региони са благим падинама. Зелена означава средњи распон надморских висина.

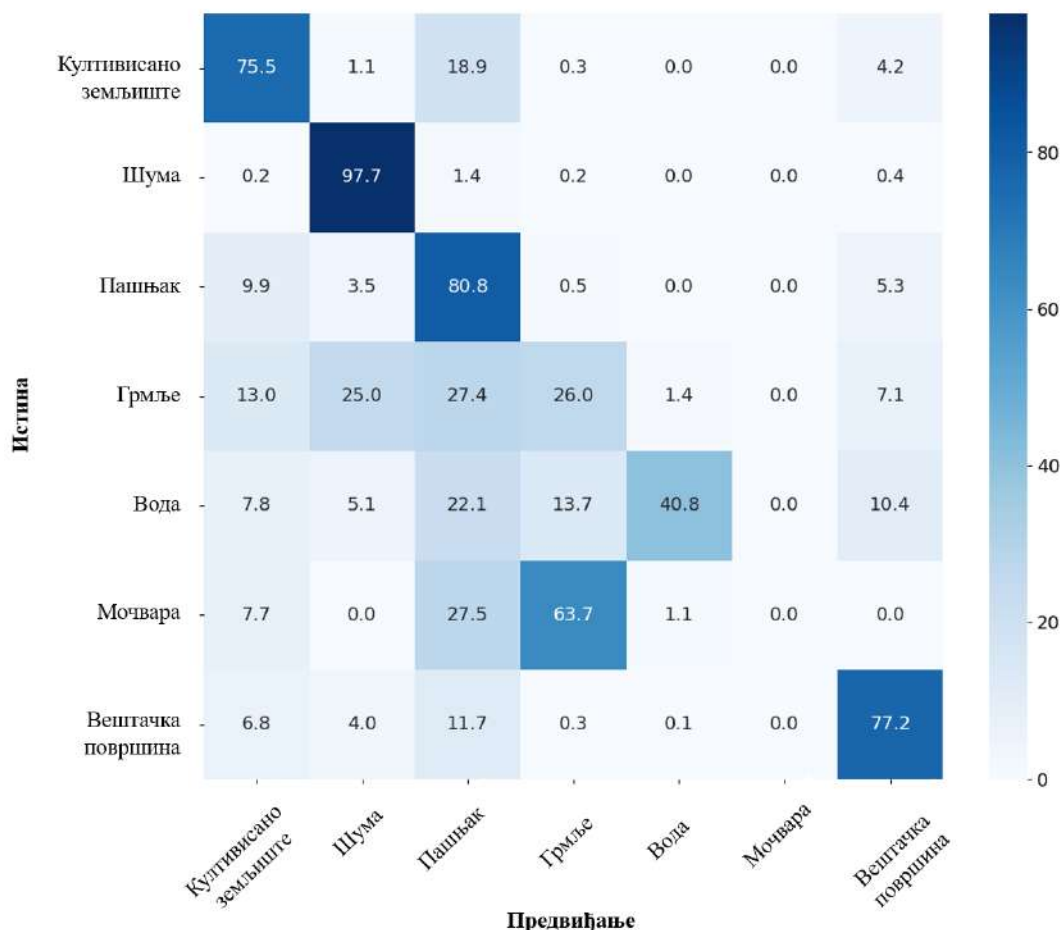
Како надморска висина наставља да расте, прелази у жуту боју. Жута означава веће надморске висине у поређењу са зеленом и често је повезана са подножјем планина или подручјима која су умерено повишена. Наранџаста представља још веће надморске висине. Користи се за приказивање делова са значајном надморском висином, као што су планине, висоравни или подручја са стрмијим падинама. Наранџаста означава значајно повећање надморске висине.

Црвена боја се обично додељује највишим надморским висинама, означавајући врхове, планинске ланце или области са највећим вредностима надморске висине. Црвена означава горњу границу елевације унутар визуелизованог подручја.

8 РЕЗУЛТАТИ ИСТРАЖИВАЊА И ДИСКУСИЈА

8.1 Евалуација LC класификације

Матрица конфузије за LC класификатор, заснован на U-Net архитектури, приказана је на Слици 56..



Слика 56: Евалуација LC класификације: матрица конфузије

Као што се види на матрици конфузије, модел је довољно прецизан за бројне LC класе. Перформансе су најбоље за шуме, које се детектују са 97,7% прецизности. Такође, за класе култивисаног земљишта и пашњака, перформансе модела су и даље добре (детектоване су са 75,5% и 80,8% тачности). Вреди напоменути да, чак и када је класификација за ове две класе погрешна, модел има тенденцију да меша ове две класе.

Једина друга класа коју модел помеша са две претходно наведене јесте вештачка површина, али то је вероватно услед карактеристика коришћеног скупа података за тренирање, који садржи део вештачких површина (углавном путева) чије су димензије мање од резолуције Sentinel- 2 и не може се поуздано идентификовати на слици.

Најизазовније за модел су класе грмље и мочваре. Ово се може приписати релативној оскудности података за тренирање који се односе на ове површине у скупу података.

Они представљају само 0,6% коришћеног скупа података. Пошто је модел успешан у одвајању класа култивисано земљиште и пашњак од остатка скупа података, закључујемо да ће издвојена латентна обележја бити довољно релевантна за крајњи циљ предложеног приступа.

С'обзиром на специфичности предложеног приступа примењена је процедура евалуације у два корака, односно евалуација за сваку фазу у приступу.

Прва фаза приступа је процењена на задатку LC класификације, како би било утврђено да су латентна обележја релевантна за задатак процене SOC. Након што је рад LC класификатора оцењен као задовољавајући, наставља се са екстракцијом латентних обележја за све релевантне узорке у LUCAS и CHLSOC скуповима података. Они су затим служили за процену перформанси целог приступа.

8.2 SOC процена Експеримент 1

Коришћењем вектора обележја изведених из U-Net енкодера, конструисани су скупови података који садрже латентна обележја и лабораторијски одређене вредности SOC-а за све узорке у CHLSOC и узорак у скуповима података LUCAS 2015. и 2018. релевантним за истраживање. За сваки унос у скупу података сачуване су географске координате.

Поред тога, процењен је утицај додавања теренских атрибута (закривљеност, нагиб, аспект, надморска висина), као и индекса даљинске детекције (NDVI, NDBI, NDWI, SAVI, BSI) како би се моделима пружиле додатне информације.

На основу њих процењено је 10 различитих алгоритама машинског учења доступних у библиотеци scikit-learn [170], са циљем да се идентификује алгоритам који показује оптималну равнотежу између тачности и генерализације.

Кандидати су укључивали: линеарну регресију, стабла квантилне регресије, логистичку регресију, регресију К-најближих суседа, регресију екстремно насумичних стабала, регресију машине вектора упоришта, регресију стабла одлучивања, регресију слушајне шуме, регресију повећавања градијента и вишеслојни перцептрон.

У оквиру овог истраживања низ модела машинског учења је естимиран за процену садржаја органског угљеника у земљишту. Ови модели су категорисани у две различите групе: базични модели и хибридни модели.

Базични модели су користили директних 13 спектралних опсега са сателитских снимака Sentinel-2, док су хибридни модели користили 192 латентна обележја екстрахована из U-Net мреже сегментације, која је тренирана за класификацију земљишног покривача коришћењем Sentinel-2 снимака.

Укупно четири базична модела и четири хибридна модела су развијена и истренирана. Базични модел 01 се ослањао искључиво на Sentinel-2 снимке за процену SOC-а. Сходно томе, хибридни модел 01 је трениран искључиво на 192 латентна обележја изведена из U-Net сегментационе мреже. Укупно је коришћено 38.559 тачака за евалуацију модела из серије 01.

Базични модел 02 је укључио Sentinel-2 снимке заједно са атрибутима терена, укључујући закривљеност, нагиб, аспект и надморску висину. Слично, хибридни модел 02 је трениран коришћењем латентних обележја U-Net заједно са наведеним атрибутима терена. За евалуацију модела из серије 02 употребљено је 35.425 тачака. До смањења у

односу на број тачака коришћених у серији 01 дошло је услед немогућности аквизиције података о атрибутима терена за одређени број тачака.

Основни модел 03 проширио је скуп функција укључивањем индекса даљинске детекције (NDVI, NDBI, NDWI, SAVI, BSI) који су комбиновани са Sentinel-2 снимцима, као и хибридни модел 03, уз разлику да су индекси комбиновани са 192 латентна обележја из U-Net мреже. За евалуацију модела из серије 03 употребљено је 38.248 тачака. Овде је такође дошло до смањења броја тачака у односу на серију 01, у мањој мери, услед немогућности аквизиције података о индексима даљинске детекције за одређени број тачака.

Најсвеобухватнији модели, базични модел 04 и хибридни модел 04, комбинују Sentinel-2 снимке са атрибутима терена и индексима даљинске детекције за базични модел, и одговарајућа латентна обележја заједно са истим додатним подацима за хибридни модел. За евалуацију модела из серије 04 употребљено је 35.126 тачака. Поново је присутан случај да је дошло до смањења броја тачака у односу на серију 01 услед немогућности аквизиције података о атрибутима терена и индексима даљинске детекције за одређени број тачака.

Да би се проценио учинак наведених модела, коришћен је 10-струки приступ унакрсној валидацији, са тачкама података измешаним из скупова података CHLSOC, LUCAS 2015 и LUCAS 2018. Процена перформанси и поређење различитих модела и алгоритама су спроведени коришћењем регресионих метрика: средња апсолутна грешка, средња квадратна грешка и коефицијент детерминације.

Поред тога, предузети су неопходни кораци за претходну обраду података, као што је нормализација Sentinel-2 снимака за базичне моделе, где су вредности пиксела из сваког опсега извучене и конвертоване у обележја.

Тачке података којима недостају информације за индексе даљинског детекције и атрибуте терена уклоњене су из свих скупова података, укључујући CHLSOC, LUCAS 2015 и LUCAS 2018, како би се осигурао интегритет и тачност процеса тренирања и евалуације модела.

Резултати експеримената за базичне моделе су приказани у табелама 3. и 4..

| Модел | Базични 01 | | | Базични 02 | | |
|----------------------------|------------|---------|---------|------------|---------|---------|
| | MAE | MSE | R^2 | MAE | MSE | R^2 |
| Линеарна регресија | 3.9242 | 50.3191 | 0.0234 | 2.9287 | 29.5733 | 0.0559 |
| Стабла квантилне регресије | 2.457 | 41.5081 | 0.1941 | 1.7013 | 20.586 | 0.3417 |
| Логистичка регресија | 3.9437 | 51.2172 | 0.0062 | 2.9394 | 29.8351 | 0.0476 |
| К најближих суседа | 2.8754 | 39.3465 | 0.2349 | 2.582 | 26.7411 | 0.1454 |
| Екстремно насумична стабла | 2.9775 | 37.7144 | 0.2664 | 1.9765 | 18.281 | 0.4139 |
| Машине вектора упоришта | 2.9568 | 50.8714 | 0.0134 | 2.5047 | 31.5548 | -0.0075 |
| Стабла одлучивања | 3.5404 | 70.0804 | -0.3675 | 2.4524 | 34.6654 | -0.1161 |
| Насумичне шуме | 2.869 | 36.457 | 0.2911 | 1.9703 | 18.2757 | 0.4137 |
| Повећавање градијента | 3.1045 | 38.5371 | 0.2519 | 2.2 | 20.9716 | 0.3293 |
| Вишеслојни перцептрон | 3.3644 | 42.8205 | 0.169 | 2.6562 | 25.3665 | 0.1902 |

Табела 3: Резултати 10-струке унакрсне валидације за базичне моделе трениране на сва 3 скупа података за базичне моделе 01 и 02

| Модел | Базични 03 | | | Базични 04 | | |
|----------------------------|------------|---------|---------|------------|---------|---------|
| | MAE | MSE | R^2 | MAE | MSE | R^2 |
| Линеарна регресија | 3.7896 | 47.8096 | 0.0741 | 2.8328 | 28.3646 | 0.0961 |
| Стабла квантилне регресије | 2.3898 | 40.4341 | 0.218 | 1.6706 | 20.3539 | 0.3529 |
| Логистичка регресија | 3.952 | 51.3172 | 0.0064 | 2.9457 | 29.8987 | 0.047 |
| К најближих суседа | 2.8639 | 39.3111 | 0.2384 | 2.5857 | 26.8376 | 0.1438 |
| Екстремно насумична стабла | 2.8945 | 35.4428 | 0.3138 | 1.91 | 17.2718 | 0.4501 |
| Машине вектора упоришта | 2.9456 | 50.8586 | 0.0157 | 2.5095 | 31.6323 | -0.0079 |
| Стабла одлучивања | 3.3712 | 64.7625 | -0.2573 | 2.3902 | 33.0685 | -0.0554 |
| Насумичне шуме | 2.8049 | 34.6424 | 0.3294 | 1.9466 | 17.9979 | 0.4271 |
| Повећавање градијента | 3.1005 | 38.5829 | 0.253 | 2.1927 | 20.844 | 0.3359 |
| Вишеслојни перцептрон | 3.3934 | 42.3942 | 0.1792 | 2.9056 | 27.0196 | 0.1329 |

Табела 4: Резултати 10-струке унакрсне валидације за базичне моделе трениране на сва 3 скупа података за базичне моделе 03 и 04

Као што показују подаци у табелама, за базичне моделе, метрике перформанси показују да су модели који су укључивали атрибуте терена и индексе даљинске детекције (базични 02 и базични 04) генерално показали побољшане перформансе у поређењу са онима који користе само снимке Sentinel-2 или индексе даљинске детекције.

Конкретно, базични модел 04, који комбинује атрибуте терена и индексе даљинске детекције, остварио је најбоље просечне перформансе. Ово сугерише да је свеобухватан приступ, који укључује низ обележја, најефикаснији за процену SOC-а у конфигурацијама базичног модела.

Три алгоритма са најбољим учинком за базичне моделе били су насумичне шуме, екстремно насумична стабла и стабла квантилне регресије. Сва три алгоритма представљају ансамбл методе стабла одлучивања.

