

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Tarkvarateaduse instituut

Priit Ulmas 178001IABM

**SATELLIIDIFOTODE SEGMENTEERIMINE U-NET
MUDELIGA MAAKASUTUSE KAARDISTAMISE NÄITEL**

Magistritöö

Juhendaja

Innar Liiv

PhD

Tallinn 2020

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Priit Ulmas

.....

(allkiri)

Kuupäev: Jaanuar 6, 2020

Annotatsioon

Käesoleva magistritöö teemaks on U-Net arhitektuuril põhineva konvolutsioonilise masinõppe mudeli loomine maakasutuse kaardistamise näitel satelliidifotodelt. Töö eesmärgiks on konvolutsiooniliste masinõppe mudelite treenimise ja analüüsi läbi hinnata masinõppe abil maakasutuse kaartide loomise teostatavust ning vaadelda võimalusi olemasoleva kaardi resolutsiooni tõstmiseks ja keskkonnamuutuste jälgimiseks.

Probleemi lahendamiseks töö autor koostas ja treenis masinõppe mudelid satelliidifotode klassifitseerimiseks ja piksli tasemel segmenteerimiseks maakasutuse klassifikatsioonide põhiselt. Töö käigus analüüsiti loodud mudelite täpsust klassifikatsiooni kolme erineva detailsusastme juures.

Satelliidifotode andmekogudest kasutas autor 2019. aastal avaldatud ja töö kirjutamise hetkeks vähe kasutatud BigEarthNet andmestikku. See põhineb Sentinel-2 satelliidifotodel, mis on pildistatud 2017. ja 2018. aasta jooksul ja hõlmab piirkondi kümnest Euroopa riigist. Lisaks kasutati töös CORINE maakasutuskaarti aastast 2018 ja sama perioodi satelliidifotot Eestist.

Töö käigus loodud mudel saavutas võrdlemisi kõrge F_2 skoori 0.793 mitmeklassilise klassifikatsiooni puhul 43 eri klassiga. Mudel ühtlasi tõi esile puuduseid BigEarthNet andmekogu täpsuses, kus piltide pealkirjad ei ole korrektsed.

Segmenteerimismudelid loodi Sentinel-2 satelliidifoto piksli tasemel klassifitseerimiseks maakasutuse klasside täpse ala tuvastamiseks. Mudelid näitasid kõrgeimat IoU täpsust eelkõige metsade (0.867), veekogude (0.813) ja põllumaade (0.757) tuvastamisel. Mudel näitas ühtlasi ka kasutusvõimalusi olemasoleva maakasutuskaardi piirjoonte korrigeerimisel ja maakasutuse muutuste jälgimisel.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 48 leheküljel, 4 peatükki, 30 joonist, 8 tabelit ja 4 lisa.

Abstract

Segmentation of Satellite Imagery using U-Net Models: a Case Study of Land Use Classification

The research topic of this master's thesis is developing a convolutional machine learning model with a U-Net structure for creating land use classification (LUC) mapping based on satellite imagery. The aim of the research is to develop and test convolutional models for automatic land use mapping and to assess their usability in increasing land cover mapping accuracy and change detection.

To solve these tasks the author comprised and trained machine learning models for land use classification and semantic segmentation from satellite images. The results were analysed on three different land classification levels.

BigEarthNet satellite image archive was selected for the research as one of two main datasets. This dataset was published in 2019 and includes Sentinel-2 satellite photos from 10 European countries made in 2017 and 2018. As a second dataset the author comprised a set containing a Sentinel-2 image and a land use map of Estonia.

The developed classification model shows a high overall F_2 score of 0.793 on multiclass land use classification with 43 possible image labels. The model also highlights noisy data in the BigEarthNet dataset, where images seem to have incorrect labels.

Segmentation models offer a solution for generating automatic land use mappings based on Sentinel-2 satellite images and show a relatively high IoU score mainly for forests (0.867), water bodies (0.813) and arable land (0.757). The models show a capability of increasing the accuracy of existing land classification maps and in land use change detection.

The thesis is written in Estonian and contains 48 pages of text, 4 chapters, 30 figures, 8 tables and 4 appendices.

Lühendite ja mõistete sõnastik

BigEarthNet	Andmekogu satelliidipildidel põhinevate masinõppe mudelite loomiseks
CLC	<i>CORINE Land Cover</i> , CORINE maakasutuse klassifitseerimise projekt, mille tulemusena on loodud Euroopa maakasutuskaart
CNN	<i>Convolutional Neural Network</i> , konvolutsiooniline tehismärgivõrk
CORINE	<i>Coordination of information on the environment</i> , Euroopa keskkonnainfo kogumiseks loodud programm
Eksimismatriks	<i>Confusion matrix</i> , matriks, milles registreeritakse katseliste näidetele mingi reeglistiku rakendamisel saadavate õigete ja väärade liigitusjuhtude arv [EVS-ISO/IEC 2382-31:1999]
FN	<i>False Negative</i> , valenegatiiv
FP	<i>False Positive</i> , valepositiiv
IoU	<i>Intersection over Union (Jaccard Index)</i> , kahte andmekogu võrdlev statistiline arvutus, ühisosa ja ühendi jagatis
LUC	<i>Land Use Classification</i> , maakasutuse klassifitseerimine
LULUC	<i>Land use and land-use changes</i> , maakasutuse klassifitseerimine ja selle muutuste jälgimine
NN	<i>Neural Network</i> , Tehismärgivõrk
ResNet	Konvolutsioonilise tehismärgivõrgu arhitektuur, mis võimaldab luua suure arvu kihtidega mudeleid
Sentinel-2	Kaugseiresatelliitide paar, mis pildistavad maapinda kõrglahutusega kaameratelega 5 päevase intervalliga
Siirdeõpe	<i>Transfer learning</i> , masinõppe meetoodika, milles kasutatakse teise valdkonna mudeli kaalusid uue valdkonna mudeli treenimise alguspunktina
TP	<i>True Positive</i> , tõeselt positiivne
U-Net	Konvolutsioonilise tehismärgivõrgu arhitektuur, mida rakendatakse pildi segmenteerimise ülesannetel

Sisukord

Jooniste loetelu	vii
Tabelite loetelu	viii
Sissejuhatus	1
1 Teoreetiline taust	3
1.1 Säästva arengu näitajad	3
1.1.1 Eesmärk 15, Maismaa ökosüsteemid	3
1.2 Kosmosesektori areng	3
1.3 Maapinna vaatlus	4
1.3.1 CORINE maakasutuse klassifikatsioon	5
1.4 Ülevaade tehnoloogiast	7
1.4.1 Juhendatud, juhendamata ja kinnitusega masinõpe	7
1.4.2 Konvolutsioonilised närvivõrgud	7
1.4.3 ResNet mudel	8
1.4.4 U-Net mudel	8
1.4.5 Siirdeõpe	9
1.4.6 Semantiline segmenteerimine	10
2 Metoodika	11
2.1 Andmekogud	11
2.1.1 BigEarthNet andmestik	11
2.1.2 Sentinel-2 andmed	13
2.1.3 CORINE maakasutuse andmed	13
2.2 Andmete eeltöötlus	13
2.3 Mudelid	17
2.3.1 Masinõppe treenimiskeskond	17
2.3.2 Kasutatavad masinõppe mudelid	18
3 Eksperimendid	20
3.1 Eksperimendi disain	20
3.2 Eksperimendi tulemuste mõõtmine	20
3.2.1 Andmekogude hindamine	20
3.2.2 Klassifitseerimismudeli hindamine	21
3.2.3 Segmenteerimismudeli hindamine	22

4 Tulemused	24
4.1 Andmekogude näidikud	24
4.2 Klassifitseerimismudeli tulemused	24
4.2.1 Maakasutuse klassifikatsiooni 1. tase	24
4.2.2 Maakasutuse klassifikatsiooni 2. tase	27
4.2.3 Maakasutuse klassifikatsiooni 3. tase	28
4.3 Segmenteerimismudeli tulemused	30
4.3.1 Maakasutuse klassifikatsiooni 1. tase	30
4.3.2 Maakasutuse klassifikatsiooni 2. tase	32
4.3.3 Maakasutuse klassifikatsiooni 3. tase	35
4.4 Tulemuste analüüs	38
4.5 Teemad edasiseks uurimistööks	41
Kokkuvõte	42
Kasutatud kirjandus	44
Summary	47
Lisa 1 - BigEarthNet klassid ja piltide arv	49
Lisa 2 - Klassifikatsioonimudelite eksimismatriksid (2. ja 3. tase)	51
Lisa 3 - Klassifikatsioonimudeli näitajad klasside põhiselt, 3. tase	56
Lisa 4 - Näiteid saadud tulemustest	58

Jooniste loetelu

1	Riskikapitali investeeringud kosmosesektorisse, 2010-2019. [5]	4
2	CORINE maakasutuskaardi näide.	6
3	ResNet arhitektuuri plokk.	8
4	U-Net mudeli struktuuri näidis.	9
5	Semantilise segmenteerimise sisend ja väljund CamVid andmekogu näitel.	10
6	Näidis BigEarthNet andmekogust.	11
7	Andmekogu loomiseks kasutatav piirkond.	14
8	Näide segmenteerimismudeli treenimiseks loodud andmekogust.	15
9	Piksli tasemel klasside jaotus 1. tasemel.	16
10	Piksli tasemel klasside jaotus 2. tasemel.	16
11	Piksli tasemel klasside jaotus 3. tasemel.	17
12	One-Cycle meetod. Treenimise käigus muutuv õpisamm ja hoog.	18
13	Mudeli treenimise progress (ResNet-50), klassifikatsiooni 1. tase.	25
14	Mudeli absoluutne täpsus 1. taseme klassifikatsiooni määramisel.	25
15	Klasside põhised eksimismatriksid 1. tasemel.	26
16	Mudeli treenimise progress (ResNet-50), klassifikatsiooni 2. tase.	27
17	Mudeli absoluutne täpsus 2. taseme klassifikatsiooni määramisel.	27
18	Mudeli treenimise progress (ResNet-50), klassifikatsiooni 3. tase.	29
19	Mudeli absoluutne täpsus 3. taseme klassifikatsiooni määramisel.	29
20	Mudeli treenimise progress (U-Net), klassifikatsiooni 1. tase.	30
21	Segmenteerimismudeli tulemuste näide, maakasutuse 1. tase.	31
22	Segmenteerimise eksimismatriks maakasutuse 1. taseme klasside lõikes.	32
23	Mudeli treenimise progress (U-Net), klassifikatsiooni 2. tase.	33
24	Segmenteerimismudeli tulemuste näide, maakasutuse 2. tase.	34
25	Segmenteerimise eksimismatriks maakasutuse 2. taseme klasside lõikes.	34
26	Mudeli treenimise progress (U-Net), klassifikatsiooni 3. tase.	35
27	Segmenteerimismudeli tulemuste näide, maakasutuse 3. tase.	37
28	Segmenteerimise eksimismatriks maakasutuse 3. taseme klasside lõikes.	37
29	Ebatäpsused BigEarthNet andmekogus.	39
30	Ebatäpsused maakasutuskaardil.	40

Tabelite loetelu

1	Sentinel-2 spektraalkanalid.	5
2	Maakatte klassifikatsioon	12
3	BigEarthNet andmete kardinaalsus ja tihedus.	24
4	Klassifikatsioonimudeli hindamine 1. tasemel.	26
5	Klassifikatsioonimudeli hindamine 2. tasemel.	28
6	Segmenteerimise täpsus maakasutuse 1. taseme klasside lõikes.	30
7	Segmenteerimise täpsus maakasutuse 2. taseme klasside lõikes.	33
8	Segmenteerimise täpsus maakasutuse 3. taseme klasside lõikes.	35

Sissejuhatus

Käesoleva töö teemaks on masinõppe meetoditel põhineva algoritmi loomine satelliidifotode töötlemisel maapinna klassifitseerimiseks. Töös luuakse masinõppe mudelid satelliidifotode klassifitseerimiseks ja piksli tasemel segmenteerimiseks maakasutuse klassifikatsiooni alusel ning analüüsitakse saadud tulemusi maakasutukaartide täpsuse tõstmisel ja keskkonnamuutuste jälgimisel.

Taust

Maapinna vaatlus on aktiivne uurimisvaldkond, mis omab suurt tähtsust looduskeskkonna säilimise ja keskkonna planeerimise vallas. Ka ÜRO säästva arengu eesmärkide saavutamise aastaks 2030 toetub maapinna ja selle erinevate vormide, näiteks metsade ja veekogude muutuste jälgimisel.

Kosmosesektor on alates 2015. aastast tugevalt kasvatanud riskikapitali investeeringud, ületades prognoositava aastase kogusummana 2019. aastal 2 miljardi dollari piiri. Satelliidivõrkude kiire kasv ja andmemahu suurenemine loob palju uusi võimalusi andmetöötlemises ja sellest uue väärtuse leidmisel. Kui varasemalt on satelliidiandmete töötlemine nõudnud suurel määral manuaalset tööd, siis tänu masinõppe arengule on võimalik seda üha rohkem automatiseerida, luues vahetu infovoo maapinnal toimivate muutuste jälgimiseks.

Masinõppe kasutamine andmetöötlemises on kiiresti tõusvas trendis tänu suurtele arenguhüpetele arvutusvõimsuses ja masinõppe meetodites, eriti pildituvastuse ja keeletöötlemise vallas. See võimaldab automatiseerida mitmeid seni manuaalseid andmetöötlemise protsesse ja ühtlasi loob hulgaliselt nii ärilisi, sotsiaalseid, kui ka keskkonda säästvaid lahendusi.

Probleem

Maakasutuse jälgimine on töömahukas tegevus, mida näiteks Euroopas on läbi viidud iga 6 aasta tagant ja mille protsess aastaks 2018 võttis aega 1,5 aastat. Teiseks probleemiks on loodavate maakasutuskaartide võrdlemise suur ebatäpsus. Kui kaasaegne satelliidifoto võimaldab pilti maapinnast 10m täpsusega, siis maakasutuskaart on 100m täpsusega ja ei võta arvesse alla 25 hektarilisi alasid. Täpne ja pidevalt uuenev maakasutuskaart annaks

enneolematu võimaluse jälgida muutuseid meid ümbritsevas keskkonnas ja võtta vastu paremaid otsuseid keskkonda planeerides.

Eesmärk

Töö eesmärgiks on luua U-Net arhitektuuril põhinev masinõppe mudel maakasutuskaartide loomiseks satelliidifotode põhjal. Töö käigus hinnatakse loodud mudelite täpsust ja võrreldakse seda üleeuroopalise maakasutuskaardiga. Lisaks vaadeldakse loodud mudeli rakendamise võimalust olemasolevate maakasutuskaartide täpsuse tõstmisel ja maakasutuse muutuste jälgimisel.

Masinõppe mudelite treenimisel kasutatakse BigEarthNet andmekogu, Sentinel-2 satelliidifotot ja CORINE maakasutuskaarti. Lisaks olemasolevale satelliidifotode andmekogule luuakse lisaks teine andmekogu segmenteerimismudeli treenimiseks.

Ülevaade tööst

Töö esimene peatükk käsitleb teema tausta ja aktuaalsust ning toob välja peamised töös rakendatud masinõppega seotud mõisted. Ühtlasi selgitatakse töös kasutatud masinõppe mudelite struktuuri.

Teine peatükk kirjeldab töö metoodilist sisu. Peatükk sisaldab ülevaadet töös kasutatud andmekogudest, rakendatud masinõppe mudelitest ja andmete eeltöötluse protsessist.

Kolmandas peatükis vaadeldakse töös teostatud eksperimentide ülesehitust ja selle tulemuste hindamise kriteeriume.

Neljas peatükk toob välja töös saadud tulemused ja nende analüüsi.

Magistritöö viimane peatükk toob kokkuvõtlikult välja töö peamised tulemused ja järeldused.

1. Teoreetiline taust

Käesoleva peatüki esimene pool annab ülevaate satelliitidel põhineva maapinnavaatluse arengust ja trendidest. Peatüki teine pool annab ülevaate magistritöös kasutatud masinõppe meetoditest ja süvaõppe peamistest mõistetest.

1.1 Säästva arengu näitajad

Aastal 2015 võeti ÜRO Säästva arengu tippkohtumisel liikmesriikide poolt vastu säästva arengu eesmärgid, millega defineeriti kokku 17 eesmärki ja 169 alaeesmärki inimeste ja planeedi kaitseks ning heaolu saavutamiseks. Eesmärkide saavutamise tähtajaks määrati hiljemalt aasta 2030, mitmete alaeesmärkide puhul ka juba 2020. aasta. Iga alaeesmärgi juurde on ühtlasi määratud ka kindlad mõõdetavad väärtused, mille abil eesmärkide täitumist analüüsida [1]. Lisaks eesmärkide seadmisele rõhutab deklaratsioon ka ressursside panustamise andmete kogumises, analüüsimises ja jagamises. [2]

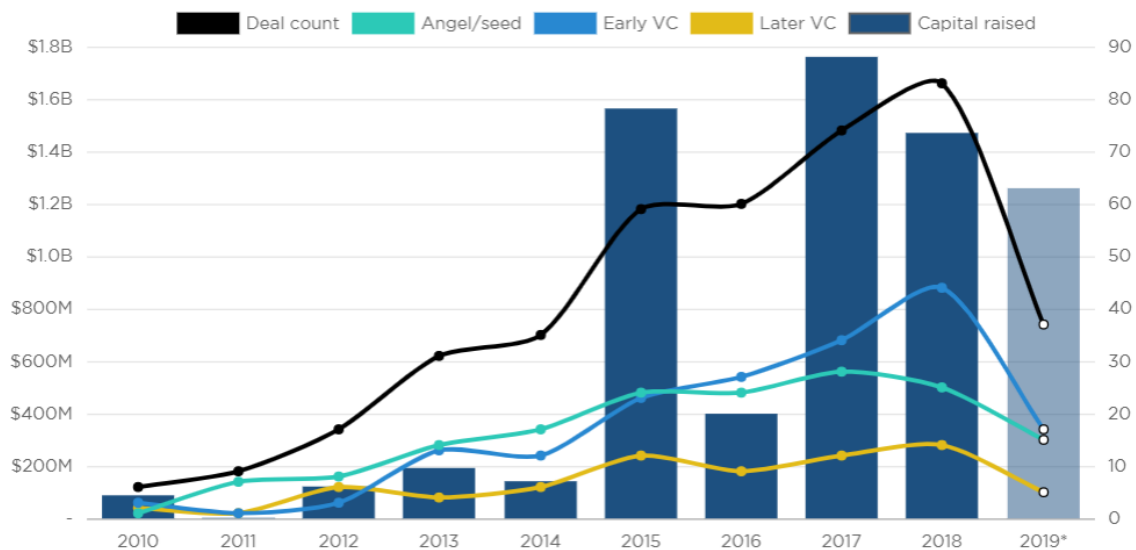
1.1.1 Eesmärk 15, Maismaa ökosüsteemid

Säästva arengu eesmärk nr. 15 kirjeldab maismaa ökosüsteemi jätkusuutlikkuse tagamise sihte ja nende jälgimise näitajaid. Eesmärgi alla on koondatud 12 alaeesmärki, mille sisu katab nii liikide säilimise tagamist kui ka metsade säästvat majandamist ja kõrbestumise peatamist [2]. Osadeks arvulisteks mõõdikuteks selle eesmärgi all on näiteks metsa-alade protsent kogu maismaast ja metsaalade muutuse protsent [1]. Säästva arengu näitajate jälgimisel on suure rolli endale võtnud ka Euroopa Liidu maa seire ja jälgimise programm Copernicus, mis jälgib ja koostab selleks olulisi andmekogusid ja ühtlasi koostab ka käesolevas töös kasutatavat CORINE maakasutuskaarti. [3]

1.2 Kosmosesektori areng

Euroopa Komisjoni ja Euroopa Investeerimispannga raport kosmosesektori arengu ja investeeringute üle Euroopas näitab sektori aktiivset tõusu ja investeerimismahtude hoogsat kasvu. Perioodide 2007-2012 ja 2012-2017 vahel 347% suurust kasvu näitav investeeringute maht tuleneb suures osas riskikapitali osakaalu suurenemisest. Investeeringute kogumaht nendel perioodidel on vastavalt 1.9 ja 8.5 miljardid eurot. Raport

toob ühtlasi ka välja, et suurim osa (49%) 2017. aastal tehtud investeeringutest komponentide ehituse ja kosmosetranspordi peale tehti maapinna vaatluse valdkonnas, millele järgnes kommunikatsiooni valdkond (18%). [4]



Joonis 1. Riskikapitali investeeringud kosmosesektorisse, 2010-2019. [5]

Viimase kümne aasta riskikapitali investeeringud kosmosesektorisse (Joonis 1) näitavad kiiret kasvu sõlmitud investeerimislepingute arvus ja investeeringute kogusummas. 2019. aasta investeeringud näitavad esimese kahe kvartali andmete põhjal esmakordselt ligi 2 miljardi dollari suurust koguinvesteeringut ühe aasta jooksul. [5]

1.3 Maapinna vaatlus

Satelliidiandmete kasutamine maa jälgimisel on tõusvas trendis ja üha rohkem rakendatakse selleks ka era satelliidivõrke operatiivselt info kogumiseks. Peamiselt tehnoloogilised edasiarendused on muutud ligipääsu kosmosesse odavamaks ja selle hind on jätkuvalt langustrendis. Seda toetab näiteks korduvkasutatavate rakettide arendamine ja tugev konkurents sellel turul. Odavam hind satelliidi üles viimisel on võimaldanud ka era satelliidivõrkude loomise. [6]

Mitmed satelliidivõrgud pakuvad kõrge resolutsiooniga ja tiheda ülelennuga maapinna jälgimise teenuseid. See võimaldab saada uut infot isegi mitmel korral päeva jooksul, mis omakorda on avanud uued ärimudelid satelliidiandmete töötlemise ja selle põhjal väärtuse pakkumisel. [7]

Avaliku sektori satelliidiprogrammidest on üks kaasaegsemaid ja võimekamaid Sentinel-2,

mille kaks satelliiti suudavad pakkuda maapinna vaatluse andmeid 13 spektraalkanaliga kaamerate abil viie päevase intervalliga [8]. Tabelis 1 on välja toodud Sentinel-2 satelliitide infokanalid [9]. Samuti on tabelis näha ka erinevate kanalite resolutsioonid, millest ühed täpsemad on traditsiooniliseks RGB värvifotoks kokku liidetavad punane, roheline ja sinine kanal, mida kasutatakse käesolevas töös.

Tabel 1. Sentinel-2 spektraalkanalid.

Nr	Kanali nimetus	Resolutsioon
1	Aerosool	60
2	Sinine	10
3	Roheline	10
4	Punane	10
5	Taimkate, punane serv	20
6	Taimkate, punane serv	20
7	Taimkate, punane serv	20
8	Lähis-infrapunane	10
8a	Lähis-infrapunane	20
9	Veeaur	60
10	Lühilaineline infrapunane - kiudpilved	60
11	Lühilaineline infrapunane	20
12	Lühilaineline infrapunane	20

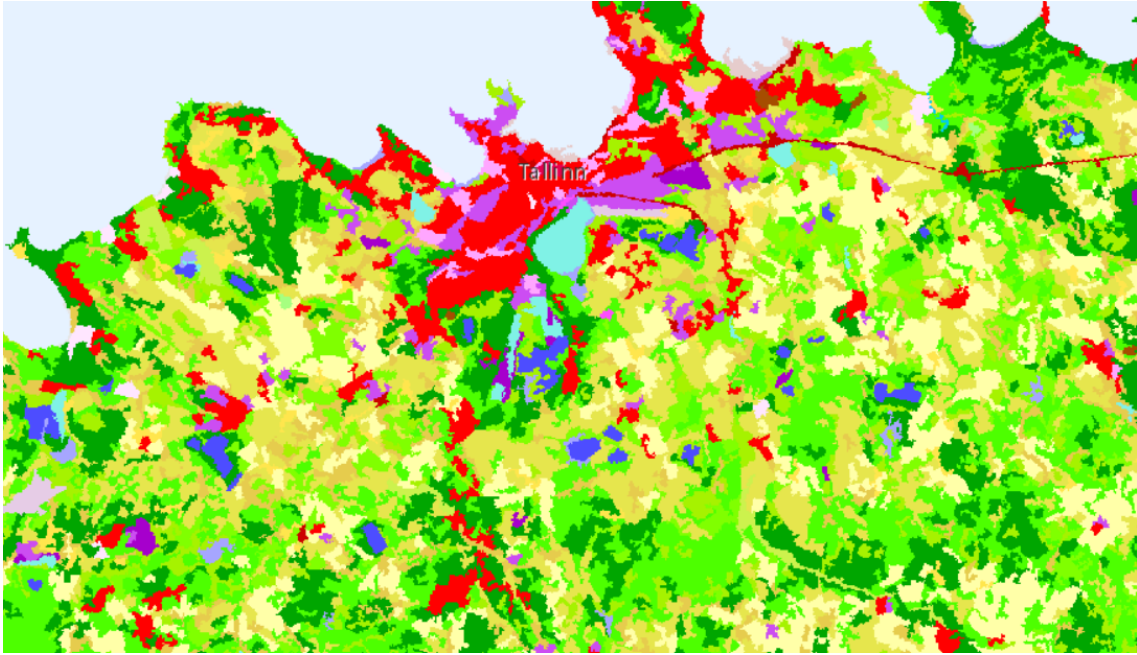
Lisaks keskkonna jälgimisele on aktiivsed arengusuunad ka näiteks looduskatastroofide ulatuste ja keskkonnariskide hindamine, ehitustegevuse jälgimine ning äriliste konkurentide tegevuse monitooring. [10]

1.3.1 CORINE maakasutuse klassifikatsioon

CORINE maakatte kaardistamise (CLC - *Corine Land Cover*) esimene referentsaasta oli 1990. Hiljem on toimunud veel neli uuendust, millest kõige hilisem valmis 2018. aastal (kaardistamise aastad: 1990, 2000, 2006, 2012, 2018). Lisaks maakatte kaardi uuendamistele on loodud ka maakatte muutuste kaardistused, mis annavad ülevaate iga kahe kaardi vahelistest muutustest.[11]

CLC klassid on defineeritud selle ametlikus juhendis ja koosnevad kolmest hierarhilisest tasemest, vastavalt 5, 15 ja 43 maakasutuse klassiga. [12]

CORINE maakasutuskaardi eeliseks saab arvestada selle meetodilist püsivust, kuid peamiseks miinuseks on siiski võrdlemisi suur üldistus (tingitud kaartide madalast resolutsioonist) ja sellest tulenev ebatäpsus. Näiteks on maa-ala kaardile kandmise eelduseks



Joonis 2. CORINE maakasutuskaardi näide.

vähemalt 25 ha suurune pindala [12], mille tõttu on paljud väiksemad alad ebatäpselt klassifitseeritud. Veel on näiteks enamus Eesti aladel toimuvaid metsaraieid väiksemad kui 5 ha ja on seega kaardistamiseks liiga väikese pindalaga, mistõttu ei ole võimalik selle kaardi abil näiteks metsade muutuseid piisava täpsusega analüüsida [13]. CORINE maakasutuskaardi koostamise nõutavaks täpsuseks on 85% ja 2012. aasta kaardi peal läbi viidud analüüsi kohaselt saadi üldiseks täpsuseks 89.7% [14].

Maakasutuse kaardistamise protsessi automatiseerimiseks CORINE maakasutuskaardi järgi on Baldao (et al., 2018) välja pakkunud lasermõõdistamise andmete kasutamist [15]. Nende loodud lahendus rakendas konvolutsioonilisi süvaõppe mudeleid ja saavutas 73.5% täpsuse esimese taseme klasside põhised. 2019. aastal on välja töötatud ka juhendamata masinõppe meetoditel töötav lahendus (Paris, 2019), mida satelliidiandmete kaudu maakasutuskaartide uuendamiseks rakendati [16].

2019. aasta detsembrikuus avalikustas CORINE masinõppe abil loodud ülemaailmse maakasutuskaardi, mis töö kirjutamise ajal oli saadaval 2015. aasta kohta. Selle kaardi loomisel rakendati andmetöötluse meetodeid satelliidipiltide ajaseeria töötlemiseks ja otsustuspuudel põhinevaid masinõppe meetodeid korrektse maakasutuse klassifikatsiooni määramiseks. Loodud kaart on täpsusega 100m. [17]

Näidis CORINE maakasutuskaardist Joonisel 2 kujutab Tallinna ja selle lähiümbruse maakasutuse jaotust.