Наведени алгоритми су показали супериорне перформансе у смислу ниже средње апсолутне грешке, ниже средње квадратне грешке и већег коефицијента детерминације, што указује на њихову ефикасност у задацима процене SOC.

Резултати експеримената за хибридне моделе представљени су у табелама 5. и 6..

| Модел | Хибридни 01 | | | Хибридни 02 | | |
|----------------------------|-------------|---------|--------|-------------|---------|--------|
| | MAE | MSE | R^2 | MAE | MSE | R^2 |
| Линеарна регресија | 3.4754 | 42.3284 | 0.1782 | 2.6789 | 25.7789 | 0.177 |
| Стабла квантилне регресије | 2.0375 | 30.2786 | 0.4132 | 1.5064 | 15.5843 | 0.5021 |
| Логистичка регресија | 3.4722 | 43.1366 | 0.1628 | 2.6904 | 26.2908 | 0.1604 |
| К најближих суседа | 3.3137 | 43.9859 | 0.1454 | 2.5396 | 26.4444 | 0.1526 |
| Екстремно насумична стабла | 2.2982 | 27.8525 | 0.4593 | 1.6738 | 14.947 | 0.5219 |
| Машине вектора упоришта | 2.9343 | 49.0894 | 0.0478 | 2.3422 | 28.6765 | 0.0848 |
| Стабла одлучивања | 2.6978 | 47.2165 | 0.0832 | 2.0351 | 26.3656 | 0.1561 |
| Насумичне шуме | 2.3998 | 27.3409 | 0.4695 | 1.7217 | 14.6514 | 0.5306 |
| Повећавање градијента | 2.9492 | 35.8304 | 0.3045 | 2.1369 | 19.7664 | 0.3682 |
| Вишеслојни перцептрон | 3.5001 | 47.5478 | 0.0806 | 2.6696 | 24.9651 | 0.1964 |

Табела 5: Резултати 10-струке унакрсне валидације за хибридне моделе трениране на сва 3 скупа података за хибридне моделе 01 и 02

| Модел | Хибридни 03 | | | Хибридни 04 | | |
|----------------------------|-------------|---------|--------|-------------|---------|--------|
| | MAE | MSE | R^2 | MAE | MSE | R^2 |
| Линеарна регресија | 3.4615 | 41.8802 | 0.1888 | 2.664 | 25.4305 | 0.1892 |
| Стабла квантилне регресије | 2.0248 | 30.0811 | 0.4184 | 1.5047 | 15.764 | 0.4992 |
| Логистичка регресија | 3.4765 | 43.1752 | 0.164 | 2.6944 | 26.3329 | 0.1605 |
| К најближих суседа | 3.317 | 44.0559 | 0.1458 | 2.541 | 26.339 | 0.1586 |
| Екстремно насумична стабла | 2.2932 | 27.5334 | 0.4674 | 1.6727 | 14.9439 | 0.5248 |
| Машине вектора упоришта | 2.9382 | 49.1632 | 0.0485 | 2.3451 | 28.7204 | 0.0849 |
| Стабла одлучивања | 2.6926 | 46.8923 | 0.0908 | 1.9933 | 25.1316 | 0.1985 |
| Насумичне шуме | 2.3924 | 27.1312 | 0.4751 | 1.7145 | 14.6738 | 0.5329 |
| Повећавање градијента | 2.9498 | 36.0232 | 0.3025 | 2.1341 | 19.8038 | 0.369 |
| Вишеслојни перцептрон | 3.5218 | 46.0292 | 0.1073 | 2.646 | 25.1262 | 0.1992 |

Табела 6: Резултати 10-струке унакрсне валидације за хибридне моделе трениране на сва 3 скупа података за хибридне моделе 03 и 04

У контексту хибридних модела, резултати су били константно бољи од основних модела. Ово побољшање се приписује укључивању латентних обележја изведених из дубоке неуронске мреже, тј. U-Net мреже сегментације.

Тренд побољшања перформанси са додатком атрибута терена и индекса даљинске детекције такође је примећен у хибридним моделима. Хибридни модел 04, који је укључивао обе врсте додатних функција, остварио је најбоље перформансе као и код базичних модела.

Такође, као и код базичних модела, алгоритми који су остварили најбољи учинак били су насумичне шуме, екстремно насумична стабла и стабла квантилне регресије.

Генерално, хибридни модели, који користе латентна обележја са више информација, надмашили су базичне моделе. Овај образац је очигледан у побољшаним метрикама перформанси (нижи MAE и MSE, већи R^2) у свим алгоритмима.

Алгоритми који су остварили најбољи учинак, посебно насумичне шуме и екстремно насумична стабла, показали су робусност у различитим конфигурацијама модела. Поред тога, додавање свеобухватних скупова обележја, тј. атрибута терена и индекса даљинске детекције, значајно побољшава перформансе модела, што је тренд примећен и у базичним и у хибридним категоријама модела.

8.3 Додатна евалуација модела са најбољим учинком за процену SOC - Експеримент 2

Пошто је претходно урађена 10-струка унакрсна валидација показала да су базични модел 04 и хибридни модел 04, остварили најбољи учинак, следећи корак је да се процени да ли се наведени модели могу генерализовати за употребу широм света (тј. CHLSOC и LUCAS скупови података).

Како би то било урађено, употребљен је посебан тип 10-струке унакрсне валидације. Сваки преклоп (енгл. fold) у 10-струкој унакрсној валидацији креиран је коришћењем 100% података из два скупа, као и 90% података из трећег скупа. Преосталих 10% података из трећег скупа употребљено је за евалуацију модела. У Експерименту 2 укупно је коришћено 35.126 тачке за евалуацију модела.

За евалуацију су примењена 3 типа алгоритама која су остварила најбоље резултате приликом класичне 10-струке унакрсне валидације. То су насумичне шуме, екстремно насумична стабла и стабла квантилне регресије.

У табелама 7. и 8. су представљени резултати за базични модел 04, као и за хибридни модел 04 са употребом различитих алгоритама.

| Модел | Базични 04 | | | | |
|----------------------------|---------------|--------|---------|--------|----------|
| Алгоритам | Скуп података | MAE | MSE | R^2 | MAPE |
| Екстремно насумична стабла | CHLSOC | 1.4561 | 4.9342 | 0.7604 | 55.2074 |
| | LUCAS 2015 | 1.7813 | 15.3941 | 0.4277 | 85.9890 |
| | LUCAS 2018 | 2.1245 | 22.5000 | 0.3916 | 104.9643 |
| Насумичне шуме | CHLSOC | 1.4164 | 4.6487 | 0.7744 | 52.4240 |
| | LUCAS 2015 | 1.8542 | 15.7807 | 0.4123 | 90.1151 |
| | LUCAS 2018 | 2.1611 | 23.5234 | 0.3604 | 101.4239 |
| Стабла квантилне регресије | CHLSOC | 1.3680 | 4.8233 | 0.7666 | 43.4362 |
| | LUCAS 2015 | 1.5298 | 18.5261 | 0.3108 | 55.9282 |
| | LUCAS 2018 | 1.8446 | 26.7573 | 0.2790 | 65.9441 |

Табела 7: Резултати евалуације базичних 04 модела са алгоритмима са најбољим учинком за процену SOC

| Модел | Хибридни 04 | | | | |
|----------------------------|---------------|--------|---------|--------|---------|
| | Скуп података | MAE | MSE | R^2 | MAPE |
| Екстремно насумична стабла | CHLSOC | 1.5991 | 5.6120 | 0.7286 | 57.9446 |
| | LUCAS 2015 | 1.4069 | 12.5886 | 0.5044 | 60.6769 |
| | LUCAS 2018 | 1.7773 | 19.0146 | 0.4884 | 77.5490 |
| Насумичне шуме | CHLSOC | 1.5221 | 5.2979 | 0.7424 | 59.2572 |
| | LUCAS 2015 | 1.5053 | 11.3710 | 0.5669 | 71.6700 |
| | LUCAS 2018 | 1.8583 | 19.4209 | 0.4677 | 86.7991 |
| Стабла квантилне регресије | CHLSOC | 1.4286 | 5.0055 | 0.7577 | 44.0793 |
| | LUCAS 2015 | 1.3090 | 12.8420 | 0.5196 | 50.8010 |
| | LUCAS 2018 | 1.6057 | 20.6066 | 0.4419 | 56.4293 |

Табела 8: Резултати евалуације хибридних 04 модела са алгоритмима са најбољим учинком за процену SOC

Естимација базичног модела 04 и хибридног модела 04 за процену органског угљеника у земљишту, коришћењем алгоритама насумичне шуме, екстремно насумична стабла и стабла квантилне регресије, открива сложене обрасце перформанси у различитим скуповима података (CHLSOC, LUCAS 2015, и LUCAS 2018).

Наведена анализа је проширена са додатном метриком, конкретно са средњом апсолутном процентуалном грешком (MAPE), како би се добио дубљи увид у тачност и поузданост предвиђања модела и њихову способност генерализације.

У контексту процене SOC-а, укључивање MAPE-а је посебно корисно јер омогућава стандардизовано поређење грешака у скуповима података са различитим скалама и величинама SOC вредности, што олакшава тумачење и упоређивање перформанси модела.

У базичном моделу 04, алгоритам екстремно насумичних стабала показује похвалну тачност на скупу података CHLSOC, са MAE од 1,4561 и R^2 од 0,7604, али његове перформансе се смањују на LUCAS скуповима података, посебно у LUCAS 2018 скупу где се MAE повећава на 2,1245 а R^2 пада на 0,3916. Ова варијација сугерише на потенцијално прекомерно прилагођавање специфичним обележјима података присутним у скупу података CHLSOC.

Алгоритам насумичне шуме, иако приказује нешто боље перформансе у односу на екстремно насумична стабла на скупу података CHLSOC, прати сличан тренд смањене ефикасности на LUCAS скуповима података.

Алгоритам стабла квантилне регресије, међутим, показује конзистентније перформансе у свим скуповима података, што указује на његову робусност и прилагодљивост различитим обележјима података.

Хибридни модел 04 показује занимљив образац где алгоритам екстремно насумичних стабала, иако показује добру тачност на CHLSOC, значајно побољшава своје перформансе на скупу података LUCAS 2015, у односу на базни модел 04, о чему сведоче нижи MAE и већи R^2 . Ово побољшање указује на способност алгорита да ефикасно искористи додатна обележја присутна у хибридном моделу. Међутим, слично као код базног модела 04, приметан је пад перформанси на скупу података LUCAS 2018.

Алгоритам насумичне шуме у хибридном моделу, за разлику од његових перформанси у базичном моделу, не одржава исти ниво тачности, посебно на LUCAS скуповима података.

Алгоритам стабла квантилне регресије се поново истиче по својој доследној тачности у свим скуповима података у хибридном моделу, појачавајући његову погодност за различите контексте процене SOC-а.

Спроведена анализа подвлачи супериорност алгоритма стабла квантилне регресије у руковању различитим скуповима података, што га чини веома погодним за генерализовање и поузданим избором за процену SOC-а у конфигурацијама базичног и хибридног модела.

Укључивање MAPE као метрике обогаћује евалуацију, нудећи јаснију перспективу релативне тачности алгоритама. Различите перформансе алгоритама екстремно насумичних стабала и насумичних шума у различитим скуповима података наглашавају важност одабира правог алгоритма који не само да добро ради на одређеном скупу података, већ и одржава своју ефикасност у различитим изворима података, што је кључно разматрање у глобалној примени модела процене SOC-а.

Доказана супериорност хибридних модела над базичним моделима у контексту процене органског угљеника у земљишту може се разумети кроз свеобухватну анализу њихових основних методологија и резултата перформанси посматраних из специјализованог процеса 10-струке унакрсне валидације.

Хибридни модели се разликују првенствено по коришћењу напредних скупова обележја добијених применом метода дубоког учења. Ова латентна обележја, извучена из U-Net сегментационе мреже, представљају софистицираније и нијансираније разумевања класификације земљишног покривача на основу Sentinel-2 снимака.

Овај напредни инжењеринг добијања обележја података омогућава хибридним моделима да ухвате сложене просторне и спектралне односе који су кључни за прецизну процену SOC-а. Латентна обележја обухватају дубљи ниво информација, надилазећи директне спектралне опсеге који се користе у базичним моделима, и стога пружају робуснији темељ за рад алгоритмима машинског учења.

Насупрот томе, базични модели се директно ослањају на 13 спектралних опсега са сателитских снимака Sentinel-2. Иако ови опсежи нуде драгоцене информације, они можда неће обухватити пуно сложеност односа између типова земљишног покривача и садржаја SOC.

Недостатак напредног инжењеринга добијања обележја података у основним моделима могао би довести до површније анализе сателитских снимака, потенцијално пропуштајући суптилне, али критичне обрасце битне за тачну процену SOC-а.

Метрике коришћене приликом евалуација додатно појачавају супериорност хибридних модела. У различитим алгоритмима, хибридни модели су доследно показивали побољшану тачност и могућност генерализације.

Ово је очигледно из нижих средњих апсолутних грешака, средњих квадратних грешака и виших коефицијената детерминације у различитим скуповима података у поређењу са базичним моделима.

Штавише, додавање средње апсолутне процентуалне грешке као метрике наглашава релативну прецизност предвиђања хибридних модела на интуитивнији начин, омогућавајући стандардизовано поређење између скупова података са различитим SOC скала-

ма.

Побољшане перформансе хибридних модела се такође могу приписати њиховој способности да ефективно интегришу додатне изворе података као што су атрибути терена и индекси даљинске детекције.

Ова интеграција омогућава моделима да размотре шири спектар фактора околине који утичу на SOC, што доводи до прецизнијих предвиђања. Алгоритми у хибридних моделима, који имају користи од богатог, латентног скупа обележја, боље су опремљени да открију сложене, нелинеарне односе који су често кључни за разумевање и предвиђање SOC садржаја.

У закључку, напредни инжењеринг добијања обележја података у хибридних моделима, заједно са њиховом способношћу да инкорпорирају шири спектар фактора околине, резултира софистициранијим, прецизнијим и генерализативнијим приступом за процену SOC-а. Ово чини хибридне моделе супериорнијим избором у поређењу са базичним моделима, посебно у применама које захтевају високу тачност и поузданост у различитим географским скуповима података.

Додавање атрибута терена и индекса даљинске детекције, како базичним тако и хибридних моделима за процену SOC-а показало је квантитативно побољшање у њиховим перформансама. О томе сведоче наведене вредности прецизираних метрика. Ово побољшање је укорењено у додатним информацијама о самом окружењу тачака узорковања земљишта, као и у контекстуалним информацијама које ова обележја пружају, допуњујући спектралне податке са Sentinel-2 снимака.

На пример, у базичним моделима који су се у почетку ослањали искључиво на снимке Sentinel-2 сателита, укључивање атрибута терена као што су закривљеност, нагиб, аспект и надморска висина довело је до приметног смањења средње апсолутне грешке и средње квадратне грешке, и повећања коефицијента детерминације.

Конкретно, у типичном базном моделу, додавање атрибута терена може смањити MAE до 20-30% и MSE за сличну или већу маргину, док повећава вредност R^2 за 10-15 процентних поена. Ово значајно побољшање тачности се приписује додатним топографским информацијама, које играју кључну улогу у разумевању услова животне средине који утичу на SOC.

Индекси даљинске детекције као што су NDVI, NDBI, NDWI, SAVI и BSI додатно повећавају могућности модела. Примећено је да њихова интеграција додатно смањује MAE и MSE и повећава вредности R^2 .

У практичном смислу, ово се огледа у додатних 10-20% смањења MAE и MSE приликом укључивања наведених индекса у обележја за тренирање модела, заједно са даљим повећањем R^2 од 5-10 процентних поена, у поређењу са моделима који су тренирани само на атрибутима терена. Ово побољшање наглашава вредност индекса даљинске детекције у пружању увида у покривач земљишта и обрасце коришћења земљишта који су кључни за процену SOC.

Код хибридних модела, утицај додавања ових обележја је још израженији. Напредна латентна обележја из U-Net сегментационе мреже, у комбинацији са подацима о терену и индексима даљинске детекције, доводе до свеобухватног модела који хвата сложене, нелинеарне односе од виталног значаја за прецизно предвиђање SOC-а.

Метрике перформанси хибридних модела са овим додатним обележјима обично показују додатно смањење од 15-25% у MAE и MSE, и повећање од 10-20 процентних

поена у R^2 , у поређењу са њиховим парњацама из групе базичних модела.

На крају, инкорпорација атрибута терена и индекса даљинске детекције има значајан нумерички утицај на перформансе модела. Смањења метрика грешке и повећање вредности R^2 не само да демонстрирају повећану тачност и поузданост модела, већ и наглашавају важност вишестраног приступа који узима у обзир различите факторе околне за ефикасну процену SOC-а. Овај приступ води ка моделима који нису само тачни већ и робусни и способни да се генерализују у различитим географским контекстима.

8.4 Процена садржаја SOC у подручјима са хомогеним типовима вегетацијског покривача - Експеримент 3

У овом делу истраживања фокус је стављен на експерименте који испитују перформансе најбољег модела, тј. хибридног модела 04, у процени садржаја органског угљеника у земљишту у областима са хомогеним насупрот хетерогеним типовима вегетационог покривача.

Циљ је да се даље испитају могућности и ограничења модела. Приликом анализе коришћен је LUCAS 2018 скуп података, који је посебно релевантан због инхерентне разноликости у типовима вегетације и броја прикупљених узорака земљишта.

Хетерогеност скупа података LUCAS 2018 у типовима вегетације представља јединствену прилику да се процени ефикасност модела у различитим условима животне средине. У областима са хомогеним вегетацијским покривачем, претпоставка је да би модел могао лакше предвидети садржај SOC због релативне униформности фактора животне средине који утичу на земљиште. Спектрални и просторни обрасци повезани са једном врстом вегетационог покривача ће вероватно бити конзистентнији, омогућавајући моделу да направи тачнија предвиђања на основу научених образаца из података за тренирање.

Одлука да се за овај експеримент фокусира на скуп података LUCAS 2018 је донесена узимајући у обзир његову разноврсну покривеност вегетацијом и обимно прикупљање узорака тла. CHLSOC није коришћен јер није толико хетероген у поређењу са LUCAS 2018 и не садржи много прикупљених узорака земљишта, који се поклапају са доступношћу снимака са Sentinel-2 сателита (лансираних у јуну 2015. и марту 2017. године). LUCAS 2015, такође није коришћен услед доступности Sentinel-2 снимака, како не би дошло до значајног губитка информација за тачке узорака земљишта због недостајућих сателитских снимака, као и података за индексе даљинског истраживања и атрибута терена.

Даље, LUCAS 2015 и LUCAS 2018 скупови података имају велико преклапање у виду локација са којих су прикупљени узорци земљишта, што би довело до редувантних података у скупу за тренинг што би негативно утицало на перформансе модела.

У овом истраживању направљене су 3 групе хомогених типова вегетацијског покривача. Групе су категорисане на основу документа LUCAS 2018 Технички референтни документ С3: Класификација (земљишни покривач и коришћење земљишта). Наведени документ пружа детаљне информације о систему класификације земљишног покривача и употреби земљишта који се користи у LUCAS истраживању.

Три групе које су издвојене за ову сврху су:

- Група 1 - В10 житарице (под В00 Пољопривредно земљиште). Ова група представља површине претежно под узгојем житарица, пружајући хомогену вегетацију која се првенствено састоји од житарица.
- Група 2 - С10 листопадне шуме (под С00 шуме). Ова група се фокусира на области прекривене широколисним шумама, нудећи посебан тип вегетационог покривача који карактеришу листопадне врсте дрвећа.
- Група 3 - Е20 травњаци без дрвећа/жбуња (под Е00 травњаци). Ова група обухвата травнате површине без делова покривених дрвећем или жбуњем, представљајући уједначен тип вегетације који се првенствено састоји од трава.

Евалуација перформанси процене SOC хибридног модела 04 у овим специфичним, хомогеним типовима вегетације омогућава разумевање ефикасности модела у различитим еколошким окружењима.

Такође, испитује осетљивост модела на варијације у вегетационом покривачу. Хомогеност унутар сваке групе помаже да се изолује утицај специфичних типова вегетационог покривача на садржај SOC, пружајући јаснији увид у однос између типа вегетације и нивоа органског угљеника у земљишту.

Исходи ових експеримената су од значаја за дубље испитивање и одређивање применљивости и тачности хибридног модела 04 у процени SOC-а за различите типове земљишног покривача. Даље, испитивање помаже у бољем разумевању начина на који модел реагује на јединствене карактеристике различитих вегетацијских покривача.

Овде је примењена модификована верзија 10-струке унакрсне валидације где је вршено селективно уклањање одређене групе, као што су нпр. житарице В10 из LUCAS 2018 скупа. Остатак скупа је употребљен као примарни скуп за тренирање модела. Као секундарни скуп за тренирање модела употребљено је 90% уклоњене групе (нпр. житарице В10). Затим је извршено тестирање перформанси модела на преосталих 10% ове групе.

За групу 1 - житарице В10 употребљено је 4.145 тачака за тренирање, за групу 2 - листопадне шуме С10 употребљено је 2.002 тачке за тренирање и за групу 3 - травњаци без дрвећа/жбуња коришћено је 2.758 тачака.

То је омогућило да се добије циљана процена тачности модела у процени SOC-а за специфични тип вегетације. Приликом експеримената и овде су коришћени алгоритми са најбољим учинком - насумичне шуме, екстремно насумична стабла и стабла квантилне регресије.

Као базни модел према којем се процењивао учинак модела у одређеним групама вегетације коришћен је модел трениран на читавом LUCAS 2018 скупу, тј. на 16.232 тачака. Учинак референтног модела је добијен применом редовне 10-струке унакрсне валидације. На овај начин успостављене су основне перформансе за поређење са учинком хибридног модела 04. Поређењем перформанси хибридног модела 04 у различитим групама вегетације из скупа података LUCAS 2018 са његовим перформансама на читавом скупу података пружаће свеобухватан поглед на свестраност и тачност модела у различитим контекстима.

У свакој селектованој групи вегетације (В10 житарице, С10 листопадне шуме и Е20 травњаци без дрвећа/жбуња), перформансе модела се мере тренирањем на 90% пода-

така из групе и тестирањем на преосталих 10%. Овај приступ пружа увид у то колико се модел прилагођава и предвиђа садржај SOC у хомогеним вегетацијским подручјима.

Поређењем резултата евалуација специфичних за селектовану групу са основним перформансама на читавом скупу података LUCAS 2018 ће нагласити ефикасност модела у хомогеним наспрам хетерогених окружења.

Даље, испитаће се да ли модел ради боље у областима са једним типом вегетационог покривача или одржава своју тачност у различитим областима земљишног покривача.

У табели 9 су приказани резултати евалуације хибридних 04 модела са алгоритмима са најбољим учинком за процену SOC у подручјима са хомогеним типовима вегетацијског покривача:

| Скуп/група | Алгоритам | MAE | MSE | R^2 | MAPE |
|---|----------------------------|--------|---------|---------|---------|
| LUCAS 2018 | Екстремно насумична стабла | 2.4319 | 26.4643 | 0.2869 | 91.6035 |
| | Насумичне шуме | 2.4078 | 26.8723 | 0.2759 | 87.4573 |
| | Стабла квантилне регресије | 2.0280 | 28.6324 | 0.2299 | 54.0013 |
| Група 1 В10 житарице | Екстремно насумична стабла | 1.0567 | 3.7180 | -0.9848 | 80.8960 |
| | Насумичне шуме | 1.0198 | 3.9826 | -1.0626 | 74.9520 |
| | Стабла квантилне регресије | 0.6597 | 2.3459 | 0.0197 | 43.2800 |
| Група 2 С10 листопадне шуме | Екстремно насумична стабла | 3.1283 | 32.1830 | 0.1307 | 94.3638 |
| | Насумичне шуме | 3.0846 | 32.6624 | 0.1416 | 90.9700 |
| | Стабла квантилне регресије | 2.7456 | 35.0728 | 0.0869 | 59.4723 |
| Група 3 Е20 травњаци без дрвећа/жбуња | Екстремно насумична стабла | 2.6227 | 24.6498 | -0.0343 | 84.5000 |
| | Насумичне шуме | 2.6454 | 26.2906 | -0.1077 | 82.1534 |
| | Стабла квантилне регресије | 2.1621 | 24.5360 | 0.0009 | 49.5587 |

Табела 9: Резултати евалуације хибридних 04 модела са алгоритмима са најбољим учинком за процену SOC у подручјима са хомогеним типовима вегетацијског покривача

Анализа перформанси хибридног модела 04 у процени садржаја SOC у различитим хомогеним типовима вегетације у скупу података LUCAS 2018 открива нијансиране увиде у могућности и ограничења модела. Метрике модела, добијене за одређене групе вегетације када се упореде са перформансама модела добијеним на целом скупу података, илуструју кључне аспекте тачности предвиђања модела.

Перформансе хибридног модела 04 за групу 1 (В10 житарице) скупа података LUCAS 2018, у поређењу са његовим укупним перформансама, откривају значајне увиде. Мучење модела са предвиђањима у овој категорији је очигледна из негативних R^2 вредности у алгоритмима, што указује на недостатак уклапања и тачности предвиђања у пејзажу којим доминирају житарице.

За алгоритам екстремно насумичне шуме, MAE од 1,0567 и MSE од 3,7179, заједно са негативним R^2 од -0,9848, указују на значајне грешке у предвиђању. Ово може произаћи из неспособности модела да адекватно ухвати јединствене карактеристике усева житарица.

Алгоритам стабла квантилне регресије је нешто бољи са MAE од 0,6597 и маргиналним R^2 од 0,0197, али MSE од 2,3459 што указује на изазове у доследној прецизности

предвиђања. Ово би могло да имплицира да, иако модел обухвата неке аспекте варијације SOC-а у житарицама, недостаје му свеобухватно разумевање.

Алгоритам насумичне шуме, који показује MAE од 1,0198 и MSE од 3,9826, има негативан R^2 од -1,0626, који наглашава његова ограничења у овом специфичном пољопривредном окружењу.

Насупрот томе, шира примена модела на целом LUCAS 2018 скупу података дала је приметно веће вредности R^2 и ниже метрике грешке. Овај диспарат наглашава потенцијалну прекомерну специјализацију модела за различите пејзаже, што захтева пречишћавање специфичних типова вегетације као што су житарице како би се побољшала његова примењивост у пољопривредној процени SOC-а.

Када се упореде перформансе хибридног модела 04 за групу 2 (C10 листопадне шуме) са укупним перформансама на скупу података LUCAS 2018, појављују се приметне разлике. У шумском окружењу, алгоритми су показали различите резултате: алгоритам екстремно насумичних стабала остварио је MAE од 3,1283, MSE од 32,1830 и низак R^2 од 0,1307, што указује на пад тачности модела.

Слично томе, алгоритам насумичне шуме остварио је MAE од 3,0846 и MSE од 32,6624, са нешто бољим R^2 од 0,1416.

Алгоритам стабла квантилне регресије је показао MAE од 2,7456, MSE од 35,0728 и релативно нижи R^2 од 0,0869, показујући нешто другачији образац перформанси у поређењу са алгоритмима екстремно насумичних стабала и насумичне шуме.

Насупрот томе, перформансе модела у целом скупу података су генерално биле робусније, са вишим вредностима R^2 и нижим метрикама грешке. Овај контраст наглашава изазове модела у прецизном предвиђању SOC-а у хомогенијем и специфичнијем шумском окружењу, наглашавајући потребу за прецизирањем модела за такве еколошке сценарије.