1.4 Ülevaade tehnoloogiast

1.4.1 Juhendatud, juhendamata ja kinnitusega masinõpe

Masinõpe aitab meil andmete põhjal leida vastuseid küsimustele, mille lahendamiseks konkreetne algoritm ei ole teada või puudub. Puuduseid meie teadmistes leevendab aga suur hulk andmeid, mille põhjal on võimalik luua arvutuslik mudel küsimuse lahendamiseks. See küll alati ei lahenda küsimust täieliku täpsusega, aga võib siiski anda piisava vastuse. Masinõppe meetodeid eristuvad üksteisest peamiselt sõltuvalt lahendatavast probleemist. [18]

Juhendatud õpe loob seoseid sisend- ja väljundandmete vahel. Selline probleem võib olla näiteks auto hinna määramine kui on teada ajaloolised andmed masina tehniliste näitajate ja müügihinna kohta. Regressioonimudelile antakse ette seega nii sisendandmeid kui oodatavaid vastuseid. Klassifikatsioonimudeli puhul võib mudel sisendandmeteks saada näiteks pildi autost ja väljundiks anda auto mudeli ja tootmisaasta. [18]

Juhendamata õppe puhul aga väljundandmeid treenimiseks ees ei ole ja mudeli ülesandeks jääb ise leida andmete erisusi. Lahendatavaks probleemiks võib olla näiteks veebipoe klientide grupeerimine kolmeks nende ostukäitumise põhjal. [18]

Kinnitusega õppimise puhul on tegemist dünaamilise keskkonnaga, milles loodav mudel peab tegema järjekorras mitmeid otsuseid, reageerides oma keskkonnale. Sellisel juhul ei ole määravaks ühe otsuse tegemise täpsus, vaid kogu otsuste tegemise loogika (*policy*), millest mudel juhendub. Kinnitusega õppimise meetodeid kasutavad näiteks arvutimänge või malet mängivad algoritmid. [18]

1.4.2 Konvolutsioonilised närvivõrgud

Konvolutsioonilised närvivõrgud on süvanärvivõrkude alamosa, mida kasutatakse näiteks pildiliste andmete puhul, milles sisendandmetel (pildi pikslitel) on omavahel teatav ruumiline seos. Näiteks pildi moodustavad pikslid on tasapinnal omavahelises seoses lähedal asuvate pikslitega ja samas koosneb värvilise pildi iga piksel veel kolmest värvikanalist (*RGB - Red, Green, Blue*). [19]

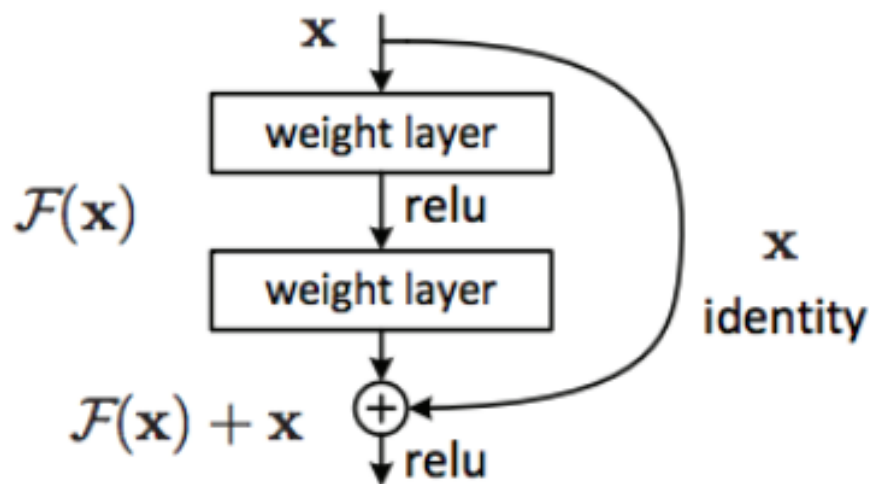
Konvolutsioon on konvolutsioonilise närvivõrgu protsessi peamine matemaatiline funktsioon, millest tuleneb ka mudeli nimetus. Selle kaks peamist osa on sisend ja ahendustuum (kernel). Konvolutsiooni väljundiks on tunnuskaart (feature map). Konvolutsioon on

protsess, mis võimaldab närvivõrgul õppida andmete struktuurist, näiteks pildi pikslite paiknemisest ja nendest joonistuvate kujunditest. [19] [20]

Konvolutsioonilistel mudelitel on mitu laiemalt levinud ja kasutatavat arhitektuuri, järgnevad kaks alapeatükki kirjeldavad lähemalt käesolevas töös rakendatavaid arhitekture.

1.4.3 ResNet mudel

Konvolutsioonilised mudelid on väga võimekad pildandmete töötlemisel ja on viinud mitmete läbimurreteni pildituvastuses. Üheks suureks probleemiks oli aga sügavamate mudelite loomine, sest senisel meetodil kihtide lisamisel mudeli täpsus ühest hetkest enam ei parane ja hakkab halvenema. He (2016) tuli välja uudse arhitektuuriga, mis lisas mudelisse kihte vahele jätvad ühendused (*skip connection*), tänu millele on võimalik luua varasemast kordades sügavamaid mudeleid. ResNet arhitektuur koosneb järjestikku kokku liidetud plokkidest, mille struktuur on näidatud Joonisel 3. Näites on näha välja pakutud ühendus (*identity*), mis ei lase ploki sisendiks olevatel väärtustel kaduda, sest varasema kihi sisend tuuakse mudeli hilisemasse kihti kaasa. [21]



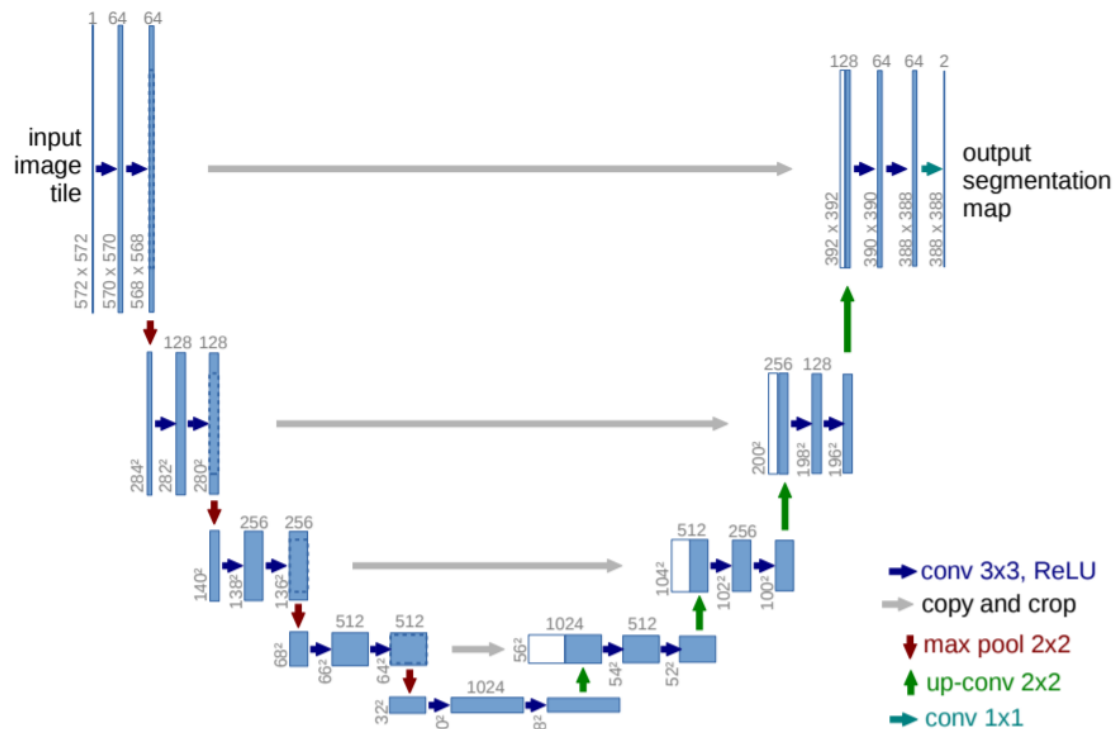
Joonis 3. ResNet arhitektuuri plokk.

1.4.4 U-Net mudel

U-Net mudel loodi esialgu meditsiiniliste piltide töötlemiseks, mille puhul treeningandmete kogus on tavaliselt vähene ja lisaks klassifitseerimisele on oluline ka pildi peal täpne lokaliseerimine. [22]

U-Net mudel koosneb kahest peamisest osast - kodeerija ja dekodeerija. Mudeli kodee-

rija ülesandeks on sisendandmetest eraldada kontekstiline teave ja sellega sümmeetrilise dekodeerija ülesandeks on kodeering väljundiks laiendada koos õpitud detailse lokaliseerimisega. [22]



Joonis 4. U-Net mudeli struktuuri näidis.

Näide U-Net mudeli struktuurist on toodud Joonisel 4. Lisaks kahele peamisele osale on mudelis kujutatud ka horisontaalsed ühendused mudeli sarnaste kihtide vahel kodeerija (kujutatud mudeli vasak pool) ja dekodeerija (parem pool) vahel. Selle ühenduse kohas kasutatakse kodeerija kihti sisendina dekodeerijale. Need ühendused annavad dekodeerijale tagasi andmeid sisendi detailsusest, mis kodeerimise protsessis kadusid. [22]

1.4.5 Siirdeõpe

Siirdeõpe (*transfer learning*) on masinõppe mudelite treenimise meetodika, mis kummutas senise arusaama, et mudeli treenimise jaoks on tarvis just konkreetse probleemi valdkonna treeningandmeid. Tihtilugu on piisava koguse treeningandmete kogumine aga väga tömahukas protsess, mis masinõppe rakendamist raskendab. Siirdeõppega kasutatakse mõne teise domeeni treeningandmete peal loodud mudelit ja treenitakse see ümber uude probleemi lahendamiseks. Mida sarnasemad on vastavad domeenid, seda väiksem on ümber treenimise vajadus. Näiteks erinevate pildituvastuse probleemide jaoks saab kasutada ImageNet andmestiku peal eeltreenitud mudeleid. Nende mudelite esimesed kihid

on väga hästi treenitud pildi madalama astme tunnuste tuvastamiseks. Sellised tunnused on näiteks kontuurjooned ja mustrid. Mudeli ümber treenimisel on aga vaja suuremat tähelepanu pöörata mudeli viimastele kihtidele, mis tuvastavad pildi kõrgema taseme tunnuseid (näiteks linnu või looma liikide tuvastamine). [23]

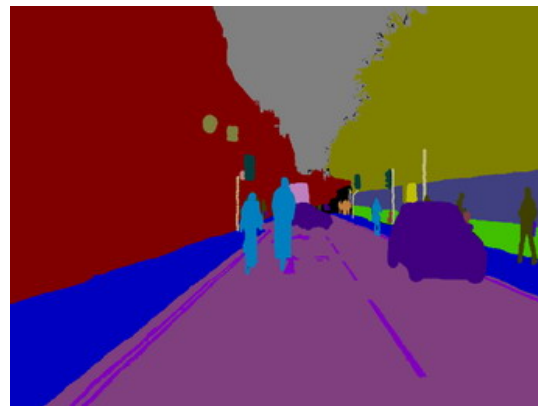
Siirdeõpe on laiemalt kasutust leidnud pildituvastuse ja keeletötluse valdkonnas, mille erinevate probleemide puhul on suur ühisosa. [23]

1.4.6 Semantiline segmenteerimine

Pildi segmenteerimine on pilditötluse protsess, mille puhul jaotatakse pilt mitmeks tähenduslikuks osaks. Ühtlasi saab seda vaadelda pildi osade vaheliste piirjoonte määramisena. Tehniliselt tähendab see pildi pikslite klassifitseerimist etteantud klasside vahel. Masinõppe vallas jaguneb segmenteerimine peamiselt kaheks meetodiks: klassipõhine (näide Joonisel 5 [24]) ja instantsipõhine segmenteerimine. Kui klassipõhise segmenteerimise puhul klassifitseeritakse iga piksel klasside vahel, siis instantsipõhise segmenteerimise puhul tuvastatakse iga individuaalne instants pildil (näiteks kolm jalgratturit pildil eraldi instantsidena tuvastatud). Segmenteerimise ülesannet lahendatakse laialdaselt näiteks meditsiini vallas, tervisehäirete tuvastamiseks pildipõhisel vaatlusel. [25]



(a) Sisend.



(b) Pildi segmentatsioon.

Joonis 5. Semantilise segmenteerimise sisend ja väljund CamVid andmekogu näitel.

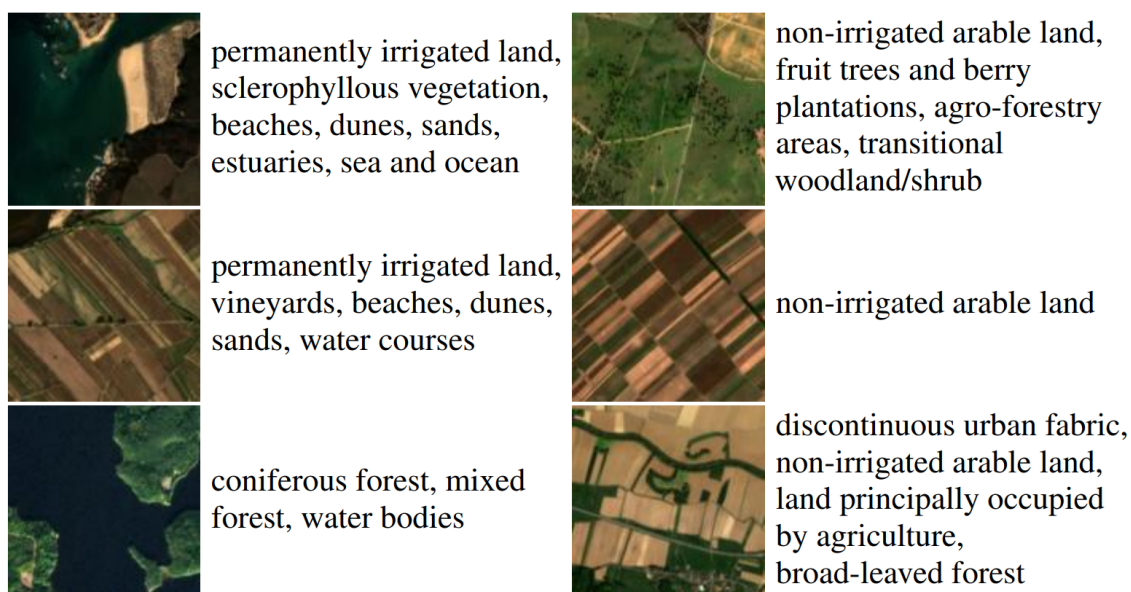
2. Metoodika

Käesolev peatükk kirjeldab lähemalt eksperimentideks kasutatud andmekogusid, loodud masinõppemudeleid ja selgitab andmete eeltötluse protsessi.

2.1 Andmekogud

2.1.1 BigEarthNet andmestik

BigEarthNet on 2019. aastal avalikustatud suuremahuline andmekogu masinõppe mudelite arendamiseks maapinna jälgimise vallas. Andmekogu koosneb 590 326 pildist, mis on võetud Sentinel-2 satelliidifotodest aastatest 2017 ja 2018. Andmestikus on iga satelliidipilt lõigatud 120x120 piksli suurusteks piltideks ja iga ruudu kohta on saadaval Sentinel-2 sensorite kõik 13 andmekanalit. [26]



Joonis 6. Näidis BigEarthNet andmekogust.

Andmestiku pildid on nendel oleva keskkonna järgi pealkirjastatud, ühel pildil võib olla mitu vastavat klassi. Kokku on andmestikus 43 eri klassi ja valdav enamus pilte sisaldavat rohkem kui ühte klassi (Joonis 6). Pildid jagunevad nelja aastaaja vahel (kevad: 189 276,

suvi: 128 951, sügis: 154 943, talv: 117 156) ja eraldi on välja märgitud 70 987 pilti, millel on kas täielik lumekate või pilvisus. [26]

Kõik BigEarthNet andmestikus kasutusel olevad maapinna klassifikatsioonid on välja toodud Lisas 1. Tabelis 2 on loetletud Eesti aladel esinevad maapinna klassid koos piltide arvuga BigEarthNet andmestikus, millel need klassid esinevad. Tabelist on ühtlasi ka näha, et kasutatavad andmed on tugevalt tasakaalust väljas. Arvestades reaalselt pinnakatte jagunemist on see erinevus selge, kuid masinõppe mudeli treenimisel tähendab see, et mudeli hindamisel ei piisa ühest arvulisest näitajast vaid tuleb vaadelda ka mudeli käitumist klasside lõikes [26].

Tabel 2. Maakatte klassifikatsioon

Tüübi kood	Maakatte tüüp	Piltide arv
1	Tehisalad	
111	Tiheda hoonestusega alad	10 784
112	Hõredalt hoonestatud alad	69 872
121	Tööstus- ja/või kaubandusterritooriumid	12 895
122	Maantee- ja raudteevõrk ja piirnev ala	3 384
123	Sadamad	509
124	Lennuväljad	979
131	Karjäärid	4 618
132	Prügiplatsid	959
133	Ehitusplatsid	1 174
141	Asula haljasalad	1 786
142	Puhkealad, pargid, kalmistud	5 353
2	Põllumajanduslikud alad	
211	Niisutuseta haritav maa	196 695
222	Puuvilja- ja marjaaiad	4 754
231	Karjamaad	103 554
242	Kompleksmaaviljelus (haritavat maad > 75%)	107 786
243	Põllumajanduslik maa (< 75%) loodusliku taimkatte osalusega	147 095
3	Metsad ja pool- looduslikud alad	
311	Heitlehised lehtmetsad	150 944
312	Okasmetsad	211 703
313	Segametsad	217 119
321	Looduslikud rohumaad	12 835
322	Loopealsed põõsastikud, nõmmraba ja nõmm	5 890

Jätkub...