У процени перформанси хибридног модела 04 у групи 3 (E20 травњаци без дрвећа/жбуња), у поређењу са његовом широм применом у скупу података LUCAS 2018, појављују се одређени обрасци. Показатељи перформанси модела у групи 3, иако показују извесну доследност у стопама грешака, показују приметан пад тачности предвиђања, посебно у погледу вредности R^2 . Ово сугерише да се модел, оптимизован за разнолик скуп вегетационих покривача, сусреће са специфичним изазовима у хомогеном окружењу травњака. Контраст у метрикама перформанси, посебно вредности R^2 , наглашава ограничења модела у таквим униформним подешавањима. Варијације у MAPE у различитим алгоритмима такође наглашавају релативне разлике у тачности предвиђања, што указује на потребу за даљим усавршавањем модела како би се побољшала његова ефикасност у специфичним еколошким сценаријима као што су травњаци.

Упоређујући добијене резултате са перформансама модела тренираном на читавом скупу података LUCAS 2018, где је постигао приметно веће вредности R^2 и ниже метрике грешке, постаје очигледно да је хибридни модел 04 бољи у руковању различитим врстама вегетације уместо да се фокусира на специфичне, хомогене категорије. Снага модела лежи у његовој способности да интегрише и анализира широку лепезу обележја различитих типова земљишног покривача, што је благо смањено када се примени на хомогена вегетациона подручја.

8.5 Конструисање SOC мапе: примена модела и мозаик приступ

У завршној фази истраживања, коришћен је посебан мозаик метод како би се направила континуирана SOC мапа високе резолуције и демонстрирала практична примена предложеног приступа.

Свака улазна слика која је унета у прву фазу предложеног приступа је 64 x 64 пиксела, при чему сваки пиксел представља површину од 10 x 10 метара на тлу. Коришћено је свих 13 спектралних опсега Sentinel-2 сателита. Модел обрађује сваку улазну слику и производи излаз исте величине од 64 x 64 пиксела.

Међутим, са сваке излазне слике је одбачено по 8 пиксела са свих страна, што резултира ефективном величином излазне слике од 48 x 48 пиксела. Наведени метод одбацивања се користи да би се објаснио опадајући капацитет модела да одржи контекст како се приближава границама улазне слике.

Да би се формирала коначна континуирана SOC мапа, употребљена је техника клизног прозора. Ова техника обезбеђује да излазне слике (плочице) величине 48 x 48 пиксела буду савршено поређане једна уз другу, што резултира беспрекорном, континуираном мапом. Примена ове технике захтева пажљиво разматрање димензија улазне слике, димензија излазне слике и допуна да би се обезбедило тачно поравнање и континуитет коначне SOC мапе.

Коришћењем наведене методологије, могуће је искористити предиктивну моћ развијеног модела у целом пејзажу, генеришући тачну, високу резолуцију и континуирано представљање SOC садржаја широм региона од интереса, у овом случају Тоскане.

Када је мапа генерисана, употребљен је CORINE скуп података како би се креирала маска валидности за добијену SOC мапу.

Конкретно, коришћен је CLC2018. Јавно доступна и ажурна верзија CORINE скупа података. Подаци CLC2018 су коришћени у свом изворном растерском формату, да би се елиминисале SOC процене за класе земљишног покривача за које модел није био истрениран. Процене су ограничене на категорије 2 и 3 као што је приказано у табели 10. (Пољопривредне површине и шуме и полуприродне површине), као регионе од интереса у оквиру проучаваног подручја, док се све остало сматра неважећим површинама података.

| CLC ниво 1 | CLC ниво 2 |
|----------------------------------|---|
| 1. Вештачке површине | 1.1. Урбана подручја 1.2. Индустијске, комерцијалне и транспортне јединице 1.3. Рудници, депоније и градилишта 1.4. Вештачке непољопривредне области под вегетацијом |
| 2. Пољопривредне површине | 2.1. Обрадива земља 2.2. Трајни усеви 2.3. Пашњаци 2.4. Хетерогене пољопривредне површине |
| 3. Шуме и полу-природне површине | 3.1. Шуме 3.2. Групаације жбунасте и/или зељасте вегетације 3.3. Отворени простори са мало или нимало вегетације |
| 4. Мочваре | 4.1. Унутрашње мочваре 4.2. Приобалне мочваре |
| 5. Водна тела | 5.1. Унутрашње водене области 5.2. Морске водене области |

Табела 10: Класе CORINE скупа података

Резултат је растерски слој података који јасно раздваја култивисане површине и површине под вегетацијом од вештачких површина и површина под водом.

8.6 Генерисање и доступност развијеног производа

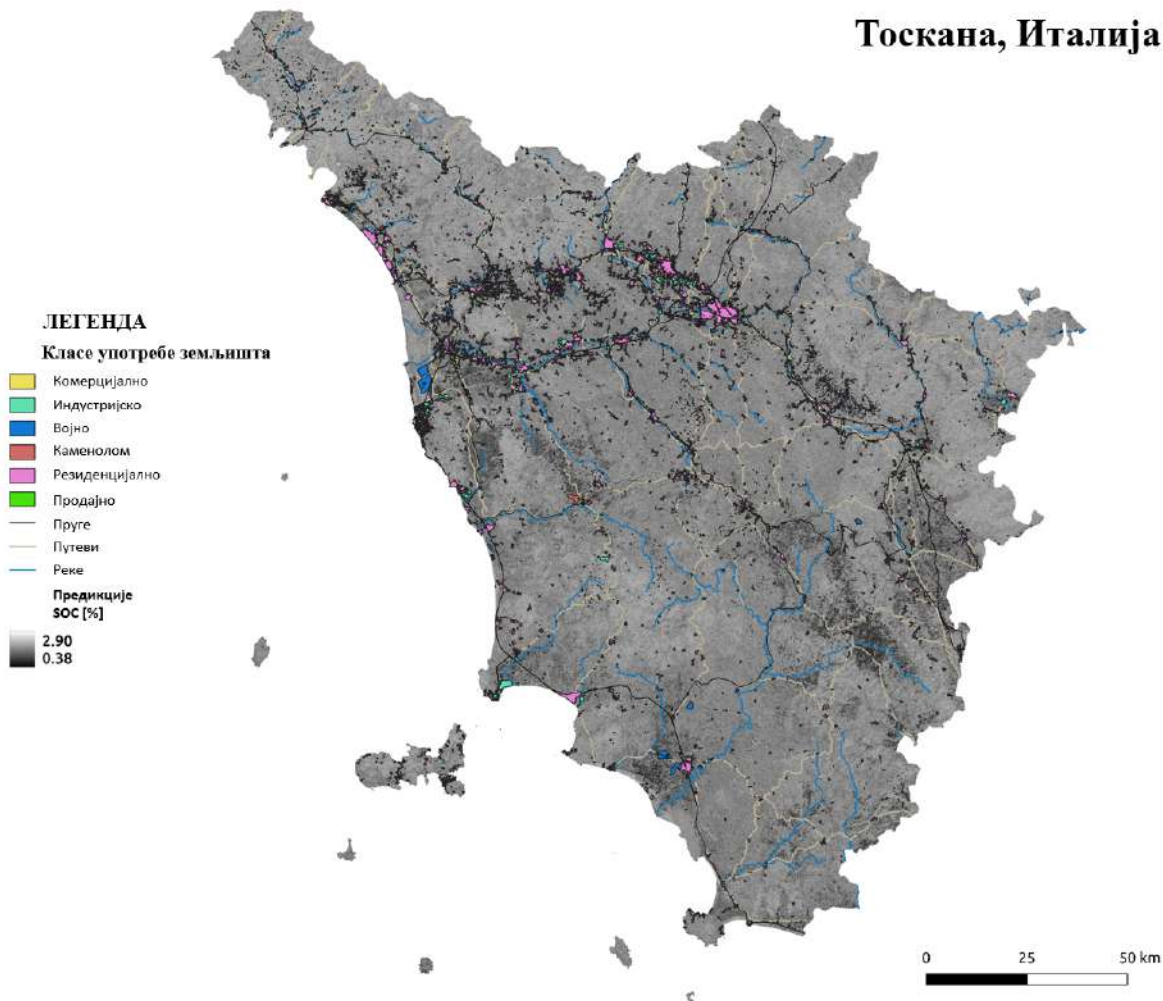
У спровођењу предложене методологије, успешно је генерисана континуирана карта органског угљеника у земљишту високе резолуције за регион Тоскане. SOC мапа, заједно са маском валидности заснованом на CORINE скупом података, представља вредан ресурс за разумевање дистрибуције SOC-а у односу на различите намене коришћења земљишта и класе земљишног покривача. Развијени производи, форматирани као GeoTIFF датотеке, могу се лако интегрисати у даље просторне анализе.

Како би се омогућило визуелно сагледавање дистрибуције SOC-а широм Тоскане, продукована је Слика 57. Наведена слика, под називом „Представљање вредности SOC-а у сивим тоновима са преклапајућом маском валидности“, илуструје садржај SOC-а широм региона са преклапањем маске валидности.

8.7 Дискусија

Иако даљинска детекција представља методологију која највише обећава у смислу постизања приступачног и глобалног праћења SOC-а, студије које се фокусирају искључиво на сателитске снимке су ограничене по обиму (тј. по броју узорака земљишта који се користе за развој и валидацију модела као и по обухваћеном подручју).

Тоскана, Италија



Слика 57: Представљање вредности SOC-а у сивим тоновима са преклапајућом маском валидности за регион Тоскане

Табела 11. приказује поређење између резултата овог истраживања и других релевантних приступа објављених у литератури. Треба напоменути да је највећи R^2 добијен у истраживању које су спровели Стајнберг и сарадници, али њихов приступ се ослања на хиперспектралне податке и ограничен је на само 81 сакупљени узорак, за који су лабораторијски одређене вредности, на површини од само 7km^2 . То чини наведени приступ веома тешким за процену колико је ово мерило учинка уопштено, тј. колико добро би генерализовало.

Слично томе, недавни приступ који су предложили Касталди и сарадници ограничава моделе на специфичне области, у распону од 2,2 до 425 ha, тако да се поставља питање како би се њихов приступ могао проширити на глобално или континентално праћење SOC-а, али оно остаје без одговора.

Предложени приступ се, међутим, скоро поклапа са перформансама које су остварили Стајнберг и сарадници на CHLSOC-у, који покрива област која је више од 100.000 пута већа и садржи 85 пута више података него што су користили Стајнберг и сарадници.

| Приступ | Број узорака | RMSE | R ² |
|----------------------------------|--------------|------|----------------|
| Гомез и сарадници [12] | 146 | 0.73 | 0.51 |
| Ванг и сарадници [160] | 705 | / | 0.47 |
| Стајнберг и сарадници [18] | 81 | 0.21 | 0.74 |
| Јанг и сарадници [23] | 733 | / | 0.26 |
| Емади и сарадници [24] | 1879 | 0.75 | 0.65 |
| Касталди и сарадници [25] | 53–131 | 0.15 | 0.67 |
| Предложени приступ на CHLSOC | 6.814 | 2.23 | 0.757 |
| Предложени приступ на LUCAS 2015 | 12.080 | 3.37 | 0.57 |
| Предложени приступ на LUCAS 2018 | 16.232 | 4.36 | 0.49 |
| Предложени приступ укупно | 35.126 | 3.83 | 0.53 |

Табела 11: Поређење са објављеним истраживањима

ци, чинећи евалуацију урађену у овде представљеном истраживању много опширнијом. Од претходних студија, најближа по величини у смислу броја узорака је она Емадија и сарадника, али њихов скуп података је и даље неколико редова величине мањи и студија је ограничена на северни Иран. Међутим, када се изврши евалуација на целокупном скупу података који се састоји од CHLCOS, LUCAS 2015 и LUCAS 2018 скупова података, учинак представљен у овом истраживању је близу учинка који су постигли Емади и сарадници у смислу варијансе објашњене развијеним моделом (R²). Ваљано поређење у смислу RMSE је практично немогуће постићи, пошто опсег SOC у различитим студијама драстично варира.

У табели 8. такође је приказана средња апсолутна процентуална грешка као додатна метрика за процену учинка предложеног приступа. Ово омогућава да закључимо да у просеку истренирани модел промашује SOC вредност за 50.44%. Иако је ово свакако велика граница грешке, то је такође значајно побољшање када се узме у обзир да тренутне глобалне процене SOC варирају између 504 и 3000 Pg C [96].

8.8 Ограничења предложеног приступа

Иако је ово истраживање дало значајне резултате, важно је признати његова ограничења.

За експеримент 1 (Хипотеза 1), приступ коришћења различитих модела машинског учења за процену органског угљеника у земљишту са сателитских снимака Sentinel-2 има неколико изазова. Разноликост модела, укључујући базичне и хибридне типове, може довести до различитих степена интерпретабилности и доследности у резултатима.

Базични модели директно користе 13 опсега са сателита, док се хибридни модели ослањају на латентна обележја екстрахована преко U-Net мреже за сегментацију. Ова бифуркација поставља питања о релативној ефикасности и робусности ова два приступа.

Штавише, процес издвајања обележја и накнадно ослањање на ограничен број (192) латентних обележја можда неће свеобухватно да обухвати сложеност и варијабилност фактора везаних за SOC у посматраним географским регионима.

Такође, кораци претходне обраде података, као што је нормализација вредности пиксела за сваки од 13 опсега Sentinel-2 снимака и уклањање тачака података са недостајућим подацима, такође могу да уведу предрасуде или умање комплетност скупа података, потенцијално утичући на перформансе модела и могућност генерализације.

Експеримент 1 (Хипотеза 2) укључивање атрибута терена и одабраних индекса даљинске детекције у модел уводи додатна ограничења. Избор специфичних атрибута терена (закривљеност, нагиб, аспект, надморска висина) и индекса даљинске детекције (NDVI, NDBI, NDVI, BSI SAVI) обликује начин како модел разуме и представља факторе животне средине.

Иако су ови атрибути и индекси вредни, они представљају ограничен избор од великог броја фактора који утичу на SOC. Они можда неће у потпуности ухватити сложене варијације терена или комплетне спектралне карактеристике које утичу на динамику SOC.

Штавише, ефикасност ових атрибута и индекса може варирати у различитим пејзажима и условима животне средине. Ова варијабилност може ограничити генерализацију и тачност модела у представљању SOC-а у различитим еколошким окружењима.