Tabel 2 – Jät kub...

Tüübi kood	Maakatte tüüp	Piltide arv
324	Üleminekuline metsaala	173 506
331	Mererand, liivaluited, liivikud	1 578
333	Hõreda taimkattega alad	1 563
4	Märgalad	
411	Sisemaa sood	6 236
412	Turbaraba	23 207
421	Rannasoolakud	1 562
5	Vesi	
511	Vooluveed	10 572
512	Veekogud	83 811
521	Rannikulaguunid	1 498
523	Meri ja ookean	81 612

2.1.2 Sentinel-2 andmed

Käesoleva töö praktilises osas kasutatakse masinõppe mudeli treeningandmete loomiseks ühte Sentinel-2 satelliidifotot. Foto on valitud aasta 2018 maikuust, et oleks ligilähedane ajaline vastavus CORINE 2018 andmestikuga.

Satelliidifoto valikul sorteeritakse välja ilma pilvekatteta foto Kesk-Eestist, et saada võimalikult palju kasutatavaid andmeid ühe pildiga. Andmete laadimiseks kasutatakse EstHub riikliku satelliidiandmete keskuse portaali [27].

2.1.3 CORINE maakasutuse andmed

CORINE andmestikust kasutatakse magistritöös kõige hilisemaid, 2018. aasta andmeid. Andmete laadimiseks kasutatakse Copernicus Land Monitoring Service portaali [28]. Euroopat katvast maakasutuse kaardist eraldatakse seejärel osa mis ühtib satelliidifotoga.

2.2 Andmete eeltöötlus

Klassifitseerimise andmekogu

Klassifitseerimismudeli loomiseks kasutati BigEarthNet andmekogu, mis laeti Tensorflow kataloogist [29].

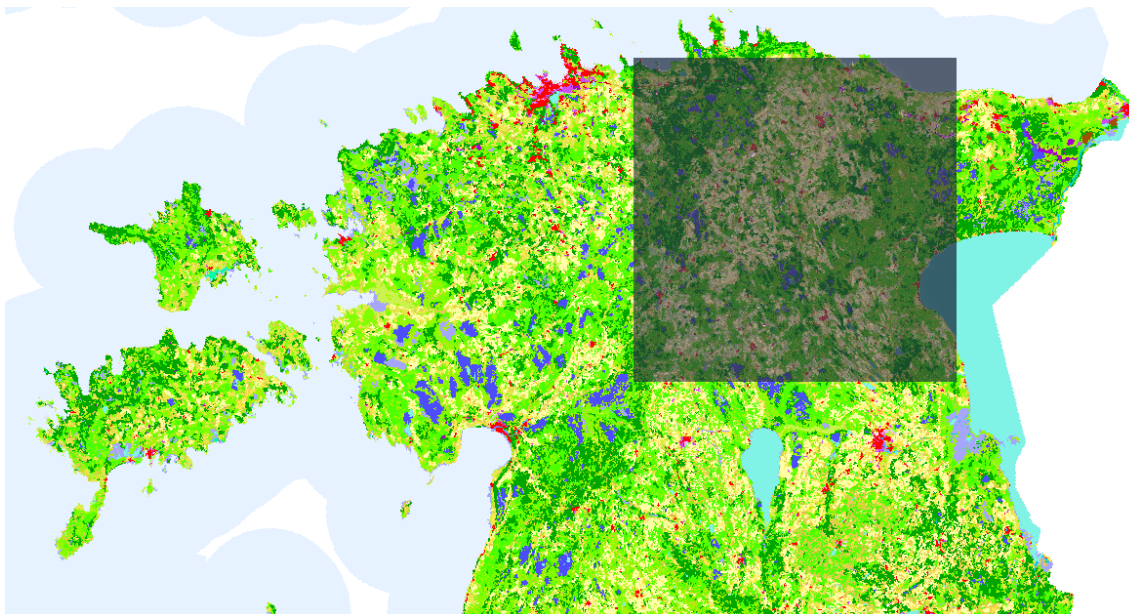
Mudeli treenimiseks RGB formaadis piltidega viidi esmalt läbi Sentinel-2 andmete kanalite 2, 3 ja 4 liitmine vastavalt .png formaadis pildifailideks. Andmete eeltötluse protsessi käigus eemaldati andmekogust 70 987 pilti, millel on kas täielik lumekate või pilvisus. Nimekiri vastavate pildifailidega ja skriptid vastavate piltide eraldamiseks on välja töötatud ja kättesaadavad BigEarthNet andmestiku autorite poolt [30]. Selle tulemusena jäi masinõppe mudeli treenimiseks andmekogu 519 339 pildiga.

Andmekogu pealkirjad on salvestatud iga pildiga kaasas olevas .json failis. Masinõppe mudeli treenimiseks salvestatakse piltide klassid eraldi .csv faili, mis võimaldab lugeda välja pildi nimele vastavad klassid. BigEarthNet andmekogu on loodud vastavalt CORINE maakasutuse kolmandale klassifikatsioonitasemele. Selleks, et võrrelda masinõppe mudelite täpsust eri tasemete vahel luuakse ühtlasi ka .csv failid esimese ja teise klassifikatsioonitaseme jaoks.

Mudelite treenimiseks valmistati seega ette pildid .png formaadis ja kolm .csv faili, milles on välja toodud pildi nimi ja sellele vastavad maakatte klassid.

Segmenteerimise andmekogu

Uue andmekogu loomiseks kasutatakse Sentinel-2 satelliidipilti ja sellega ühtiva ala maakasutuskaarti CORINE üleeuroopalisest andmekogust (illustreeritud Joonisel 7). Satelliidifoto ja maakasutuskaart jagatakse 120x120 piksli suurusteks piltideks ja salvestatakse eraldi kaustadesse. Igale ühtivale pildipaarile seatakse sama nimi, mis võimaldab mudeli treenimisel kasutada satelliidifotot sisendina ja maakasutuse pilti väljundina.



Joonis 7. Andmekogu loomiseks kasutatav piirkond.

Sentinel-2 foto:

S2A_MSIL1C_20180510T094031_N0206_R036_T35VMF_20180510T114819

CORINE maakasutuse kaardi versioon: 2018

Masinõppe mudeli treenimiseks on ühtlasi vaja ka maakasutuse kaardi piksli tasemel klassifikatsioonid ümber nimetada järjestikkusteks arvudeks 0-4, 0-14 ja 0-42 vastavalt CORINE klassifikatsiooni tasemetele 1-3. Sellest tulenevalt loodi kolm tekstifaili vastavalt piksli väärtuste ja maakasutuse klassifikatsioonidega.

Sarnaselt eelnevale töödeldakse ümber ka maakasutuse kaardi pildid, mis algselt on 3-kihilises RGB formaadis. Need töödeldakse ümber ühekihilisteks pildifailideks, mille piksliväärtused on samuti vastavalt vahemikkudes 0-5, 0-15 ja 0-43. Joonisel 8a on näha ühte treeningandmeteks kasutatud satelliidipilti. Sellele pildile vastav maakasutuskaardi väljalõige on Joonisel 8b, ning nende kahe pildi ülekate Joonisel 8c. Piltide ülekate vaadeldes on ühtlasi võimalik juba näha ka mõningaid probleeme andmekogu täpsuses. Näiteks võib eeldada, et kontuurid maakasutuskaardil peaksid langema kokku satelliidipildi kontuuridega. See on küll ligilähedane pildi alumises vasakus nurgas, kuid täpsus on kaheldav pildi keskosas ja ülemises vasakus nurgas.



(a) Sentinel-2 satelliidipilt

(b) CORINE maakasutuskaart

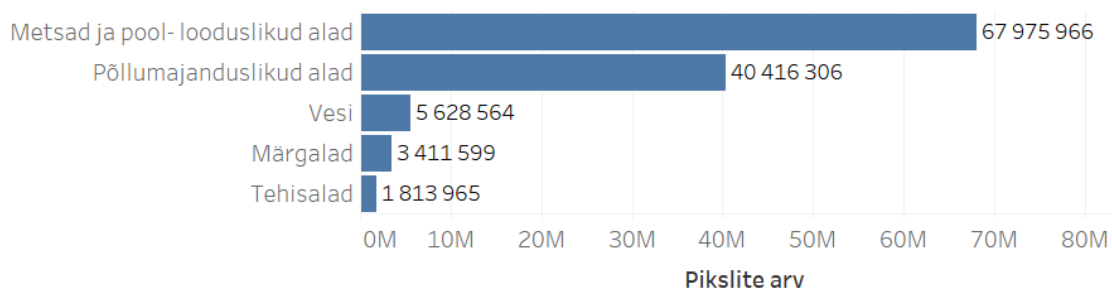
(c) Piltide ülekate

Joonis 8. Näide segmenteerimismudeli treenimiseks loodud andmekogust.

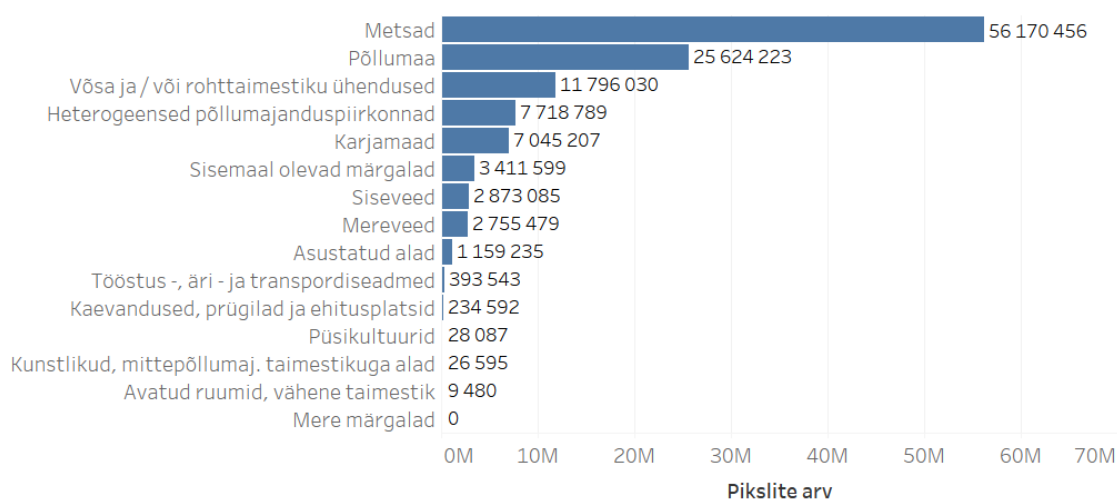
Treeningandmeteks valmistati seega ette üks pildikogu 120x120 suuruses satelliidipiltidega, kolm pildikogu samas suuruses ja ühtivate failinimedega segmentatsioonimaskidega, milles iga pildi piksel märgistab kindlat maakasutuse klassi ja kolm tekstifaili maakasutuse klassidega, mis on järjestatud samas järjekorras segmentatsioonimaskide pikslite arvuliste väärtustega.

Analüüsidest loodud andmekogu piksliväärtuste tasemel (Joonised 9, 10 ja 11) näitab kuivõrd palju on andmed tasakaalust väljas. Igal klassifikatsioonitaseme klasside jaotus annab ühtlasi ka märku loodavate mudelite täpsuse võimalikust jaotumisest. Vähem

esindatud alade puhul võib oodata ka väiksemat täpsust.

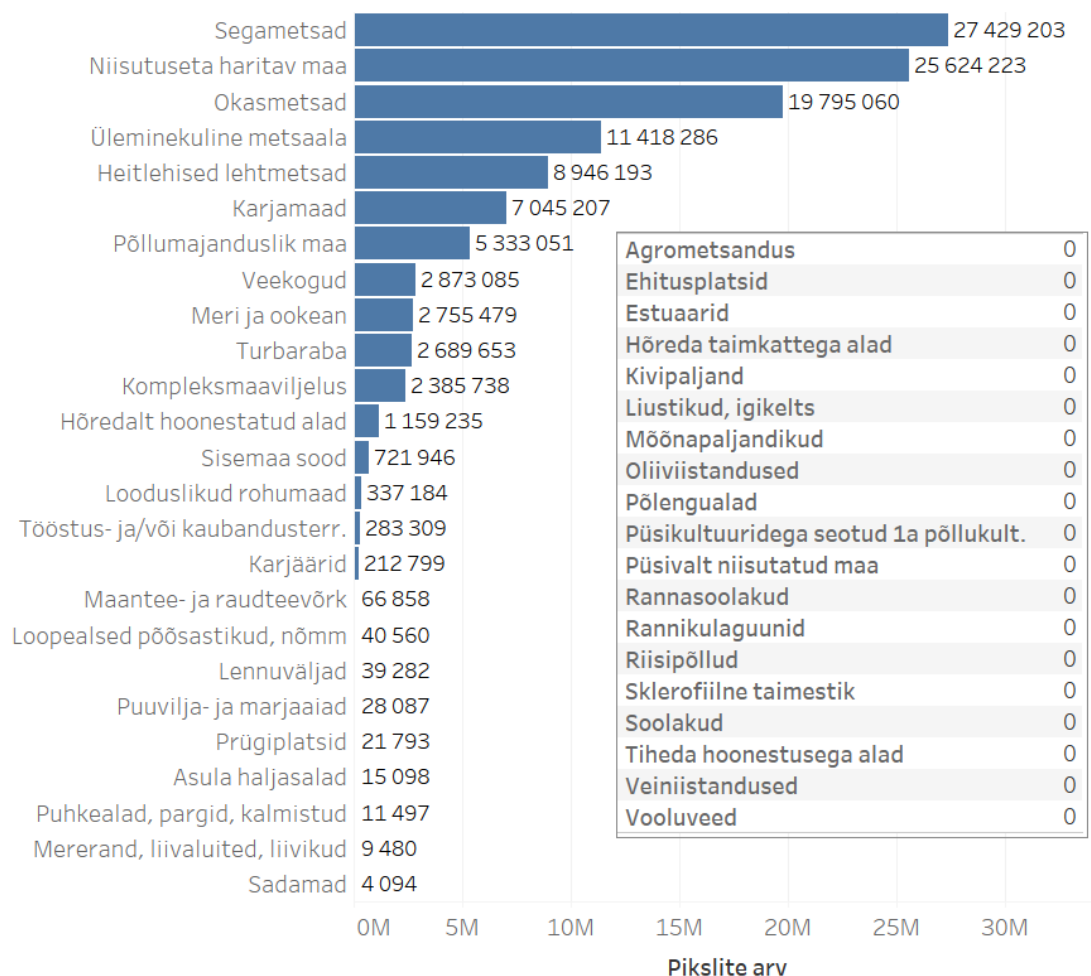


Joonis 9. Piksli tasemel klasside jaotus 1. tasemel.



Joonis 10. Piksli tasemel klasside jaotus 2. tasemel.