Да би се ово решило, модел би могао имати користи од ширег дијапазона теренских и спектралних обележја, што би довело до свеобухватније репрезентације фактора животне средине који утичу на SOC.

У Експерименту 2 (Хипотеза 1 и 2) фокус се помера на генерализацију модела са најбољим учинком из Експеримента 1. Међутим, процена способности генерализације модела је ограничена на употребу три алгорита и три скупа података (CHLSOC, LUCAS 2015 и LUCAS 2018).

Овај ограничени обим можда неће у потпуности обухватити варијабилност и хетерогеност глобалних типова и услова земљишта. Специјализована 10-струка унакрсна валидација, иако је иновативна, још увек можда неће на адекватан начин да се позабави потенцијалом претераног прилагођавања или недовољне заступљености одређених услова тла, земљишних покривача или географских подручја.

Експеримент 3 (Хипотеза 3) истражује перформансе најбољег укупног модела, тачније хибридног модела 04, у областима са хомогеним типовима вегетационог покривача. Овај фокус на хомогене типове вегетације представља значајно ограничење, јер не узима у обзир сложеност и разноврсност природног окружења, где су мешовити типови вегетације уобичајени.

Штавише, искључивање скупова података као што су CHLSOC и LUCAS 2015 због проблема са доступношћу података додатно сужава обим студије, потенцијално утичући на робусност и применљивост налаза. Ослањање на скуп података LUCAS 2018 за овај експеримент, иако је разумљиво, можда неће пружити свеобухватан поглед на перформансе модела у различитим еколошким окружењима.

У глобалу, ослањање на специфичне скупове података, ограничен обим тестираних алгоритама и фокус на одређене типове земљишног покривача могу ограничити применљивост студије на шире, разноврсније географске и еколошке контексте.

Поред тога, сложеност и варијабилност фактора који утичу на SOC, као што су клима, праксе управљања земљиштем и типови земљишта, можда неће бити у потпуности обухваћени коришћеним моделима.

Репрезентативност одабраних скупова података за глобалну процену SOC-а је тако-

ђе ограничавајућа јер обухватају само земље Европске Уније и једну земљу из јужне Америке, као и потенцијал за проблеме преносивости модела у различитим регионима или условима.

Приступ предложен у овом истраживању заснива се на коришћењу једног сателитског снимка за процену SOC-а, што занемарује временске варијације органског угљеника у земљишту. SOC није статички атрибут, он флукутира због различитих фактора као што су промене годишњих доба, пољопривредне праксе, климатски услови и услед природних догађаја попут поплава или суше. Ови динамички процеси могу значајно изменити садржај SOC у краткорочним и дугорочним периодима.

Недостатак временске анализе значи да су предвиђања модела у суштини снимак одређеног тренутка у времену, што може потенцијално да доведе до погрешног представљања праве, флукутирајуће природе нивоа SOC.

Овај приступ коришћења једног снимка може бити погрешан у сценаријима у којима су недавни догађаји драстично променили нивое SOC-а, али такве промене нису обухваћене анализом једног снимка.

Штавише, ослањање на један снимак не узима у обзир историјске промене коришћења земљишта или претходне пољопривредне праксе, које би могле имати дуготрајне утицаје на нивое SOC.

Укључивање временске динамике у модел, што би било могуће путем уврштавања временских серија сателитских података или историјских записа о коришћењу земљишта, могло би значајно побољшати његову тачност предвиђања и релевантност, посебно у контексту климатских промена и њиховог утицаја на плодност земљишта и секвестрацију угљеника.

Ово побољшање би омогућило динамичнији модел који је способан да се прилагоди стално променљивој природи нивоа SOC-а у различитим окружењима.

Употреба сателитских снимака Sentinel-2 представља инхерентна ограничења због своје просторне резолуције. Иако нуди равнотежу између покривености и детаља, можда неће бити довољно за прецизно хватање малих, сигнозних варијација у SOC-у. Ово је посебно релевантно у хетерогеним пределима где дистрибуција SOC може значајно да варира на кратким удаљеностима.

Таква ограничења могу резултирати превеликим поједностављењем процена SOC-а, потенцијално занемарујући микро-размерне интеракције животне средине које утичу на нивое SOC-а. Побољшање резолуције или допуна подацима веће резолуције тамо где је то потребно може побољшати тачност модела у различитим пејзажима.

Даљинска детекција претежно хвата информације на површинском нивоу, које могу бити индиректно повезане са SOC садржајем. Међутим, даљинском детекцијом се не мере директно својства тла испод земље као што су текстура, минерални састав и микробна активност, а све су то битни фактори за тачну процену SOC.

Поред тога, спољни фактори као што су праксе управљања земљиштем, историјско коришћење земљишта и климатске промене играју значајну улогу у одређивању нивоа SOC, али се не могу лако уочити путем сателитских снимака. Одсуство ових фактора у анализи може довести до јаза између посматраних и стварних нивоа SOC-а, што утиче на тачност и применљивост модела.

Потенцијал предложеног приступа би могао бити додатно побољшан интеграцијом сателитских података са другим релевантним изворима података. Укључивање додат-

них мерења тла на земљи, историјских података о коришћењу земљишта и локалних климатских записа могло би да пружи холистичкији поглед на динамику SOC.

Ова интеграција би могла да помогне у потврђивању и побољшању процена заснованих на даљинском сензору, што би довело до прецизнијих и свеобухватнијих SOC модела. То би такође омогућило моделу да узме у обзир шири спектар еколошких и антропогених фактора који утичу на SOC, чиме би се побољшала његова робусност и применљивост у различитим условима.

9 ЗАКЉУЧНА РАЗМАТРАЊА

Истраживање уводи иновативну, хибридную методологију која комбинује различите моделе машинског учења, укључујући дубоке неуронске мреже и традиционалне алгоритме за машинско учење, за прецизну процену органског угљеника у земљишту употребом сателитских снимака.

Овај нови приступ је усидрен у двослојном оквиру, где почетна фаза укључује софистицирано издвајање обележја помоћу DNN-а, посебно користећи U-Net архитектуру за сегментацију. Ова фаза користи моћ мултиспектралног снимања са сателита Sentinel-2, повезујући ове високодимензионалне податке са аутентичним мапама сегментације како би се проценило коришћење земљишта у циљаним географским областима.

Ефикасност DNN-а у обради сложених сателитских података је јасно демонстрирана, означавајући значајан напредак у техникама даљинске детекције.

Друга фаза модела примењује екстраховане векторе обележја за процену SOC-а, у тренирању низа ансамбл метода за машинско учење. Овај корак представља пример трансформације замршених и индиректних обележја података у практичне увиде о одређеним географским регионима, наглашавајући прилагодљивост и моћ модела машинског учења у анализи животне средине.

Ефикасност модела је ригорозно процењена кроз десетоструку унакрсну валидацију у односу на постојећа SOC мерења, показујући његове високе перформансе и поузданост.

Битно достигнуће овог истраживања представља и стварање просторно континуиране SOC карте високе резолуције (10 метара) за регион Тоскане, у северној Италији. Ова детаљна SOC процена, интегрисана са званичном мапом класификације коришћења земљишта у Тоскани, показује потенцијал модела машинског учења у пружању свеобухватних и прецизних SOC евалуација. Такво мапирање високе резолуције представља значајан скок у могућностима праћења животне средине.

Додатну вредност овог истраживања представља чињеница да не подразумева само академски допринос већ и практичан алат – метод и софтверско решење дизајнирано за ефикасну и тачну процену SOC-а. Развијени алат користи моћ дубоких неуронских мрежа и података добијених путем даљинске детекције, нудећи ресурсно ефикаснију алтернативу преовлађујућим методама процене SOC.

Истраживање означава велики корак ка глобално скалабилним, исплативим и даљински управљаним системима за праћење SOC-а. Представља пример интеграције напредних сателитских снимака и савремених метода машинског учења у науци о животној средини, утирући пут за опсежна будућа истраживања и практичне примене.

Импликације овог истраживања су дубоке, са потенцијалом до доведу до значајних промена у праксама праћења и управљања SOC-ом на глобалном нивоу, и значајно доприносећи очувању животне средине и одрживом планирању коришћења земљишта.

Надовезивањем на почетни успех у Тоскани, ово истраживање поставља терен за ширу примену у различитим географским регионима. Може се истражити прилагодљивост методе различитим пејзажима и типовима тла, потенцијално нудећи универзално применљиво средство за управљање земљиштем.

Идентификовањем варијација у SOC-у на различитим теренима и климатским условима, модел може подржати циљане напоре очувања животне средине, доприносећи

одрживим праксама управљања земљиштем на глобалном нивоу.

Истраживање наглашава кључну улогу SOC-а у глобалном циклусу угљеника. Израђене детаљне SOC карте могу бити саставни део развоја свеобухватних климатских акционих планова.

Одређивањем области са значајним потенцијалом секвестрације угљеника, модел може да се употреби као путоказ за пројекте пошумљавања или обнове земљишта, директно доприносећи смањењу нивоа CO₂ у атмосфери. Интеграција ових SOC мапа са климатским моделима могла би понудити нове увиде у интеракцију између здравља тла и климатских промена.

Детаљно SOC мапирање може довести до значајних промена у пракси прецизне пољопривреде. Пољопривредници и пољопривредни планери могу да користе ове мапе да прилагоде пољопривредне праксе специфичним условима земљишта, оптимизују употребу ђубрива, побољшају стратегије ротирања усева и побољшају управљање водом.

Овај унапређени ниво прецизности може значајно повећати приносе усева уз минимизирање утицаја на животну средину, доприносећи тако глобалној безбедности хране.

Способност прецизног праћења и процене SOC-а има значајне импликације на доношење закона и формулисање политика. Владе и међународна тела могу да користе резултате модела за доношење одлука о легислативама у вези са коришћењем земљишта, пољопривредним субвенцијама и програмима очувања животне средине.

Ово би могло довести до ефикаснијег креирања политика заснованих на доказима и чињеницама, обезбеђујући да се ресурси ефикасно додељују областима којима су најпотребнији напори за очување и побољшање земљишта.

Гледајући унапред, постоји неколико путева за даља истраживања у оквиру развоја модела. Интеграција новијих сателитских технологија са вишом резолуцијом и додатним спектралним опсезима могла би побољшати тачност модела.

Истраживање инкорпорације вештачке интелигенције и аналитике великих података могло би додатно да побољша процес процене SOC-а. Поред тога, развој динамичнијег модела који укључује временске промене у SOC би решио једно од значајних ограничења тренутног приступа.

Осим научне и практичне примене, истраживање може бити значајно у контексту образовања и ангажовања шире заједнице. Чинећи податке и налазе доступним јавности, може се подићи свест о важности здравља земљишта и његовој повезаности са ширим питањима животне средине. Сарадња са школама, универзитетима и локалним заједницама може подстаћи боље разумевање и ангажовање у пракси одрживог коришћења земљишта.

Сprovedено истраживање представља значајан напредак у области даљинске детекције и науке о земљишту. Употребом моћи машинског учења и сателитских снимака, нуди нови приступ разумевању и управљању земљишним ресурсима.

Практичне примене овог истраживања су огромне, у распону од повећања пољопривредне продуктивности до информисања о стратегијама климатских акција. Како модел наставља да се развија и побољшава, он ће имати битан утицај на наше разумевање и управљање планетом.

10 ЛИТЕРАТУРА

- [1] J. P. Scharlemann, E. V. Tanner, R. Hiederer и V. Kapos, “Global soil carbon: understanding and managing the largest terrestrial carbon pool,” *Carbon Management*, св. 5, бр. 1, стр. 81–91, 2014.
- [2] N. Batjes, “Harmonized soil profile data for applications at global and continental scales: updates to the WISE database,” *Soil Use and Management*, св. 25, бр. 2, стр. 124–127, 2009.
- [3] N. H. Batjes, “Total carbon and nitrogen in the soils of the world,” *European journal of soil science*, св. 47, бр. 2, стр. 151–163, 1996.
- [4] V. Meentemeyer, J. Gardner и E. O. Box, “World patterns and amounts of detrital soil carbon,” *Earth Surface Processes and Landforms*, св. 10, бр. 6, стр. 557–567, 1985.
- [5] R. Mäkipää, M. Häkkinen, P. Muukkonen и M. Peltoniemi, “The costs of monitoring changes in forest soil carbon stocks,” 2008.
- [6] B. Ellert, H. Janzen, B. Vandenbygaart и E. Bremer, “Measuring Change in Soil Organic Carbon Storage,” *Soil Sampling and Methods of Analysis*, стр. 25–38, ян. 2008. DOI: 10.1201/9781420005271.ch3.
- [7] J. Li, Y. Pei, S. Zhao, R. Xiao, X. Sang и C. Zhang, “A review of remote sensing for environmental monitoring in China,” *Remote Sensing*, св. 12, бр. 7, стр. 1130, 2020. DOI: 10.3390/rs12071130.
- [8] J. T. Overpeck, G. A. Meehl, S. Bony и D. R. Easterling, “Climate data challenges in the 21st century,” *science*, св. 331, бр. 6018, стр. 700–702, 2011. DOI: 10.1126/science.1197869.
- [9] M. Pavlovic, S. Ilic, N. Antonic и D. Culibrk, “Monitoring the Impact of Large Transport Infrastructure on Land Use and Environment Using Deep Learning and Satellite Imagery,” *Remote Sensing*, св. 14, бр. 10, стр. 2494, 2022.
- [10] T. Kattenborn, J. Leitloff, F. Schiefer и S. Hinz, “Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, св. 173, стр. 24–49, 2021. DOI: 10.1016/j.isprsjprs.2020.12.010.
- [11] C. Toth и G. Józków, “Remote sensing platforms and sensors: A survey,” *ISPRS Journal of Photogrammetry and Remote Sensing*, св. 115, стр. 22–36, 2016. DOI: <https://doi.org/10.1016/j.isprsjprs.2015.10.004>.
- [12] C. Gomez, R. A. V. Rossel и A. B. McBratney, “Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study,” *Geoderma*, св. 146, бр. 3-4, стр. 403–411, 2008.
- [13] B. Wang, C. Waters, S. Orgill и др., “Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia,” *Ecological indicators*, св. 88, стр. 425–438, 2018.
- [14] L. Breiman, “Random forests,” *Machine learning*, св. 45, стр. 5–32, 2001.