Klassifikatsiooni kolmanda taseme puhul (Joonis 11) eemaldas autor treenimise klasside nimekirjast need klassid, mida andmestikus ei esine (tabelis märgitud 0 väärtustega). Need klassid jääksid ka automaatselt mudeli treenimisest eemale ja nende eemaldamine eelnevalt võimaldab tulemustes paremini keskenduda andmestikus olevatele klassidele. Andmestikku jääb mudeli treenimiseks seega kokku 25 klassi.



Joonis 11. Piksli tasemel klasside jaotus 3. tasemel.

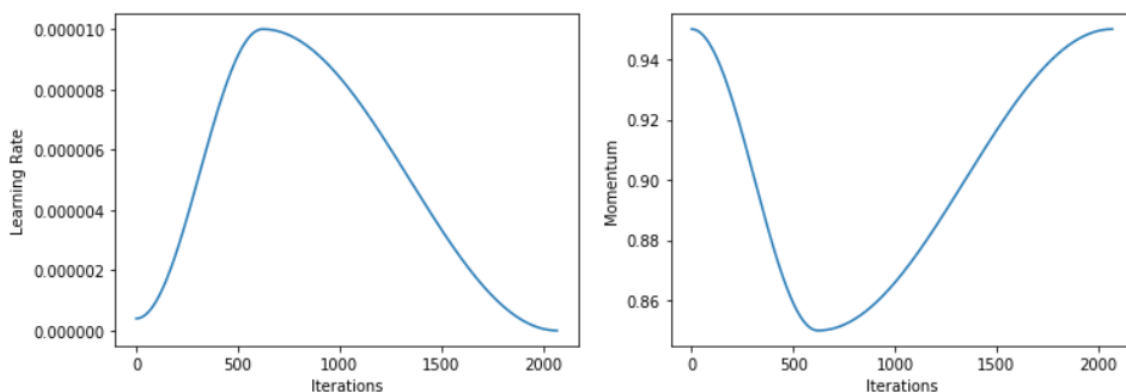
2.3 Mudelid

2.3.1 Masinõppe treenimiskeskond

Eksperimendid viidi läbi kasutades Fast.ai masinõppe teeki, mis on ehitatud PyTorch platvormile, kasutades ära selle võimekust teostada mudeli treenimise maatriksarvutused graafikakaardi peal.

Fast.ai teek on arendatud valdavalt kaasaegsete ja eksperimentaalsete masinõppemeetodite rakendamiseks. Näiteks on selles loodud meetodid mudeli õpismuutuse süstemaatiliseks leidmiseks [31] ja rakendatud õpismuutuse dünaamilist muutmist mudeli treenimise vältel (*One-cycle learning*), mida illustreerib Joonis 12. Sellelt näeme, et õppimise esimeses osas õpismuutus kasvab ja hoog väheneb, samas kui teises osas toimub vastupidine protsess [32]. Fast.ai on kaasaegne ja hästi sobiv just eksperimentaalsete ja akadeemiliste projektidel

rakendamiseks.



Joonis 12. One-Cycle meetod. Treenimise käigus muutuv õpisamm ja hoog.

Eksperimendid viiakse läbi Google Cloud virtuaalmasinas, kasutades JupyterLab arenduskeskkonda. Virtuaalmasin valitakse järgnevate tehniliste näitajatega: 8x vCPU, 30GB mälu, Nvidia K80 videokaart (12 GB). Mudelite treenimise protsessi kiirendamise eesmärgil kasutatakse ka Nvidia P100 graafikakaarti (16 GB).

2.3.2 Kasutatavad masinõppe mudelid

ResNet50

Magistritöö praktilise osa esimeses faasis treenitakse masinõppe mudel satelliidipiltide mitmeklassiliseks klassifitseerimiseks. Selle puhul rakendatakse ResNet50 mudeli struktuuri, eeltreenituna ImageNet andmekogu peal. Loodava kolme mudeli sisenditeks on BigEarthNet andmestikust võetud satelliidifotod ja väljundiks üks või mitu maakasutuse klassi, vastavalt kolmele klassifikatsiooni tasemele.

Kuna tegemist on eeltreenitud mudelitega ja rakendatakse siirdeõppe meetodikat viiakse mudelite treenimine läbi kahes etapis: esmalt treenitakse ainult mudeli viimaseid kihte ja hiljem tervet mudelit. Selline lähenemine kiirendab mudeli treenimise protsessi.

Mudeli treenimise **esimeses pooles** on valdav enamus treenitavaid parameetreid lukustatud:

Parameetreid kokku: 25,617,477

Treenitavaid parameetreid: 2,162,565

Lukustatud parameetreid: 23,454,912

Mudeli treenimise **teisese pooles** on kõik treenitavad parameetrid avatud:

Parameetreid kokku: 25,617,477

Treenitavaid parameetreid: 25,617,477

Lukustatud parameetreid: 0

U-Net

Praktilise osa teises faasis luuakse U-Net masinõppe mudel satelliidipiltide semantiliseks segmenteerimiseks. Mudeli loomisel kasutatakse esimeses faasis loodud segmenteerimismudelit kodeerija poolena ja sellega ühtivat tagurpidi mudelit dekodeerijana. Loodava kolme mudeli sisendiks on satelliidipilt ja väljundiks samas suuruses pilt, mille pikslid on klassifitseeritud väärtustes vahemikus 1-n, vastavalt maakasutuse klassifikatsiooni tasemetele.

Sarnaselt ResNet mudeli treenimisele viiakse ka U-Net mudeli puhul läbi siirdeõppe protsess, milles kasutatakse ära esimeses etapis eeltreenitud ResNet mudeleid. Need mudelid on selleks hetkeks treenitud satelliidipiltide lugemiseks ja neid rakendatakse U-Net mudeli kodeerija poolena.

U-Net mudeli treenimise **esimeses pooles** hoitakse mudeli kodeerija poole parameetrid lukus ja treenitakse vaid dekodeerija poolt:

Parameetreid kokku: 339,070,876

Treenitavaid parameetreid: 315,615,964

Lukustatud parameetreid: 23,454,912

Treenimise **teises pooles** avatakse treenimiseks ka mudeli kodeerija pool:

Parameetreid kokku: 339,070,876

Treenitavaid parameetreid: 339,070,876

Lukustatud parameetreid: 0

3. Eksperimendid

3.1 Eksperimendi disain

1. Klassifikatsioonimudelid BigEarthNet andmekogu põhjal

Eksperimendi esimeses osas luuakse kolm masinõppe klassifikatsioonimudelit kasutades BigEarthNet andmekogu. Need mudelid võimaldavad satelliidipildi klassifitseerimist vastavalt CORINE kolmele maakasutuse klassifikatsiooni tasemele. Need mudelid on ühtlasi ka sisendiks eksperimendi teisele osale.

2. Segmenteerimismudelid maakasutuse andmekogu põhjal

Eksperimendi teises osas kasutatakse esimeses osas loodud klassifikatsioonimudeleid, et luua U-Net struktuuriga segmenteerimismudelid kõigi kolme klassifikatsioonitaseme jaoks. Seda mudelit treenitakse käesoleva töö käigus loodava andmekogu põhjal maakasutuse segmenteerimise eesmärgil.

3.2 Eksperimendi tulemuste mõõtmine

3.2.1 Andmekogude hindamine

BigEarthNet andmekogu puhul on tegemist mitmeklassiliste andmetega, milles ühele pildile võib vastata rohkem kui üks pealkiri. Andmekogu mitmeklassilisuse hindamiseks kasutab autor järgnevat kahte valemit. [33]

Pealkirjade kardinaalsuse näitaja annab keskmise pealkirjade arvu ühe elemendi kohta ja see arvutatakse valemiga 3.1.

$$Kardinaalsus = \frac{1}{N} \sum_{i=1}^N |Y_i| \quad (3.1)$$

Pealkirjade tiheduse arvutamiseks leitakse pildiga seonduvate pealkirjade arvu ja kogu

pealkirjade arvu jagatis, keskmistatult üle kogu andmestiku (valem 3.2). Mida väiksem on kardinaalsus, seda keerulisem valik on mudelil vaja teha, sest võimalike pealkirjade valik on suurem.

$$Tihedus = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i|}{|L|}, \quad \text{milles } L = \bigcup_{i=1}^N Y_i \quad (3.2)$$

Bernardini [34] toob välja, et mitmeklassilise andmekogu puhul on klassifikatsioonimudeli tulemused tugevalt seotud andmete kardinaalsuse ja tihedusega. Kardinaalsuse vähenedes või tiheduse suurenedes masinõppe mudeli tulemused paranevad.

3.2.2 Klassifitseerimismudeli hindamine

Klassifitseerimismudeli täpsust hinnatakse BigEarthNet andmekogu abil, sellest eraldatud valideerimisandmete abil. Valideerimisandmestikku eraldatakse juhuslikkuse alusel 20% kogu andmestikust, 103 867 pilti. Seda osa andmestikult ei kasutatud mudeli treenimisel. Et tagada eksperimentide võrreldavus, kasutatakse iga mudeli puhul sama validatsiooniandmestikku.

Uurimuse esimese osa puhul on tegemist mitmeklassilise klassifitseerimisprobleemiga, mille treeningandmed on tugevalt tasakaalust väljas (Tabel 2). Sellest tulenevalt on vajalik analüüsida lisaks üldisele täpsusele ka iga klassi täpsust eraldiseisvalt.

Täpsus (*Precision*) - Kui suur osa kõigist mudeli poolt välja valitud pealkirjadest on korrektsed (Valem 3.3). [35]

$$Precision = \frac{TP}{TP + FP} \quad (3.3)$$

TP - Tõeselt positiivne (*True Positive*); *FP* - Valepositiiv (*False Positive*)

Saagis (*Recall*) - Kui suur osa kõigist korrektsetest pealkirjadest on mudeli poolt välja valitud (3.4). [35]

$$Recall = \frac{TP}{TP + FN} \quad (3.4)$$

TP - Tõeselt positiivne (*True Positive*); *FN* - Valenegatiiv (*False Negative*)

Klassifitseerimisprobleemi lahenduse täpsuse jälgimise mõõdikuna mudeli treenimisel kasutatakse F_β väärtust, mis võtab arvesse nii täpsust kui saagist. Selle valemis β väärtus näitab kui mitu korda arvestatakse täpsus olulisemaks saagisest. Käesolevas töös vaadeldakse nii F_1 väärtust klassifikatsiooni klasside omavahelisel võrdlusel kui ka F_2 väärtust kogu mudeli peale üldistatuna. Väärtus 2 tähendab, et korrektsete pealkirjade määramine loetakse kaks korda olulisemaks sellest, kui mitu välja pakutud pealkirja on korrektsed. Valem 3.5 kujutab üldistatud F_β valemit. F väärtus on vahemikus [0-1], optimaalse väärtusega 1. [35]

$$F_\beta = (1 + \beta^2) \frac{precision * recall}{(\beta^2 * precision) + recall} \quad (3.5)$$

Käes olevas töös kasutatakse $\beta = 2$, ehk tegemist on F_2 väärtusega.

Lisaks rakendatakse mõõdikuna absoluutset täpsust. Kuna tegemist on mitmeklassilise klassifikatsiooniülesandega, siis absoluutne täpsus ei anna mudelit täielikku ülevaadet. See näitab vaid osakaalu valideerimisandmetest, mille puhul kõik pildi pealkirjad on korrektselt määratud, jättes kõrvale osaliselt korrektsed juhud.

$$Abs.tapsus = \frac{1}{N} \sum_{i=1}^N (Y_i = Z_i) \quad (3.6)$$

3.2.3 Segmenteerimismudeli hindamine

Segmenteerimismudeli hindamine toimub U-Net mudeli treenimiseks loodud andmekogu abil selleks eraldatud valideerimisandmete põhjal. Valideerimisandmestikku eraldata juhuslikkuse alusel 20% kogu andmestikust, 1 656 pilti. Seda osa andmestikult ei kasutatud mudeli treenimisel. Et tagada eksperimentide võrreldavus, kasutatakse iga mudeli puhul sama valideerimisandmestikku.

Mudeli treenimise käigus hinnatakse mudeli täpsust esiteks piksli tasemel täpsusega (Valem 3.7), ehk keskmiselt kui suur osa pikslitest on korrektselt klassifitseeritud. See näitaja kujutab küll üldist täpsust, kuid selle peamiseks puuduseks on klasside ebavõrdsuse mitte arvestamine.

$$\text{Segmenteerimise täpsus} = \frac{\text{Korrektselt klassifitseeritud pikslite arv}}{\text{Kogu pikslite arv}} \quad (3.7)$$

Kuna tegemist on mitmeklassilise segmenteerimise probleemiga, mille andmestik on tugevalt tasakaalust väljas, siis analüüsib autor saadud tulemusi ka eksimismatriksi abil, milles väljendub mudeli täpsus klasside põhisel. Segmenteerimise täpsuse hindamiseks kasutab autor IoU skoori (*IoU - Intersection over Union*) (Valem 3.8). See on ühtlasi tuntud ka nimetusega *Jaccard Index*. [36]

$$\text{Jaccard Index} = \frac{TP}{TP + FN + FP} \quad (3.8)$$

TP - Tõeselt positiivne (True Positive); FN - Valenegatiiv (False Negative); FP - Valepositiiv (False Positive)

4. Tulemused

4.1 Andmekogude näidikud

Klassifitseerimismudeli treenimiseks kasutatud BigEarthNet andmekogu kirjeldamiseks arvutati kaks peamist näitajat, mis kirjeldavad andmekogu mitmeklassilisust - andmete kardinaalsus ja tihedus. Seda tehti kõigi kolme maakasutuse klassifikatsiooni taseme puhul. Saadud tulemused on välja toodud Tabelis 3.

Tabel 3. BigEarthNet andmete kardinaalsus ja tihedus.

Klassifikatsiooni tase	Kardinaalsus	Tihedus
1. tase	1.78	0.357
2. tase	2.48	0.165
3. tase	2.96	0.069

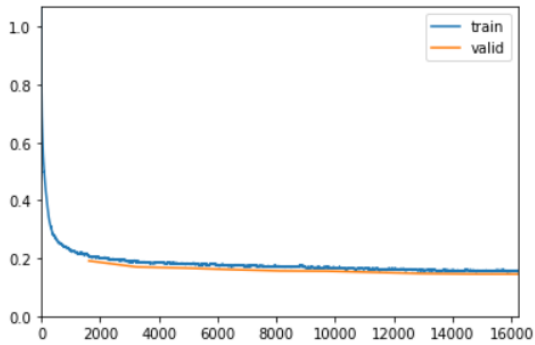
Tulemustest on näha, et kui klassifikatsiooni esimest taset jälgides on keskmiselt ühe pildi kohta 1.78 pealkirja, siis kolmandal tasemel on see ligi 3. Vastupidiselt sellega aga tihedus väheneb väärtuselt 0.357 esimesel tasemel väärtuseni 0.069 kolmandal tasemel. Muutus andmete tiheduses on märgatavalt suurem muutusega kardinaalsuses (vastavalt 1.66x ja 5.17x). See näitab, et liikudes suurema klassifikatsiooni taseme peale pealkirjade detailsus suureneb tugevalt, kuid ühe pildi peal olevate pealkirjade arv kasvab väiksemal määral.