- [15] J. Elith, J. R. Leathwick и Т. Hastie, “A working guide to boosted regression trees,” *Journal of animal ecology*, св. 77, бр. 4, стр. 802–813, 2008.
- [16] С. Cortes и V. Vapnik, “Support-vector networks,” *Machine learning*, св. 20, стр. 273–297, 1995.
- [17] F. Castaldi, A. Palombo, F. Santini, S. Pascucci, S. Pignatti и R. Casa, “Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon,” *Remote Sensing of Environment*, св. 179, стр. 54–65, 2016.
- [18] A. Steinberg, S. Chabrillat, A. Stevens, K. Segl и S. Foerster, “Prediction of common surface soil properties based on Vis-NIR airborne and simulated EnMAP imaging spectroscopy data: Prediction accuracy and influence of spatial resolution,” *Remote Sensing*, св. 8, бр. 7, стр. 613, 2016.
- [19] I. Goodfellow, Y. Bengio и A. Courville, *Deep learning*. MIT press, 2016.
- [20] Q. Yuan, H. Shen, T. Li и др., “Deep learning in environmental remote sensing: Achievements and challenges,” *Remote Sensing of Environment*, св. 241, стр. 111–116, 2020.
- [21] R. Girshick, J. Donahue, T. Darrell и J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” у *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014., стр. 580–587.
- [22] A. Rakhlin, A. Davydov и S. Nikolenko, “Land cover classification from satellite imagery with u-net and lovasz-softmax loss,” у *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018., стр. 262–266.
- [23] L. Yang, Y. Cai, L. Zhang, M. Guo, A. Li и C. Zhou, “A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables,” *International Journal of Applied Earth Observation and Geoinformation*, св. 102, стр. 102–113, 2021.
- [24] M. Emadi, R. Taghizadeh-Mehrjardi, A. Cherati, M. Danesh, A. Mosavi и T. Scholten, “Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran,” *Remote Sensing*, св. 12, бр. 14, стр. 2234, 2020.
- [25] F. Castaldi, M. H. Koparan, J. Wetterlind и др., “Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands,” *ISPRS Journal of Photogrammetry and Remote Sensing*, св. 199, стр. 40–60, 2023.
- [26] M. Von Unger и I. Emmer, “Carbon market incentives to conserve, restore and enhance soil carbon,” *Arlington: Silvestrum and The Nature Conservancy*, 2018.
- [27] S. Fuss, W. F. Lamb, M. W. Callaghan и др., “Negative emissions—Part 2: Costs, potentials and side effects,” *Environmental Research Letters*, св. 13, бр. 6, стр. 063002, 2018.
- [28] K. Hamrick и M. Gallant, “Fertile ground: State of forest carbon finance 2017,” *Forest Trends’ Ecosystem Marketplace*, Washington, DC, св. 20036, 2017.

- [29] S. Vermeulen, D. Bossio, J. Lehmann и др., “A global agenda for collective action on soil carbon,” *Nature Sustainability*, св. 2, бр. 1, стр. 2–4, 2019.
- [30] M. V. Chiriaco, L. Perugini, M. Bellotta, L. Kaugure и M. Bernoux, *Koronivia Joint Work on Agriculture: Analysis of submissions on topics 2 (e) and 2 (f)*. Food & Agriculture Org., 2021.
- [31] P. Smith, “Soil carbon sequestration and biochar as negative emission technologies,” *Global change biology*, св. 22, бр. 3, стр. 1315–1324, 2016.
- [32] W. Ertel, *Introduction to artificial intelligence*. Springer, 2018.
- [33] Y. Lu, “Artificial intelligence: a survey on evolution, models, applications and future trends,” *Journal of Management Analytics*, св. 6, бр. 1, стр. 1–29, 2019.
- [34] I. Asimov, “Runaround,” *Astounding science fiction*, св. 29, бр. 1, стр. 94–103, 1942.
- [35] M. Haenlein и A. Kaplan, “A brief history of artificial intelligence: On the past, present, and future of artificial intelligence,” *California management review*, св. 61, бр. 4, стр. 5–14, 2019.
- [36] K. Marx, “Encyclopedia britannica,” *Encyclopaedia Britannica Ultimate Reference Suite [M/CD]*. Chicago: Encyclopaedia Britannica, 2012.
- [37] E. Rich, *Artificial intelligence*. McGraw-Hill, Inc., 1983.
- [38] P. N. Russell, “Artificial Intelligence: A Modern Approach by Stuart,” Russell and Peter Norvig contributing writers, Ernest Davis...[et al.], 2010.
- [39] P. Jackson, “Introduction to expert systems,” 1986.
- [40] H. Jiang, *Machine Learning Fundamentals: A Concise Introduction*. Cambridge University Press, 2021.
- [41] T. M. Mitchell и T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997., св. 1.
- [42] A. L. Samuel, “Some studies in machine learning using the game of checkers. II—Recent progress,” *IBM Journal of research and development*, св. 11, бр. 6, стр. 601–617, 1967.
- [43] R. Bellman, “Dynamic programming,” *Science*, св. 153, бр. 3731, стр. 34–37, 1966.
- [44] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE transactions on information theory*, св. 14, бр. 1, стр. 55–63, 1968.
- [45] R. D. Hof, *Breakthrough Technologies 2013: Deep Learning*, 2013.
- [46] X. X. Zhu, D. Tuia, L. Mou и др., “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geoscience and Remote Sensing Magazine*, св. 5, бр. 4, стр. 8–36, 2017.
- [47] J. Lintz и D. Simonett, “Remote Sensing of environment Addison Wesley,” Reading, Mass, 1976.
- [48] J. B. Campbell и R. H. Wynne, *Introduction to remote sensing*. Guilford Press, 2011.
- [49] R. S. Sutton и A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

- [50] J. Rissanen, “Modeling by shortest data description,” *Automatica*, св. 14, бр. 5, стр. 465–471, 1978.
- [51] D. H. Wolpert, “The lack of a priori distinctions between learning algorithms,” *Neural computation*, св. 8, бр. 7, стр. 1341–1390, 1996.
- [52] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, св. 55, бр. 10, стр. 78–87, 2012.
- [53] A. Borkar, M. Hayes и M. T. Smith, “A novel lane detection system with efficient ground truth generation,” *IEEE Transactions on Intelligent Transportation Systems*, св. 13, бр. 1, стр. 365–374, 2011.
- [54] A. H. Maslow, “The psychology of science a reconnaissance,” 1966.
- [55] S. Sayad, Logistic Regression, 01 15, 2024, 2019. адреса: https://saedsayad.com/logistic_regression.htm.
- [56] J. R. Quinlan, “Decision trees and decision-making,” *IEEE Transactions on Systems, Man, and Cybernetics*, св. 20, бр. 2, стр. 339–346, 1990.
- [57] J. Ali, R. Khan, N. Ahmad и I. Maqsood, “Random forests and decision trees,” *International Journal of Computer Science Issues (IJCSI)*, св. 9, бр. 5, стр. 272, 2012.
- [58] S. L. Mirtaheri и R. Shahbazian, *Machine Learning: Theory to Applications*. CRC Press, 2022.
- [59] Y. LeCun, Y. Bengio и G. Hinton, “Deep learning,” *nature*, св. 521, бр. 7553, стр. 436–444, 2015.
- [60] A. Ng и др., “Sparse autoencoder,” *CS294A Lecture notes*, св. 72, бр. 2011, стр. 1–19, 2011.
- [61] I. Goodfellow, J. Pouget-Abadie, M. Mirza и др., “Generative adversarial networks,” *Communications of the ACM*, св. 63, бр. 11, стр. 139–144, 2020.
- [62] A. Voulodimos, N. Doulamis, A. Doulamis и E. Protopapadakis, “Deep learning for computer vision: A brief review,” *Computational intelligence and neuroscience*, св. 2018, 2018.
- [63] M. Nabati, H. Navidan, R. Shahbazian, S. A. Ghorashi и D. Windridge, “Using synthetic data to enhance the accuracy of fingerprint-based localization: A deep learning approach,” *IEEE Sensors Letters*, св. 4, бр. 4, стр. 1–4, 2020.
- [64] L. Deng, G. Hinton и B. Kingsbury, “New types of deep neural network learning for speech recognition and related applications: An overview,” *у 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, 2013.*, стр. 8599–8603.
- [65] D. Shen, G. Wu и H.-I. Suk, “Deep learning in medical image analysis,” *Annual review of biomedical engineering*, св. 19, стр. 221, 2017.
- [66] D. T. Nguyen, F. Alam, F. Ofi и M. Imran, “Automatic image filtering on social networks using deep learning and perceptual hashing during crises,” *arXiv preprint arXiv:1704.02602*, 2017.

- [67] S. Min, B. Lee и S. Yoon, “Deep learning in bioinformatics,” *Briefings in bioinformatics*, св. 18, бр. 5, стр. 851–869, 2017.
- [68] D. Silver, A. Huang, C. J. Maddison и др., “Mastering the game of Go with deep neural networks and tree search,” *nature*, св. 529, бр. 7587, стр. 484–489, 2016.
- [69] Y. Bengio и др., “Learning deep architectures for AI,” *Foundations and trends® in Machine Learning*, св. 2, бр. 1, стр. 1–127, 2009.
- [70] C. Yeola, Convolutional Neural Network (CNN) In Deep Learning Image Processing Using Python, Accessed: 03.01.2024., 2022. адреса: <https://python.plainenglish.io/convolution-neural-network-cnn-in-deep-learning-77f5ab457166>.
- [71] F. V. Veen и S. Leijnen, The Neural Network Zoo, Accessed: 03.01.2024., 2019. адреса: <https://www.asimovinstitute.org/neural-network-zoo>.
- [72] A. Vaswani, N. Shazeer, N. Parmar и др., “Attention is All you Need,” у *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio и др., уредници, св. 30, Curran Associates, Inc., 2017. адреса: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [73] M. A. Kramer, “Nonlinear principal component analysis using autoassociative neural networks,” *AIChE journal*, св. 37, бр. 2, стр. 233–243, 1991.
- [74] G. Litt, ENHANCE!: Upscaling images CSI-style with generative adversarial neural networks, Accessed: 03.01.2024., 2017. адреса: <https://www.geoffreylytt.com/2017/06/04/enhance-upscaling-images-with-generative-adversarial-neural-networks>.
- [75] O. Ronneberger, P. Fischer и T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” у *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015., стр. 234–241.
- [76] M. F. L. Ornelas, “The Mexican Water Forest: benefits of using remote sensing techniques to assess changes in land use and land cover,” докторска теза, University of San Francisco, 2016.
- [77] E. S. Agency, Sentinel-2 Operations, Accessed: 04.01.2024., 2015. адреса: https://www.esa.int/Enabling_Support/Operations/Sentinel-2_operations.
- [78] H. Janzen, “Carbon cycling in earth systems—a soil science perspective,” *Agriculture, ecosystems & environment*, св. 104, бр. 3, стр. 399–417, 2004.
- [79] U. C. for Science Education, Carbon Cycle Diagram, Accessed: 08.01.2024., 2024. адреса: <https://scied.ucar.edu/image/carbon-cycle>.
- [80] R. Lal, “Soil carbon sequestration impacts on global climate change and food security,” *science*, св. 304, бр. 5677, стр. 1623–1627, 2004.
- [81] R. Jandl, M. Rodeghiero, C. Martinez и др., “Current status, uncertainty and future needs in soil organic carbon monitoring,” *Science of the total environment*, св. 468, стр. 376–383, 2014.

- [82] L. Zhang, Y. Liu, X. Li и др., “Effects of soil map scales on simulating soil organic carbon changes of upland soils in Eastern China,” *Geoderma*, св. 312, стр. 159–169, 2018.
- [83] U. Stockmann, J. Padarian, A. McBratney и др., “Global soil organic carbon assessment,” *Global Food Security*, св. 6, стр. 9–16, 2015.
- [84] E. Berryman, J. Hatten, D. S. Page-Dumroese и др., “Soil carbon,” у *Forest and rangeland soils of the United States under changing conditions*, Springer, Cham, 2020., стр. 9–31.
- [85] A. Koch, A. McBratney, M. Adams и др., “Soil security: solving the global soil crisis,” *Global Policy*, св. 4, бр. 4, стр. 434–441, 2013.
- [86] S. H. Schoenholtz, H. Van Miegroet и J. Burger, “A review of chemical and physical properties as indicators of forest soil quality: challenges and opportunities,” *Forest ecology and management*, св. 138, бр. 1-3, стр. 335–356, 2000.
- [87] E. E. Oldfield, S. A. Wood и M. A. Bradford, “Direct effects of soil organic matter on productivity mirror those observed with organic amendments,” *Plant and Soil*, св. 423, бр. 1, стр. 363–373, 2018.
- [88] B. Minasny, B. P. Malone, A. V. McBratney и др., “Soil carbon 4 per mille,” *Geoderma*, св. 292, стр. 59–86, 2017.
- [89] S. S. Andrews, D. L. Karlen и C. A. Cambardella, “The soil management assessment framework: A quantitative soil quality evaluation method,” *Soil Science Society of America Journal*, св. 68, бр. 6, стр. 1945–1962, 2004.
- [90] P. Smith, D. Martino, Z. Cai и др., “Greenhouse gas mitigation in agriculture,” *Philosophical transactions of the royal Society B: Biological Sciences*, св. 363, бр. 1492, стр. 789–813, 2008.
- [91] D. Bossio, S. Cook-Patton, P. Ellis и др., “The role of soil carbon in natural climate solutions,” *Nature Sustainability*, св. 3, бр. 5, стр. 391–398, 2020.
- [92] P. Smith, J. Adams, D. J. Beerling и др., “Land-management options for greenhouse gas removal and their impacts on ecosystem services and the sustainable development goals,” *Annual Review of Environment and Resources*, св. 44, стр. 255–286, 2019.
- [93] J. Sanderman, T. Hengl и G. J. Fiske, “Soil carbon debt of 12,000 years of human land use,” *Proceedings of the National Academy of Sciences*, св. 114, бр. 36, стр. 9575–9580, 2017.
- [94] D. Jenkinson, D. Adams и A. Wild, “Model estimates of CO₂ emissions from soil in response to global warming,” *Nature*, св. 351, бр. 6324, стр. 304–306, 1991.
- [95] C. E. Hicks Pries, C. Castanha, R. Porras и M. Torn, “The whole-soil carbon flux in response to warming,” *Science*, св. 355, бр. 6332, стр. 1420–1423, 2017.
- [96] K. E. Todd-Brown, J. T. Randerson, W. M. Post и др., “Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations,” *Biogeosciences*, св. 10, бр. 3, стр. 1717–1736, 2013.

- [97] R. Hiederer, M. Köchy и др., “Global soil organic carbon estimates and the harmonized world soil database,” *EUR*, св. 79, бр. 25225, стр. 10–2788, 2011.
- [98] S. Solomon, D. Qin, M. Manning и др., “Climate change 2007: Synthesis Report. Contribution of Working Group I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers.,” *Climate change 2007: Synthesis Report. Contribution of Working Group I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers.*, 2007.
- [99] R. A. Houghton, J. I. House, J. Pongratz и др., “Carbon emissions from land use and land-cover change,” *Biogeosciences*, св. 9, бр. 12, стр. 5125–5142, 2012.
- [100] G. P. Peters, G. Marland, C. Le Quéré, T. Boden, J. G. Canadell и M. R. Raupach, “Rapid growth in CO₂ emissions after the 2008–2009 global financial crisis,” *Nature climate change*, св. 2, бр. 1, стр. 2–4, 2012.
- [101] A. Ruesch и H. Gibbs, “New global biomass carbon map for the year 2000 based on IPCC Tier-1 methodology,” Oak Ridge National Laboratory’s Carbon Dioxide Information Analysis Center: Oak Ridge, USA. Available online from the Carbon Dioxide Information Analysis Center, URL: <http://cdiac.ornl.gov>, 2008.
- [102] S. S. Saatchi, N. L. Harris, S. Brown и др., “Benchmark map of forest carbon stocks in tropical regions across three continents,” *Proceedings of the national academy of sciences*, св. 108, бр. 24, стр. 9899–9904, 2011.
- [103] A. Baccini, S. Goetz, W. Walker и др., “Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps,” *Nature climate change*, св. 2, бр. 3, стр. 182–185, 2012.
- [104] N. L. Harris, S. Brown, S. C. Hagen и др., “Baseline map of carbon emissions from deforestation in tropical regions,” *Science*, св. 336, бр. 6088, стр. 1573–1576, 2012.
- [105] G. R. Van der Werf, D. C. Morton, R. S. DeFries и др., “CO₂ emissions from forest loss,” *Nature geoscience*, св. 2, бр. 11, стр. 737–738, 2009.
- [106] C. Le Quéré, R. J. Andres, T. Boden и др., “The global carbon budget 1959–2011,” *Earth System science data*, св. 5, бр. 1, стр. 165–185, 2013.
- [107] S. Lamtom и R. Savidge, “A reassessment of carbon content in wood: variation within and between 41 North American species,” *Biomass and Bioenergy*, св. 25, бр. 4, стр. 381–388, 2003.
- [108] D. Gianelle, W. Oechel, F. Miglietta, M. Rodeghiero и M. Sottocornola, “Cataloguing soil carbon stocks,” *Science*, св. 330, бр. 6010, стр. 1476–1476, 2010.
- [109] W. M. Post и K. C. Kwon, “Soil carbon sequestration and land-use change: processes and potential,” *Global change biology*, св. 6, бр. 3, стр. 317–327, 2000.
- [110] L. B. Guo и R. M. Gifford, “Soil carbon stocks and land use change: a meta analysis,” *Global change biology*, св. 8, бр. 4, стр. 345–360, 2002.
- [111] H. Eggleston, L. Buendia, K. Miwa, T. Ngara и K. Tanabe, “2006 IPCC guidelines for national greenhouse gas inventories,” 2006.

- [112] L. Mann, “Changes in soil carbon storage after cultivation,” *Soil Science*, св. 142, бр. 5, стр. 279–288, 1986.
- [113] S. M. Ogle, F. J. Breidt и К. Paustian, “Agricultural management impacts on soil organic carbon storage under moist and dry climatic conditions of temperate and tropical regions,” *Biogeochemistry*, св. 72, стр. 87–121, 2005.
- [114] Y. Pan, R. A. Birdsey, J. Fang и др., “A large and persistent carbon sink in the world’s forests,” *Science*, св. 333, бр. 6045, стр. 988–993, 2011.
- [115] E. G. Jobbágy и R. B. Jackson, “The vertical distribution of soil organic carbon and its relation to climate and vegetation,” *Ecological applications*, св. 10, бр. 2, стр. 423–436, 2000.
- [116] S. Liu, Y. Wei, W. M. Post, R. B. Cook, K. Schaefer и M. M. Thornton, “The Unified North American Soil Map and its implication on the soil organic carbon stock in North America,” *Biogeosciences*, св. 10, бр. 5, стр. 2915–2930, 2013.
- [117] USGCRP, Second state of the carbon cycle report (SOCCR2): A sustained assessment report, 2018.
- [118] S. Liu, J. Liu, Y. Wu и др., “Baseline and projected future carbon storage, carbon sequestration, and greenhouse-gas fluxes in terrestrial ecosystems of the eastern United States,” *Baseline and projected future carbon storage and greenhouse-gas fluxes in ecosystems of the eastern United States*. Reston, VA: US Department of the Interior, US Geological Survey Professional Paper, 1804.
- [119] N. H. Batjes, “Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks,” *Geoderma*, св. 269, стр. 61–68, 2016.
- [120] R. B. Harrison, P. W. Footen и B. D. Strahm, “Deep soil horizons: contribution and importance to soil carbon pools and in assessing whole-ecosystem response to management and global change,” *Forest Science*, св. 57, бр. 1, стр. 67–76, 2011.
- [121] P. S. Homann, B. T. Bormann и J. R. Boyle, “Detecting treatment differences in soil carbon and nitrogen resulting from forest manipulations,” *Soil Science Society of America Journal*, св. 65, бр. 2, стр. 463–469, 2001.
- [122] M. F. Jurgensen, D. S. Page-Dumroese, R. E. Brown и др., “Estimating carbon and nitrogen pools in a forest soil: influence of soil bulk density methods and rock content,” *Soil Science Society of America Journal*, св. 81, бр. 6, стр. 1689–1696, 2017.
- [123] P. Smith, J.-F. Soussana, D. Angers и др., “How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal,” *Global Change Biology*, св. 26, бр. 1, стр. 219–241, 2020.
- [124] C. Poeplau, “Measuring and modelling soil carbon stocks and stock changes in livestock production systems: guidelines for assessment; Version 1-Advanced copy,” 2019.
- [125] H. Throop, S. Archer, H. Monger и S. Waltman, “When bulk density methods matter: Implications for estimating soil organic carbon pools in rocky soils,” *Journal of Arid Environments*, св. 77, стр. 66–71, 2012.

- [126] C. Poeplau, C. Vos и A. Don, “Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content,” *Soil*, св. 3, бр. 1, стр. 61–66, 2017.
- [127] E. Vanguelova, E. Bonifacio, B. De Vos и др., “Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales—review and recommendations,” *Environmental monitoring and assessment*, св. 188, стр. 1–24, 2016.
- [128] D. Angers и N. Eriksen-Hamel, “Full-inversion tillage and organic carbon distribution in soil profiles: A meta-analysis,” *Soil Science Society of America Journal*, св. 72, бр. 5, стр. 1370–1374, 2008.
- [129] K. Paustian, J. Lehmann, S. Ogle, D. Reay, G. P. Robertson и P. Smith, “Climate-smart soils,” *Nature*, св. 532, бр. 7597, стр. 49–57, 2016.
- [130] D. of Primary Industries, R. D. Agriculture и G. o. W. A. Food, Measuring and reporting soil organic carbon, Accessed: 08.01.2024., 2024. адреса: <https://www.agric.wa.gov.au/soil-carbon/measuring-and-reporting-soil-organic-carbon>.
- [131] J.-F. Soussana, T. Tallec и V. Blanfort, “Mitigating the greenhouse gas balance of ruminant production systems through carbon sequestration in grasslands,” *Animal*, св. 4, бр. 3, стр. 334–350, 2010.
- [132] D. D. Baldocchi, “Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future,” *Global change biology*, св. 9, бр. 4, стр. 479–492, 2003.
- [133] M. Aubinet, T. Vesala и D. Papale, *Eddy covariance: a practical guide to measurement and data analysis*. Springer Science & Business Media, 2012.
- [134] C. Rebmann, M. Aubinet, H. Schmid и др., “ICOS eddy covariance flux-station site setup: a review,” *International Agrophysics*, св. 32, бр. 4, стр. 471–494, 2018.
- [135] G. Fratini и M. Mauder, “Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3,” *Atmospheric Measurement Techniques*, св. 7, бр. 7, стр. 2273–2281, 2014.
- [136] D. Franz, M. Acosta, N. Altimir и др., “Towards long-term standardised carbon and greenhouse gas observations for monitoring Europe’s terrestrial ecosystems: a review,” *International Agrophysics*, св. 32, бр. 4, стр. 439–455, 2018.
- [137] D. Baldocchi, H. Chu и M. Reichstein, “Inter-annual variability of net and gross ecosystem carbon fluxes: A review,” *Agricultural and Forest Meteorology*, св. 249, стр. 520–533, 2018.
- [138] E. S. I. S. B. -. E. S. Bhd., *Eddy Covariance*, Accessed: 08.01.2024., 2019. адреса: <https://esi.com.my/inst-home/eddy-covariance/>.
- [139] C. A. Gotway, R. B. Ferguson, G. W. Hergert и T. A. Peterson, “Comparison of kriging and inverse-distance methods for mapping soil parameters,” *Soil Science Society of America Journal*, св. 60, бр. 4, стр. 1237–1247, 1996.

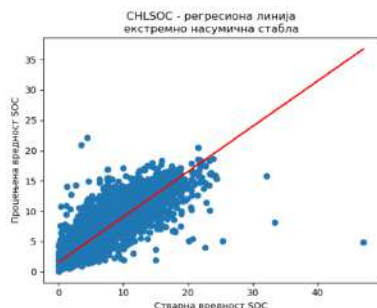
- [140] J. Triantafyllis, A. Huckel и I. Odeh, “Comparison of statistical prediction methods for estimating field-scale clay content using different combinations of ancillary variables,” *Soil Science*, св. 166, бр. 6, стр. 415–427, 2001.
- [141] A. B. McBratney, I. O. Odeh, T. F. Bishop, M. S. Dunbar и Т. М. Shatar, “An overview of pedometric techniques for use in soil survey,” *Geoderma*, св. 97, бр. 3-4, стр. 293–327, 2000.
- [142] P. Goovaerts и др., *Geostatistics for natural resources evaluation*. Oxford University Press on Demand, 1997.
- [143] V. Bellon-Maurel и A. McBratney, “Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils—Critical review and research perspectives,” *Soil Biology and Biochemistry*, св. 43, бр. 7, стр. 1398–1410, 2011.
- [144] K. D. Shepherd и M. G. Walsh, “Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries,” *Journal of Near Infrared Spectroscopy*, св. 15, бр. 1, стр. 1–19, 2007.
- [145] R. V. Rossel, T. Behrens, E. Ben-Dor и др., “A global spectral library to characterize the world’s soil,” *Earth-Science Reviews*, св. 155, стр. 198–230, 2016.
- [146] L. Wielopolski, R. Izaurrealde, C. Rice и др., “Evaluation of Three Field-Based Methods for Quantifying Soil Carbon,” *PLoS ONE*, св. 8, бр. BNL-102606-2013-JA, 2013.
- [147] J. L. (Optix), *Spectroscopy*, Accessed: 08.01.2024., 2020. адреса: <https://www.zeiss.com/spectroscopy/products/spectrometer-systems/aura.html>.
- [148] A. Stevens, B. van Wesemael, H. Bartholomeus, D. Rosillon, B. Tychon и E. Ben-Dor, “Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils,” *Geoderma*, св. 144, бр. 1-2, стр. 395–404, 2008.
- [149] V. Mulder, S. De Bruin, M. E. Schaepman и T. Mayr, “The use of remote sensing in soil and terrain mapping—A review,” *Geoderma*, св. 162, бр. 1-2, стр. 1–19, 2011.
- [150] K. Usha и B. Singh, “Potential applications of remote sensing in horticulture—A review,” *Scientia horticultrae*, св. 153, стр. 71–83, 2013.
- [151] A. Stevens, T. Udelhoven, A. Denis и др., “Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy,” *Geoderma*, св. 158, бр. 1-2, стр. 32–45, 2010.
- [152] C. Hbirkou, S. Pätzold, A.-K. Mahlein и G. Welp, “Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale,” *Geoderma*, св. 175, стр. 21–28, 2012.
- [153] H. Bartholomeus, L. Kooistra, A. Stevens и др., “Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy,” *International Journal of Applied Earth Observation and Geoinformation*, св. 13, бр. 1, стр. 81–88, 2011.
- [154] K. Whitehead и C. H. Hugenholtz, “Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges,” *Journal of Unmanned Vehicle Systems*, св. 2, бр. 3, стр. 69–85, 2014.