4.2 Klassifitseerimismudeli tulemused

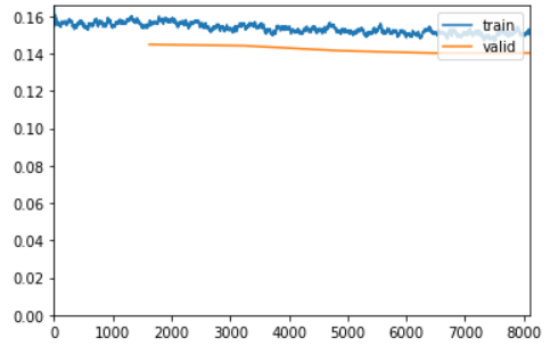
4.2.1 Maakasutuse klassifikatsiooni 1. tase

Klassifitseerimismudeli treeningu protsess toimus kahes faasis, kokku 15 epohhi vältel (10 esimeses ja 5 teises etapis). Esimeses etapis (Joonis 13a) treeniti vaid mudeli viimaseid kihte, teises faasis (Joonis 13b) aga kogu mudelit.

Treenimise protsessi analüüsidest on näha mudeli kaofunktsiooni kiire lähenemine madalale ja stabiilsele positsioonile juba esimese epohhi vältel. Sellest edasi mudel küll paranes, kuid vähesel määral. Teise etapi käigus oli samuti näha, et mudel enam paremaks ei muutunud.



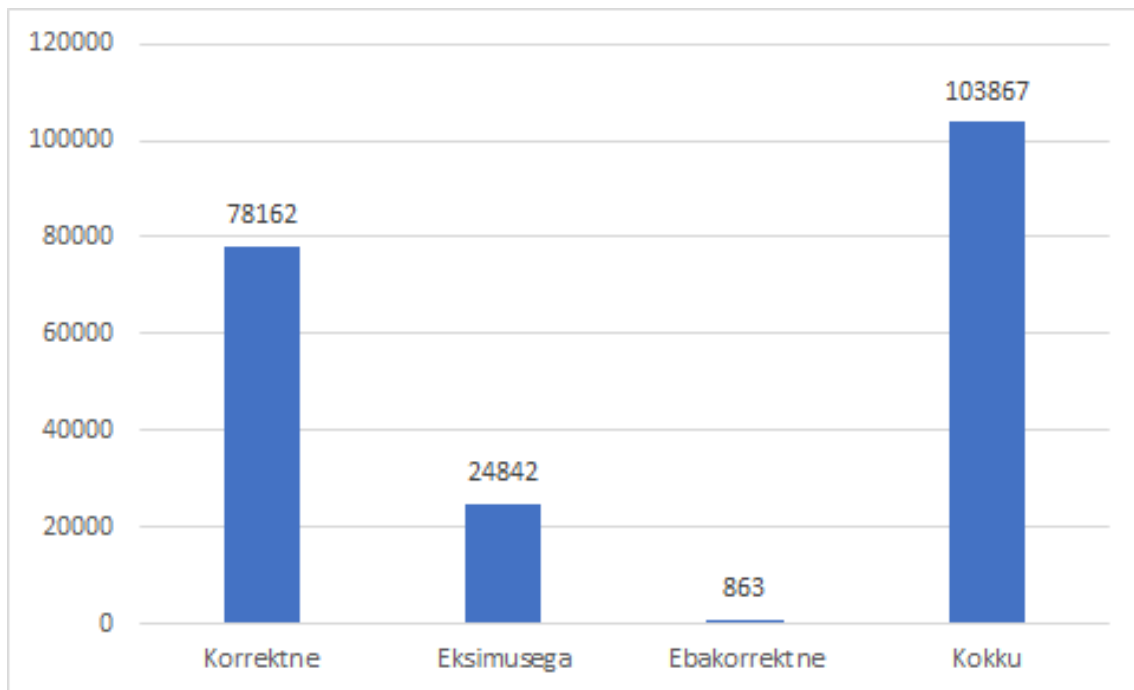
(a) Esimene faas.



(b) Teine faas.

Joonis 13. Mudeli treenimise progress (ResNet-50), klassifikatsiooni 1. tase.

Klassifikatsioonimudeli absoluutset täpsust, kõige rangemat kriteeriumit, iseloomustab Joonis 14, mille on näha validatsiooniandmestiku klassifitseerimise tulemused pildi kõikide pealkirjade kohta. Suurema osa piltidest (75.3%) klassifitseerib mudel täielikult õigesti, 23.9% puhul teeb eksimuse pildi mõne pealkirja osas ja 0.8% piltide puhul eksib mudel pealkirjadega täielikult.



Joonis 14. Mudeli absoluutne täpsus 1. taseme klassifikatsiooni määramisel.

Loodud mudeli F_2 skoor kõikide klasside peale kokku on 0.930, mis näitab mudeli kõrget täpsust. Kui vaatleme mudeli täpsuse näitajaid klasside lõikes (Joonis 4) siis näeme, et kõige nõrgemalt õnnestub mudelil märgalade tuvastamine. Saagis 0.45 näitab, et kõikidest piltidest millel on kujutatud märgalad vaid 45% suudab mudel tuvastada. Suuremate klasside puhul on aga näitajad väga head, mudel näitab kõrget täpsust ja saagist.

Tabel 4. Klassifikatsioonimudeli hindamine 1. tasemel.

Klass	Täpsus	Saagis	F ₁	Pilte
Põllumajanduslikud alad	0.95	0.93	0.94	64 182
Tehisalad	0.85	0.71	0.77	17 689
Mets	0.94	0.95	0.95	70 138
Vesi	0.95	0.91	0.93	28 570
Märgalad	0.85	0.45	0.58	4 869

Klassifikatsioonimudeli täpsuse analüüsimisel annab ülevaadet klasside põhiste eksimismatriksite vaatlemine (Joonis 15). Selles näeme, et mudeli tulemused klasside lõikes on võrdlemisi sarnased ja ei ole näha märkimisväärselt suuremaid eksimusi ühegi puhul.

Põllumajanduslikud alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	36 692	2 993
	Jah	4 581	59 601

Tehisalad

		Pakutud	
		Ei	Jah
Korrektne	Ei	83 964	2 214
	Jah	5 164	12 525

Metsad ja pool-looduslikud alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	29 781	3 948
	Jah	3 684	66 454

Vesi

		Pakutud	
		Ei	Jah
Korrektne	Ei	74 011	1 286
	Jah	2 627	25 943

Märgalad

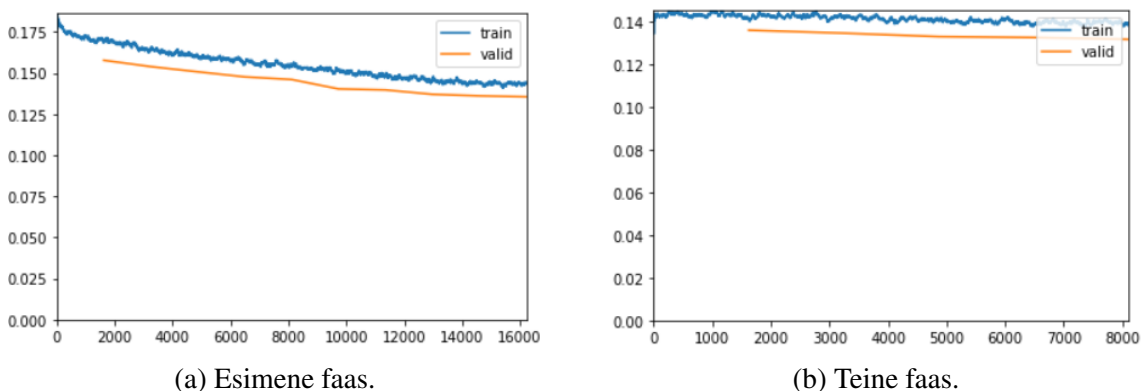
		Pakutud	
		Ei	Jah
Korrektne	Ei	98 608	390
	Jah	2 700	2 169

Joonis 15. Klasside põhised eksimismatriksid 1. tasemel.

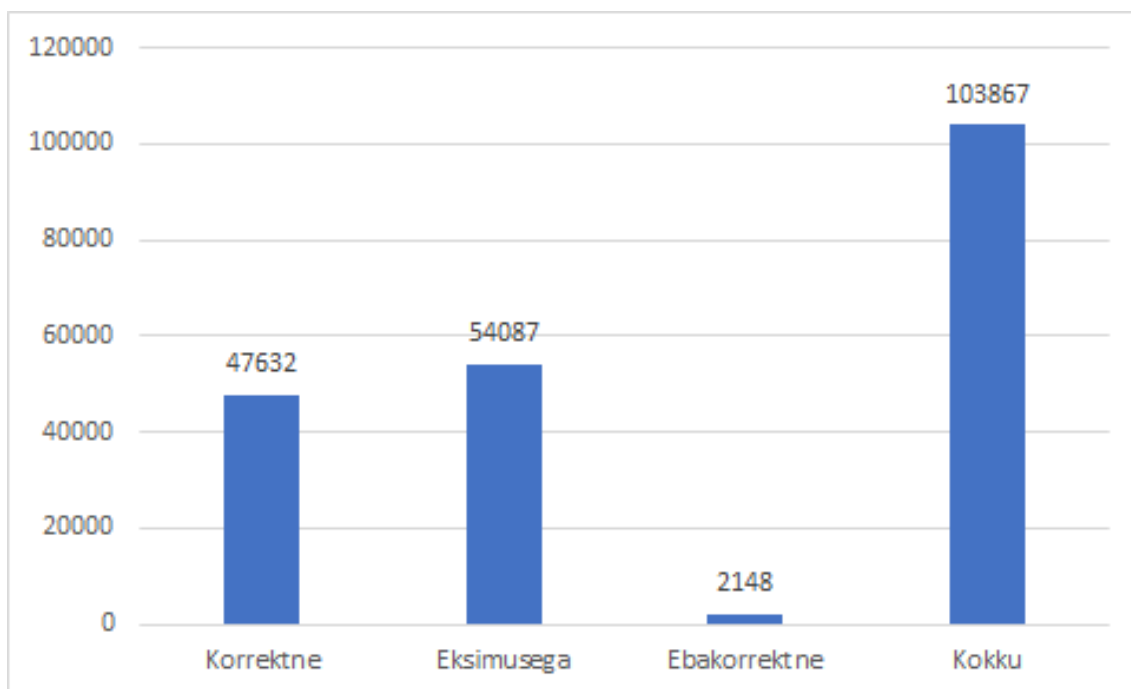
4.2.2 Maakasutuse klassifikatsiooni 2. tase

Treeningu progress toimus kahes faasis, kokku 15 epohhi vältel (10 esimeses ja 5 teises etapis). Esimeses etapis (Joonis 16a) trenniti vaid mudeli viimaseid kihte, teises faasis (Joonis 16b) aga kogu mudelit.

Mudeli kaofunktsioonid näitasid esimeses etapis võrdlemisi stabiilset langust ning saavutasid stabiilse positsiooni teises etapis.



Joonis 16. Mudeli treenimise progress (ResNet-50), klassifikatsiooni 2. tase.



Joonis 17. Mudeli absoluutne täpsus 2. taseme klassifikatsiooni määramisel.

Klassifikatsioonimudeli absoluutset täpsust, kõige rangemat kriteeriumit, iseloomustab Joonis 17, mille on näha validatsiooniandmestiku klassifitseerimise tulemused pildi kõikide pealkirjade kohta. 45.9% piltidest klassifitseerib mudel täielikult õigesti, 52.1% puhul

teeb eksimuse pildi mõne pealkirja osas ja 2.1% piltide puhul eksib mudel pealkirjadega täielikult.

Loodud mudeli F_2 täpsus kõikide klasside peale kokku on 0.853, mis näitab mudeli võrdlemisi kõrget täpsust. Kui vaatleme mudeli täpsuse näitajaid klasside lõikes (Joonis 5) siis näeme, et kõige nõrgemalt õnnestub mudelil kunstlike mittepõllumajandusliku taimeistikuga alade määratlemine, mille puhul suutis vaid 11% tuvastada (saagis 0.11). Tulemused suuremate klasside puhul on kõrged (põllumaa, metsad ja merevesi kõik üle 85%).

Tabel 5. Klassifikatsioonimudeli hindamine 2. tasemel.

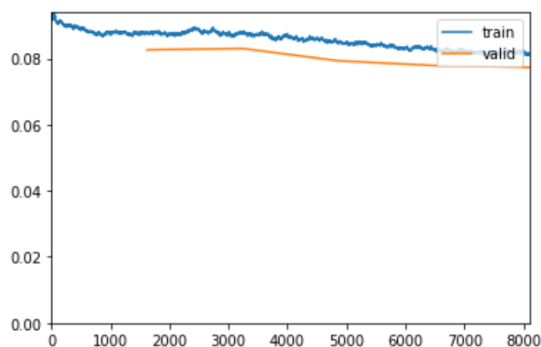
Klass	Täpsus	Saagis	F_1	Pilte
Põllumaa	0.87	0.85	0.86	38 755
Kunstlikud, mittepõllumaj. taimeistikuga alad	0.74	0.11	0.20	1 299
Metsad	0.93	0.94	0.94	64 364
Heterogeensed põllumaj. piirkonnad	0.83	0.77	0.80	42 671
Tööstus -, äri - ja transpordialad	0.73	0.44	0.55	3 050
Siseveed	0.90	0.79	0.84	13 542
Sisemaal olevad märgalad	0.84	0.45	0.58	4 424
Mereveed	0.99	0.98	0.98	15 083
Mere märgalad	0.84	0.42	0.56	455
Kaevandused, prügilad ja ehitusplatsid	0.81	0.16	0.26	1 197
Avatud ruumid, kus taimeistik on vähene või puudub üldse	0.74	0.44	0.55	1 199
Põllumaa	0.83	0.67	0.74	19 867
Püsikultuurid	0.75	0.43	0.54	4 968
Võsa ja / või rohttaimeistiku ühendused	0.75	0.59	0.66	31 820
Asustatud alad	0.83	0.72	0.77	15 118

Klassifikatsioonimudeli klasside põhised eksimismatriksid on toodud Lisas 2. Teistest suuremaid eksimusi võime märkida selle põhjal heterogeense põllumajanduspiirkonna, võsa / rohttaimeistiku ühenduste ja põllumaade puhul, kõigi kolme klassi puhul eksimusega piltide arv validatsiooniandmestikus üle 10 000.

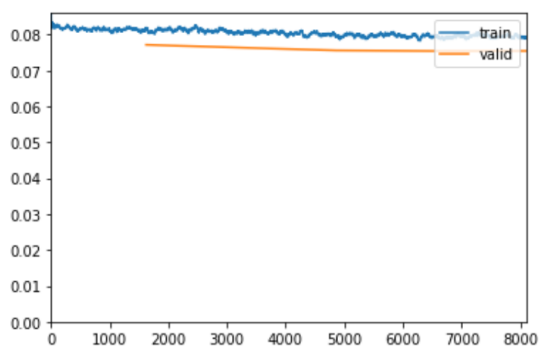
4.2.3 Maakasutuse klassifikatsiooni 3. tase

Treeningu progress toimus kahes faasis, kokku 15 epohhi vältel (5 esimeses ja 5 teises etapis). Esimeses etapis (Joonis 18a) treeniti vaid mudeli viimaseid kihte, teises faasis (Joonis 18b) aga kogu mudelit.

Mudeli kaofunktsioonid saavutasid esimese epohhi jooksul stabiilse positsiooni ja sellest edasi paranes vaid vähesel määral.