- [155] C. Zhang и J. M. Kovacs, “The application of small unmanned aerial systems for precision agriculture: a review,” *Precision agriculture*, св. 13, стр. 693–712, 2012.
- [156] A. C. Watts, V. G. Ambrosia и E. A. Hinkley, “Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use,” *Remote sensing*, св. 4, бр. 6, стр. 1671–1692, 2012.
- [157] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis и D. Bochtis, “Remote sensing techniques for soil organic carbon estimation: A review,” *Remote Sensing*, св. 11, бр. 6, стр. 676, 2019.
- [158] B. Frazier и Y. Cheng, “Remote sensing of soils in the eastern Palouse region with Landsat Thematic Mapper,” *Remote Sensing of Environment*, св. 28, стр. 317–325, 1989.
- [159] F. Castaldi, R. Casa, A. Castrignanò, S. Pascucci, A. Palombo и S. Pignatti, “Estimation of soil properties at the field scale from satellite data: a comparison between spatial and non-spatial techniques,” *European Journal of Soil Science*, св. 65, бр. 6, стр. 842–851, 2014.
- [160] B. Wang, C. Waters, S. Orgill и др., “High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia,” *Science of the Total Environment*, св. 630, стр. 367–378, 2018.
- [161] J. A. M. Demattê, C. T. Fongaro, R. Rizzo и J. L. Safanelli, “Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images,” *Remote Sensing of Environment*, св. 212, стр. 161–175, 2018.
- [162] A. Gholizadeh, D. Žižala, M. Saberioon и L. Borvka, “Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging,” *Remote Sensing of Environment*, св. 218, стр. 89–103, 2018.
- [163] Eurostat. “Sampling system for LUCAS 2015.” (2015.), адреса: https://ec.europa.eu/eurostat/documents/205002/6786255/LUCAS+2015+sampling_20160922.pdf (посећено 26. 8. 2023.).
- [164] M. Pfeiffer, J. Padarian, R. Osorio и др., “CHLSOC: the Chilean Soil Organic Carbon database, a multi-institutional collaborative effort,” *Earth System Science Data*, св. 12, бр. 1, стр. 457–468, 2020.
- [165] Copernicus, LandCopernicus, 05 15, 2023, 2023. адреса: <https://land.copernicus.eu/user-corner/technical-library/>.
- [166] A. Dabija, M. Kluczek, B. Zagajewski и др., “Comparison of support vector machines and random forests for corine land cover mapping,” *Remote Sensing*, св. 13, бр. 4, стр. 777, 2021.
- [167] D. García-Álvarez и M. T. C. Olmedo, “Analyzing the uncertainty of the CORINE Land Cover time series (1990-2018) for Spain,” 2022.
- [168] O. G. Varga, Z. Kovács, L. Bekő и др., “Validation of visually interpreted corine land cover classes with spectral values of satellite images and machine learning,” *Remote Sensing*, св. 13, бр. 5, стр. 857, 2021.

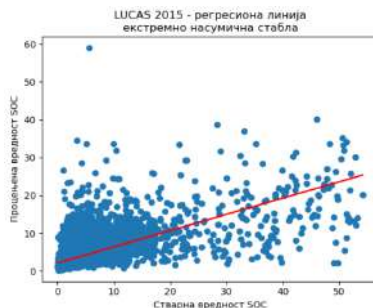
- [169] eo-learn, 2023. адреса: <https://github.com/sentinel-hub/eo-learn/>.
- [170] F. Pedregosa, G. Varoquaux, A. Gramfort и др., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, св. 12, стр. 2825–2830, 2011.

11 ПРИЛОЗИ

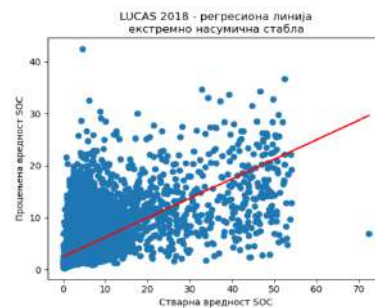
Прилог 1: Графикони регресионих линија за базични модел 04



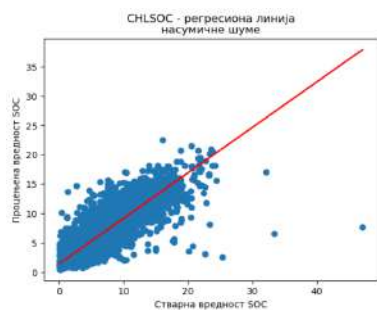
(А) CHLSOC - екстремно насумична стабла



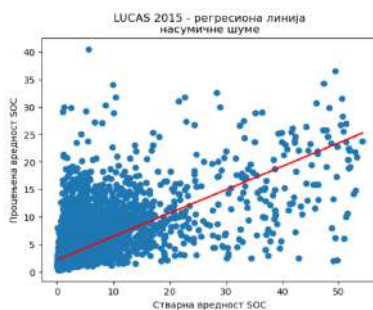
(Б) LUCAS 2015 - екстремно насумична стабла



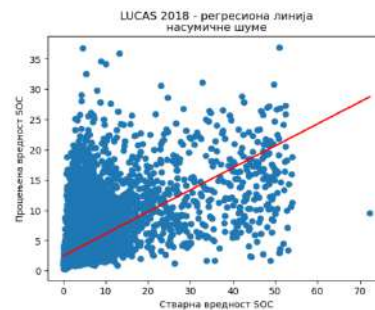
(В) LUCAS 2018 - екстремно насумична стабла



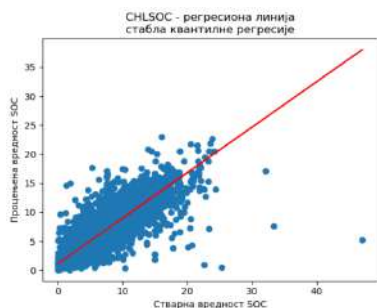
(Г) CHLSOC - насумичне шуме



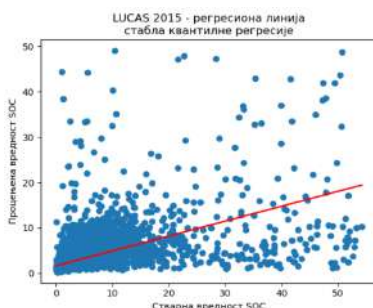
(Д) LUCAS 2015 - насумичне шуме



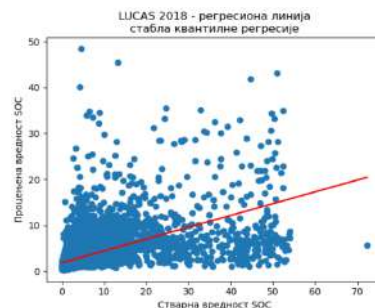
(Ђ) LUCAS 2018 - насумичне шуме



(Е) CHLSOC - стабла квантилне регресије

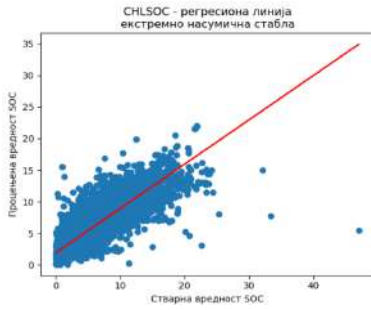


(Ж) LUCAS 2015 - стабла квантилне регресије

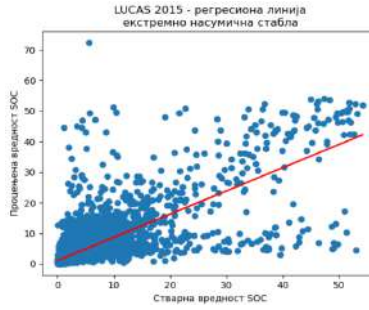


(З) LUCAS 2018 - стабла квантилне регресије

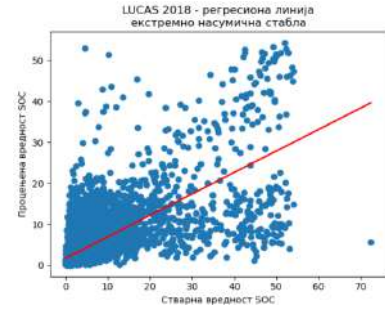
Прилог 2: Графициони регресионих линија за хибридни модел 04



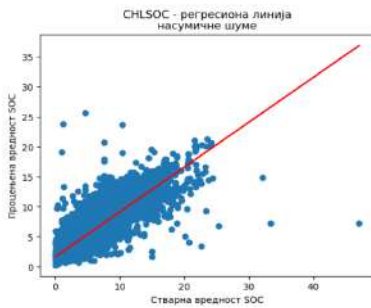
(А) CHLSOC - екстремно насумична стабла



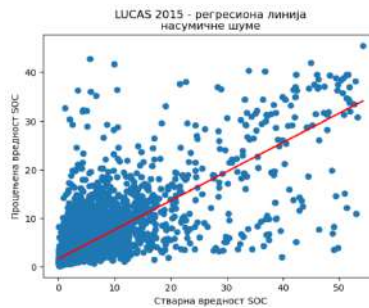
(Б) LUCAS 2015 - екстремно насумична стабла



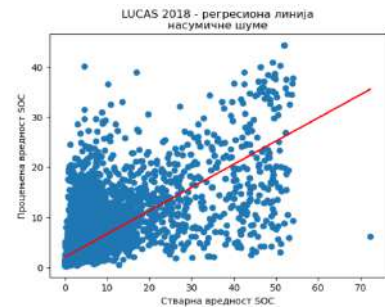
(В) LUCAS 2018 - екстремно насумична стабла



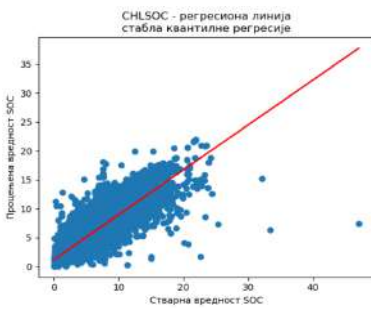
(Г) CHLSOC - насумичне шуме



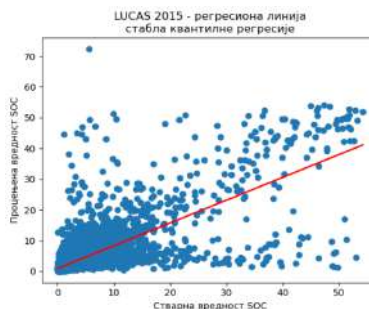
(Д) LUCAS 2015 - насумичне шуме



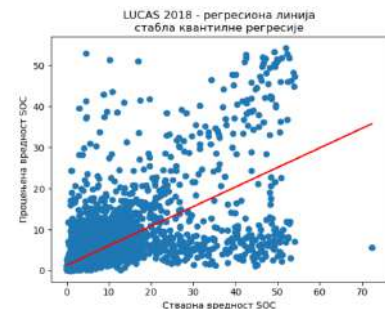
(Ђ) LUCAS 2018 - насумичне шуме



(Е) CHLSOC - стабла квантилне регресије



(Ж) LUCAS 2015 - стабла квантилне регресије



(З) LUCAS 2018 - стабла квантилне регресије

План третмана података

| |
|--|
| Назив пројекта/истраживања |
| Даљинска детекција садржаја органског угљеника у земљишту на бази дубоког учења |
| Назив институције/институција у оквиру којих се спроводи истраживање |
| а) Факултет техничких наука, Универзитет у Новом Саду б) в) |
| Назив програма у оквиру ког се реализује истраживање |
| |
| 1. Опис података |
| <p>1.1 Врста студије</p> <p><i>Укратко описати тип студије у оквиру које се подаци прикупљају</i></p> <p>__ У овој студији нису прикупљани подаци _____</p> <p>_____</p> <p>_____</p> |
| <p>1.2 Врсте података</p> <p>а) квантитативни б) квалитативни</p> |
| <p>1.3. Начин прикупљања података</p> <p>а) анкете, упитници, тестови б) клиничке процене, медицински записи, електронски здравствени записи в) генотипови: навести врсту _____ г) административни подаци: навести врсту _____ д) узорци ткива: навести врсту _____ ђ) снимци, фотографије: навести врсту _____</p> |

е) текст, навести врсту _____

ж) мапа, навести врсту _____

з) остало: описати _____

1.3 Формат података, употребљене скале, количина података

1.3.1 Употребљени софтвер и формат датотеке:

а) Excel фајл, датотека _____

б) SPSS фајл, датотека _____

с) PDF фајл, датотека _____

д) Текст фајл, датотека _____

е) JPG фајл, датотека _____

ф) Остало, датотека _____

1.3.2. Број записа (код квантитативних података)

а) број варијабли _____

б) број мерења (испитаника, процена, снимака и сл.) _____

1.3.3. Поновљена мерења

а) да

б) не

Уколико је одговор да, одговорити на следећа питања:

а) временски размак између поновљених мера је _____

б) варијабле које се више пута мере односе се на _____

в) нове верзије фајлова који садрже поновљена мерења су именоване као _____

Напомене: _____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?

а) Да

б) Не

Ако је одговор не, образложити _____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?

а) експеримент, навести тип _____

б) корелационо истраживање, навести тип _____

ц) анализа текста, навести тип _____

д) остало, навести шта _____

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

2.2 Квалитет података и стандарди

2.2.1. Третман недостајућих података

а) Да ли матрица садржи недостајуће податке? Да Не

Ако је одговор да, одговорити на следећа питања:

а) Колики је број недостајућих података? _____

б) Да ли се кориснику матрице препоручује замена недостајућих података? Да Не

в) Ако је одговор да, навести сугестије за третман замене недостајућих података

2.2.2. На који начин је контролисан квалитет података? Описати

2.2.3. На који начин је извршена контрола уноса података у матрицу?

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у _____ репозиторијум.

3.1.2. URL адреса _____

3.1.3. DOI _____

3.1.4. Да ли ће подаци бити у отвореном приступу?

- a) Да
- б) Да, али после ембарга који ће трајати до _____
- в) Не

Ако је одговор не, навести разлог _____

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани.

Образложење

3.2 Метаподаци и документација података

3.2.1. Који стандард за метаподатке ће бити примењен? _____

3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

3.3 Стратегија и стандарди за чување података

3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? _____

3.3.2. Да ли ће подаци бити депоновани под шифром? Да Не

3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да Не

3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?

Да Не

Образложити

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.

4.1 Формални стандарди за сигурност информација/података

Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности (https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html) и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Да Не

Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да Не

Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:

- a) Подаци нису у отвореном приступу
- б) Подаци су анонимизирани
- ц) Остало, навести шта

5. Доступност података

5.1. Подаци ће бити

- a) јавно доступни
- б) доступни само уском кругу истраживача у одређеној научној области
- ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима