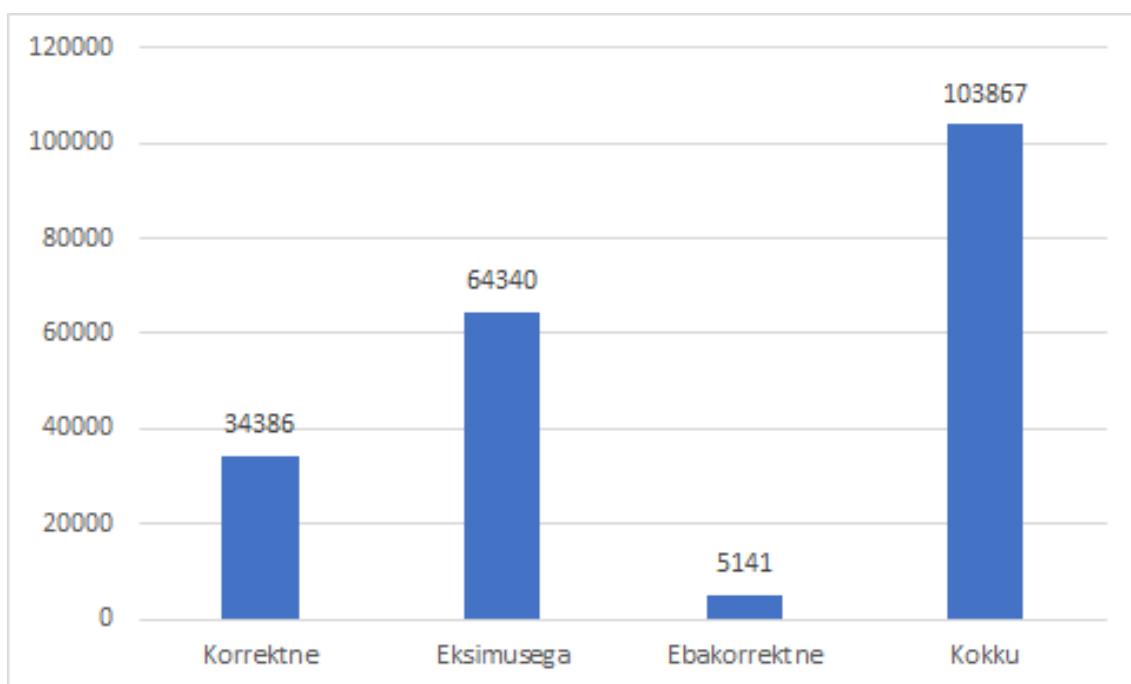
(a) Esimene faas.



(b) Teine faas.

Joonis 18. Mudeli treenimise progress (ResNet-50), klassifikatsiooni 3. tase.

Klassifikatsioonimudeli absoluutset täpsust, kõige rangemat kriteeriumit, iseloomustab Joonis 19, mille on näha validatsiooniandmestiku klassifitseerimise tulemused pildi kõikide pealkirjade kohta. 33.1% piltidest klassifitseerib mudel täielikult õigesti, 61.9% puhul teeb eksimuse pildi mõne pealkirja osas ja 4.9% piltide puhul eksib mudel pealkirjadega täielikult.



Joonis 19. Mudeli absoluutne täpsus 3. taseme klassifikatsiooni määramisel.

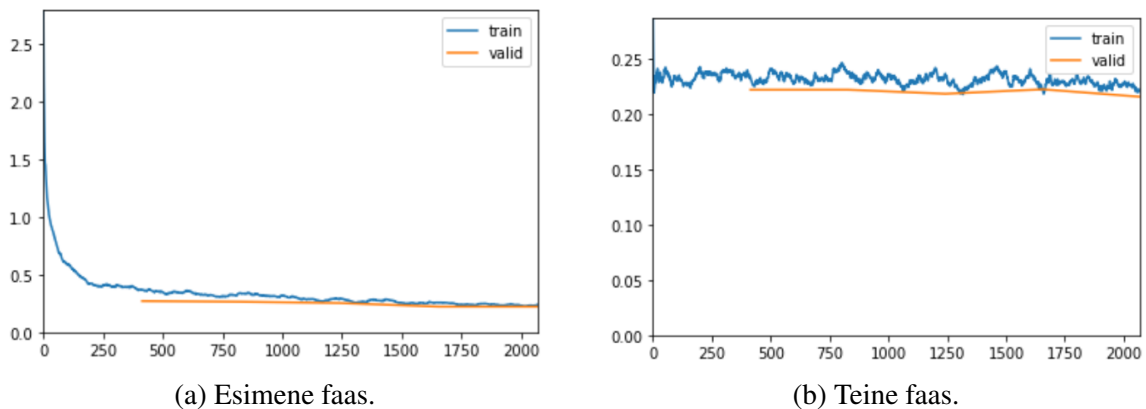
Loodud mudeli F_2 täpsus kõikide klasside peale kokku on 0.793, mis on nõrgem eelneva kahe mudeliga võrreldes, kuid lahendatav ülesanne on samuti keerukam. Kui vaatleme mudeli täpsuse näitajaid klasside lõikes (vt Lisa 3) siis näeme, et tulemusi on mõlemas äärmuses. Erinevate metsatüüpide puhul on F_1 tulemus (arvestab võrdse kaaluga nii täpsust kui saagist) kõigil üle 0.80 ja samuti veekogudega seotud klassid ja põllumaad.

Klassifikatsioonimudeli klasside põhised eksimismatriksid on toodud **Lisas 2**. Selles on näha mudeli käitumist vähe esindatud klasside puhul. Näiteks ehitusplatside ja põlen-gualade puhul mudel neid välja ei pakkunudki (lennuväljade puhul pakkus kahel korral ja ühe puhul oli pakkumine täpne).

4.3 Segmenteerimismudeli tulemused

4.3.1 Maakasutuse klassifikatsiooni 1. tase

Mudelit treeniti kokku 10 epohhi, mille jooksul jõudis mudel stabiilsesse seisu. Joonis 20 näitab mudeli kadu treening- ja validatsiooniandmestiku puhul terve treenimise protsessi vältel. Graafikult on näha, et mudeli kadu on oma miinumile lähenemas ning suuremat täpsuse kasvu valitud mudeli arhitektuuri ja hüperparameetrite puhul ei ole oodata. Mudeli üldine piksli tasemel täpsus on 91.4%.



Joonis 20. Mudeli treenimise progress (U-Net), klassifikatsiooni 1. tase.

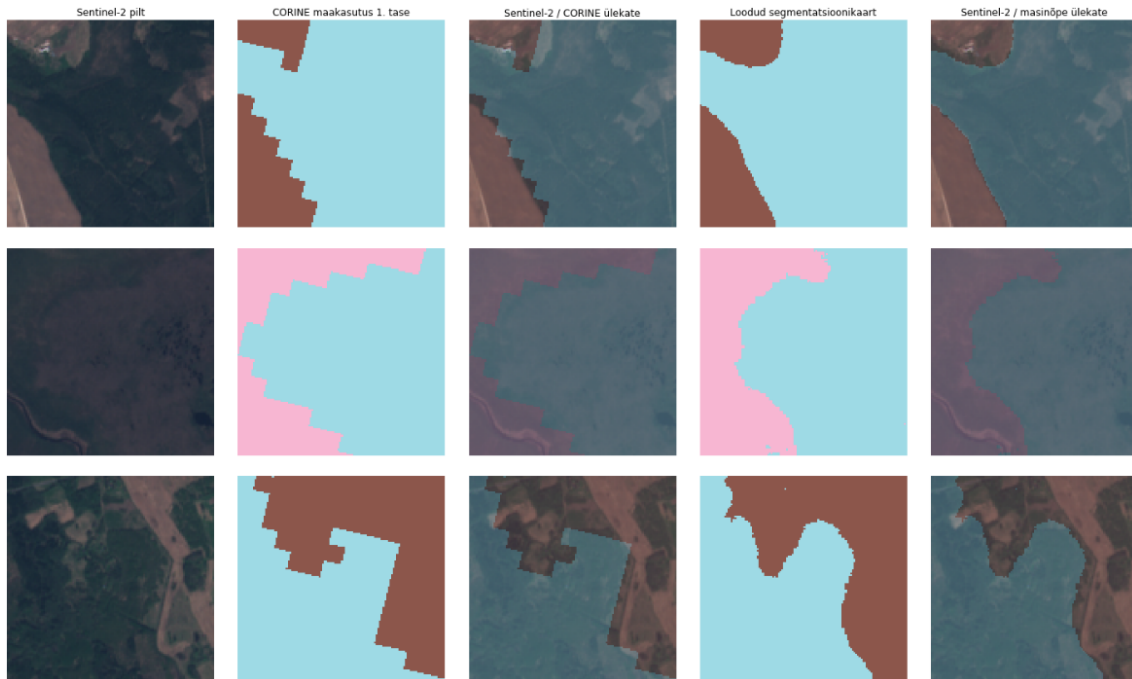
Segmenteerimismudeli täpsused klasside lõikes (Tabel 6.) näitavad üle 70% täpsust metsade, veekogude ja põllumaade puhul ja madalat täpsust tehislake ja soode puhul.

Tabel 6. Segmenteerimise täpsus maakasutuse 1. taseme klasside lõikes.

Maakasutuse klass	IoU skoor
Metsad ja pool- looduslikud alad	0.870
Vesi	0.813
Põllumajanduslikud alad	0.730
Tehisalad	0.426
Märgalad	0.343

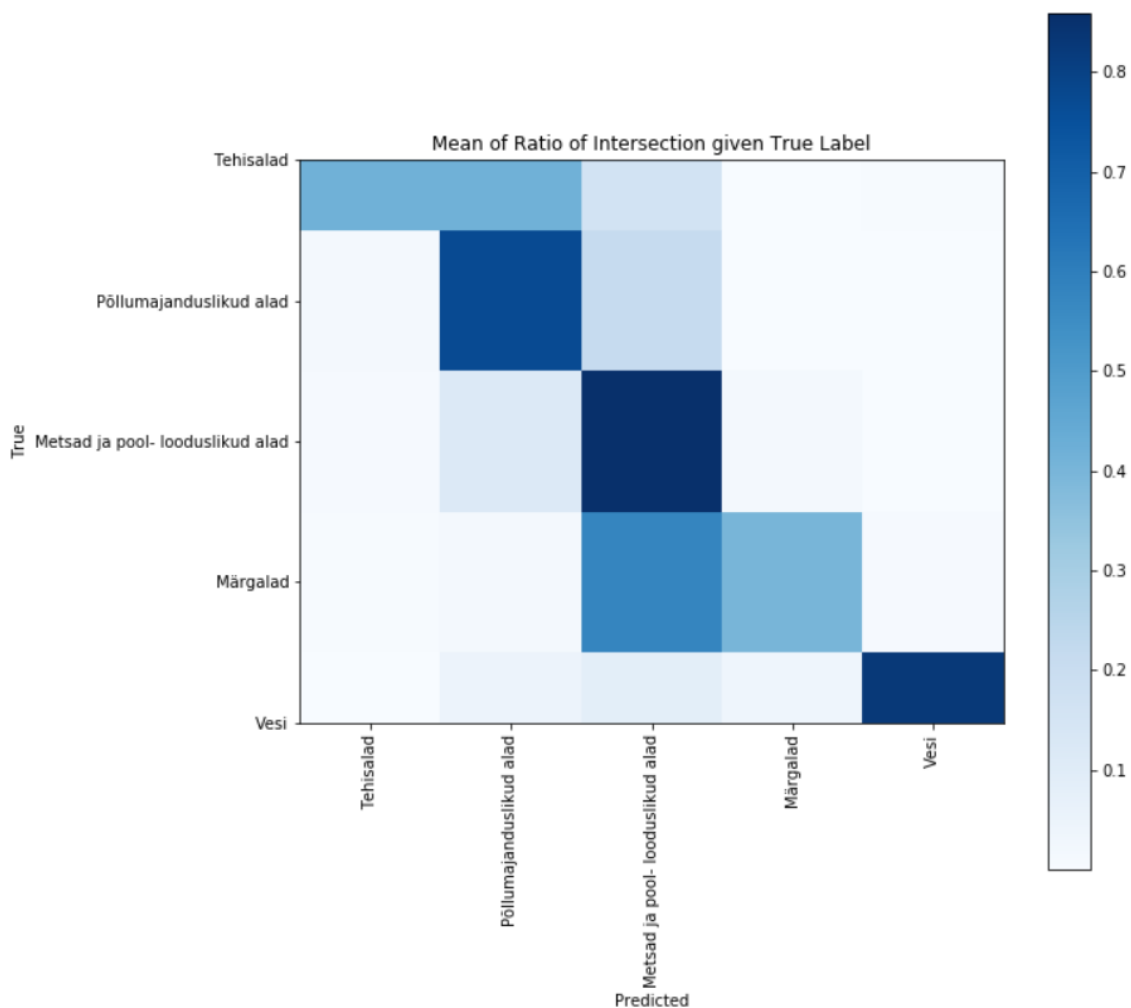
Näited segmenteerimismudeli tulemustest Joonisel 24 kõrvutavad segmenteerimismudeli tulemused (neljandas tulbas) sisendandmeteks kasutatud maakasutuskaardiga (teises tulbas). Välja toodud näidetes on võimalik ära tunda sarnaseid kontuure kahe kaardi vahel ja on võimalik näha masinõppe mudeli poolt tehtud eralduste kõrgemat täpsust. Kuna

maakasutuse esimesel tasemel on viis klassi, siis ühe pildi peal on erinevate klasside arv väike ja pilte vaadates on näha ka võrdlemisi täpseid piirjooni.



Joonis 21. Segmenteerimismudeli tulemuste näide, maakasutuse 1. tase.

Eksimismaatriks Joonisel 22 näitab, et mudeli suuremad eksimised tulenevad märgalade valesti klassifitseerimisel metsaks ja tehnilike alade klassifitseerimisel põllumajanduslikuks maaks. Kuna mets on ülekaalukalt kõige enam esindatud maakatte liik, siis on sellest selgitatav ka trend teiste klasside valesti klassifitseerimisel metsaks.

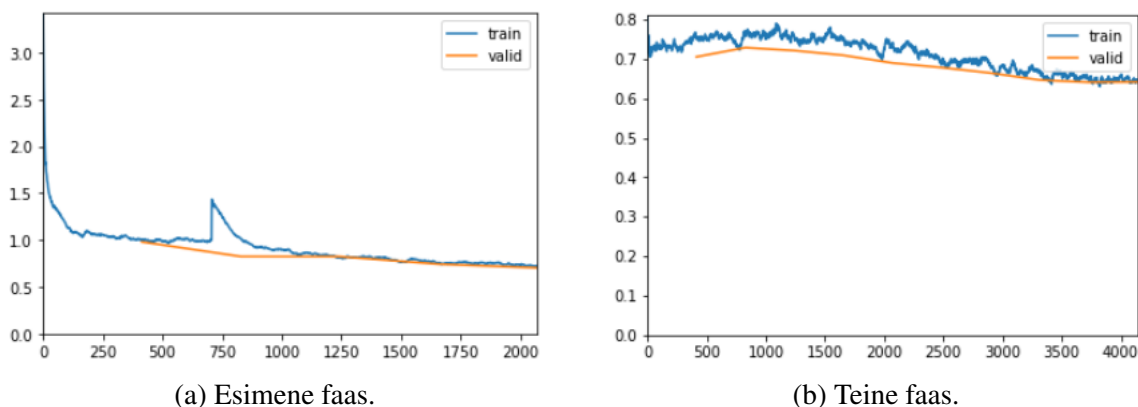


Joonis 22. Segmenteerimise eksimismatriks maakasutuse 1. taseme klasside lõikes.

4.3.2 Maakasutuse klassifikatsiooni 2. tase

Mudelit treeniti kokku 15 epohhi (5+10), mille jooksul jõudis mudel stabiilsesse seisusse. Joonis 23 näitab mudeli kadu treening- ja validatsiooniandmestiku puhul treenimise protsessi vältel esimeses ja teises faasis. Graafikult on näha, et mudeli kadu on oma miinimumile lähenemas ning suuremat täpsuse kasvu valitud mudeli arhitektuuri ja hüperparameetrite puhul ei ole oodata. Mudeli üldine piksli tasemel täpsus on 75.8%.

Segmenteerimismudeli tulemused klasside lõikes (Tabel 7.) näitavad kõrgemat täpsust metsade (86.7%), sisevete (80.5%) ja põllumaa (75.7%) puhul, millele järgnevad sisemaa märgalad (46.5%) ja asustatud alad (41.6%). Kokku õnnestus mudelil õigesti klassifitseerida 11 erineva klassi puhul (kõige madalam täpsus 13.4% kaevandused, prügilad ja ehitusplatsid). Kolme klassi puhul mudel eksis segmenteerimisega täielikult (IoU skoor 0). Valideerimise andmekogust oli puudu üks klass - mereäärsed märgalad, mis eesti aladel on



Joonis 23. Mudeli treenimise progress (U-Net), klassifikatsiooni 2. tase.

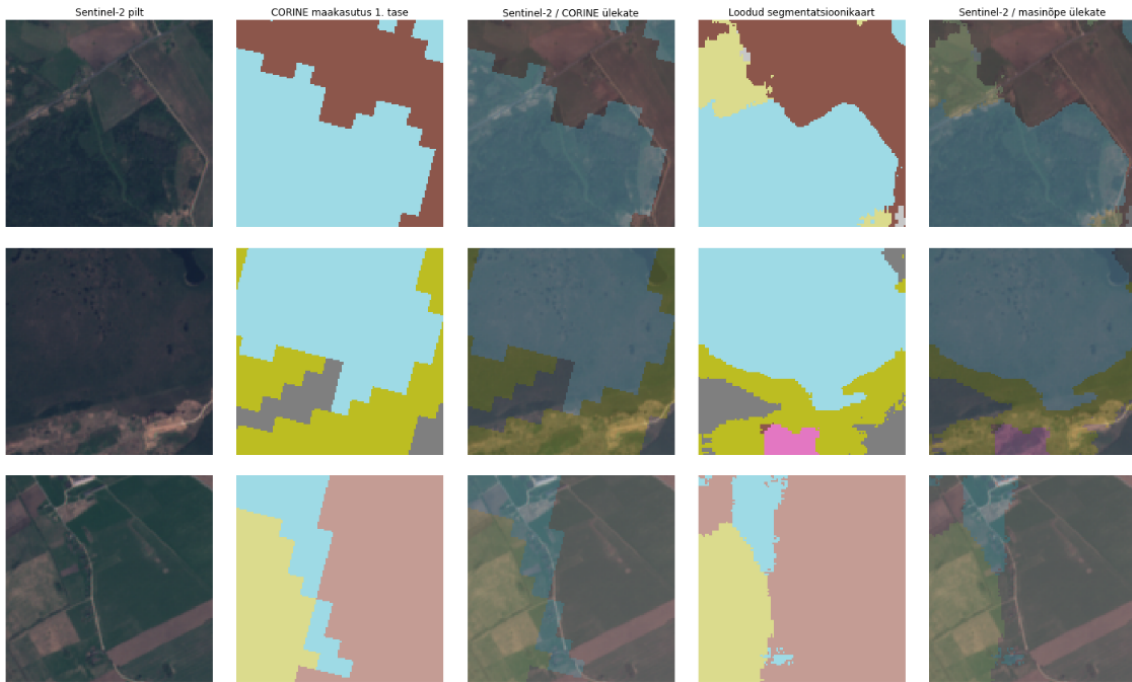
esindatud ainult rannasoolakutena ja mis ei olnud valideerimisandmestikus esindatud.

Tabel 7. Segmenteerimise täpsus maakasutuse 2. taseme klasside lõikes.

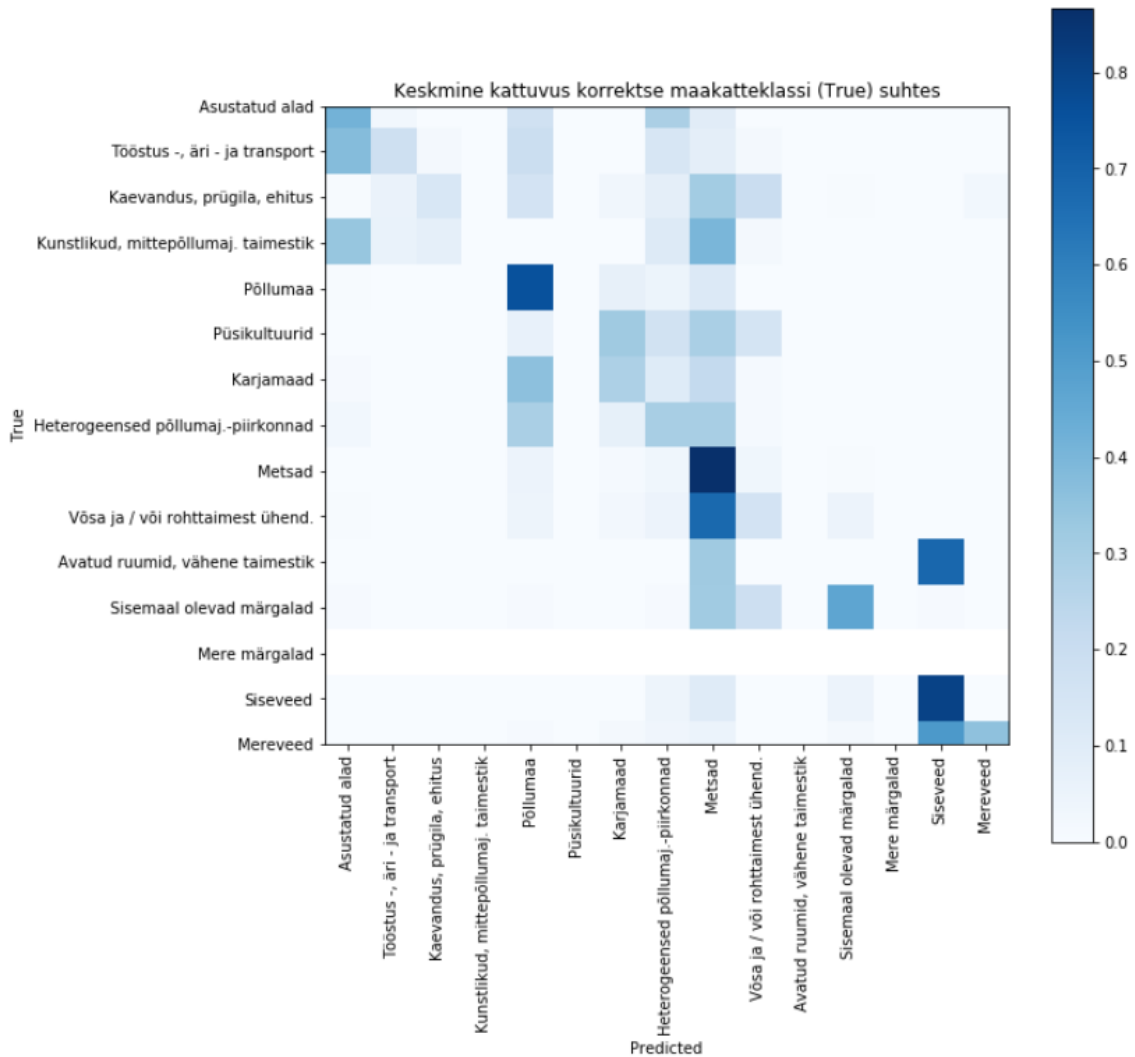
Maakasutuse klass	IoU skoor
Metsad	0.867
Siseveed	0.805
Põllumaa	0.757
Sisemaal olevad märgalad	0.465
Asustatud alad	0.416
Mereveed	0.357
Heterogeensed põllumajanduspiirkonnad	0.298
Karjamaad	0.282
Tööstus -, äri - ja transpordialad	0.179
Võsa ja / või rohttaimestiku ühendused	0.156
Kaevandused, prügilad ja ehitusplatsid	0.134
Kunstlikud, mittepõllumajanduslikud taimestikuga alad	0
Püsimatid	0
Avatud ruumid, kus taimestik on vähene või puudub üldse	0
Mere märgalad	-

Näited segmenteerimismudeli tulemustest Joonisel 24 kõrvutatavad segmenteerimismudeli tulemused (neljandas tulbas) sisendandmeteks kasutatud maakasutuskaardiga (teises tulbas). Välja toodud näidetes on võimalik ära tunda sarnaseid kontuure kahe kaardi vahel ja on võimalik näha masinõppe mudeli poolt tehtud eralduste kõrgemat täpsust.

Eksimismatriks Joonisel 25 näitab, et mudelil on kõrgem täpsus haritava maa, metsade ja siseveekogude segmenteerimisel. Mudel eksib selgelt näiteks tehilike mittepõllumajanduslike taimkattega alade ja avatud vähese või olematu taimkattega alade tuvastamisel.



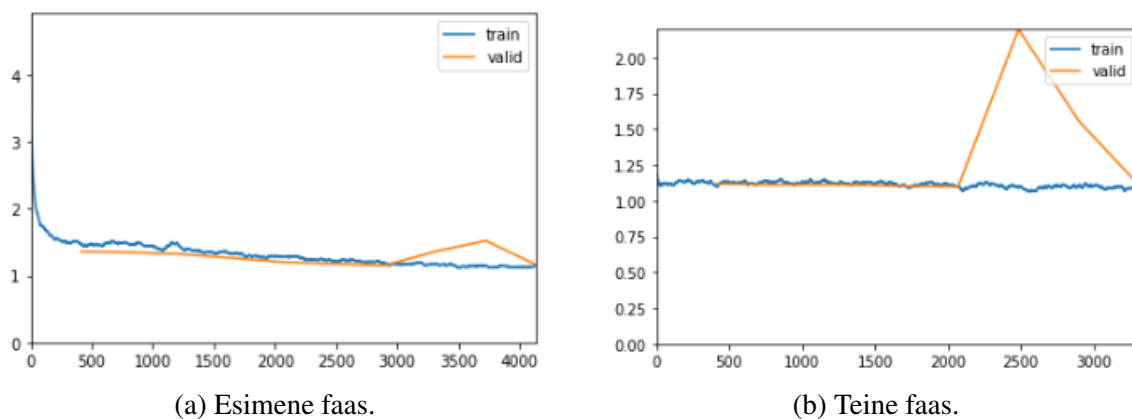
Joonis 24. Segmenteerimismudeli tulemuste näide, maakasutuse 2. tase.



Joonis 25. Segmenteerimise eksimismatriksi maakasutuse 2. taseme klasside lõikes.

4.3.3 Maakasutuse klassifikatsiooni 3. tase

Mudelit treeniti kokku 18 epohhi (10+8), mille jooksul jõudis mudel stabiilsesse seisusse. Joonis 26 näitab mudeli kadu treening- ja validatsiooniandmestiku puhul treenimise protsessi vältel esimeses ja teises faasis. Graafikult on näha, et mudeli kadu on oma miinimumile lähenemas ning suuremat täpsuse kasvu valitud mudeli arhitektuuri ja hüperparameetrite puhul ei ole oodata. Mudeli üldine piksli tasemel täpsus on 59.7%. Treenimise protsessi graafik teises faasis näitab, et valitud mudel on optimaalse tulemuse lähedal, mudel hakkab näitas valideerimiskao suurenemist 6. ja 7. epohhi puhul.



Joonis 26. Mudeli treenimise progress (U-Net), klassifikatsiooni 3. tase.

Segmenteerimismudeli tulemused klasside lõikes (Tabel 8.) näitavad üle 78.7% täpsust haritava maa puhul, millele järgnevad veekogud (65.6%), meri (63.8%) ja segametsad (61.9%). Kokku õnnestus mudelil õigesti klassifitseerida (kõige madalam täpsus 1.1% kompleksmaaviljelus) 16 erineva klassi puhul. 8 klassi puhul mudel segmenteerimisega eksis täielikult (IoU skoor 0). Juhuslikkuse alusel valideerimise andmekogust oli puudu sadamate klass, mis on ka kogu andmestikus nõrgalt esindatud.

Tabel 8. Segmenteerimise täpsus maakasutuse 3. taseme klasside lõikes.

Maakatte tüüp	IoU skoor
Niisutuseta haritav maa	0.787
Veekogud	0.657
Meri ja ookean	0.638
Segametsad	0.619
Turbaraba	0.511
Okasmetsad	0.490
Hõredalt hoonestatud alad	0.484
Karjamaad	0.285

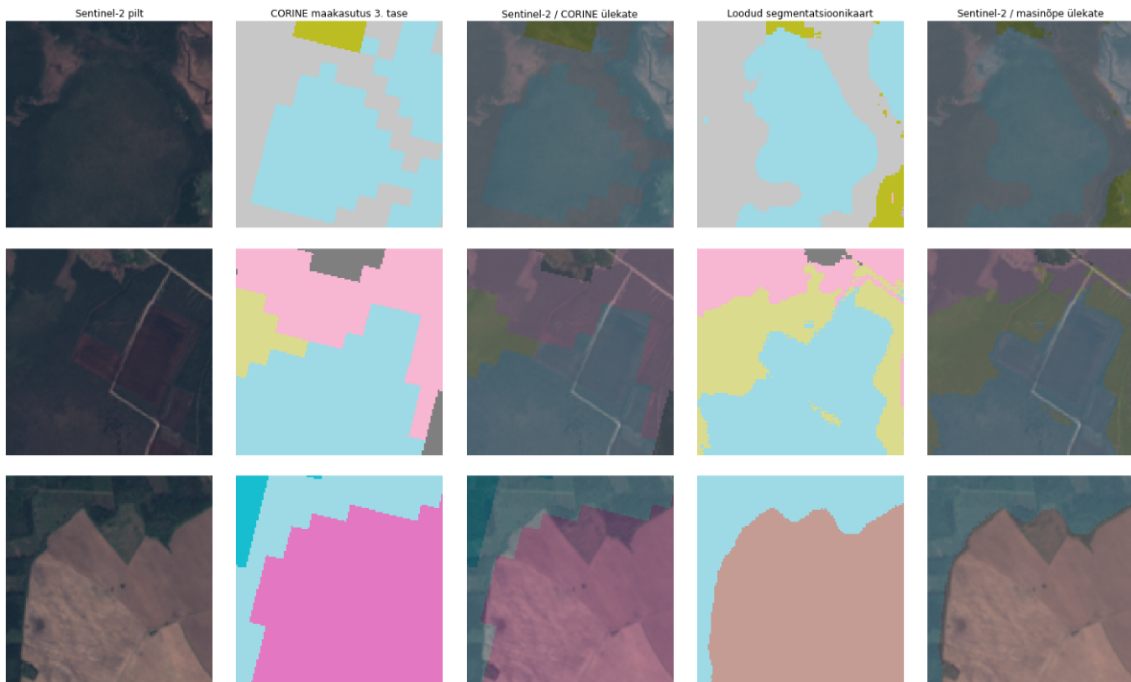
Jät kub...

Tabel 8 – Jät kub...

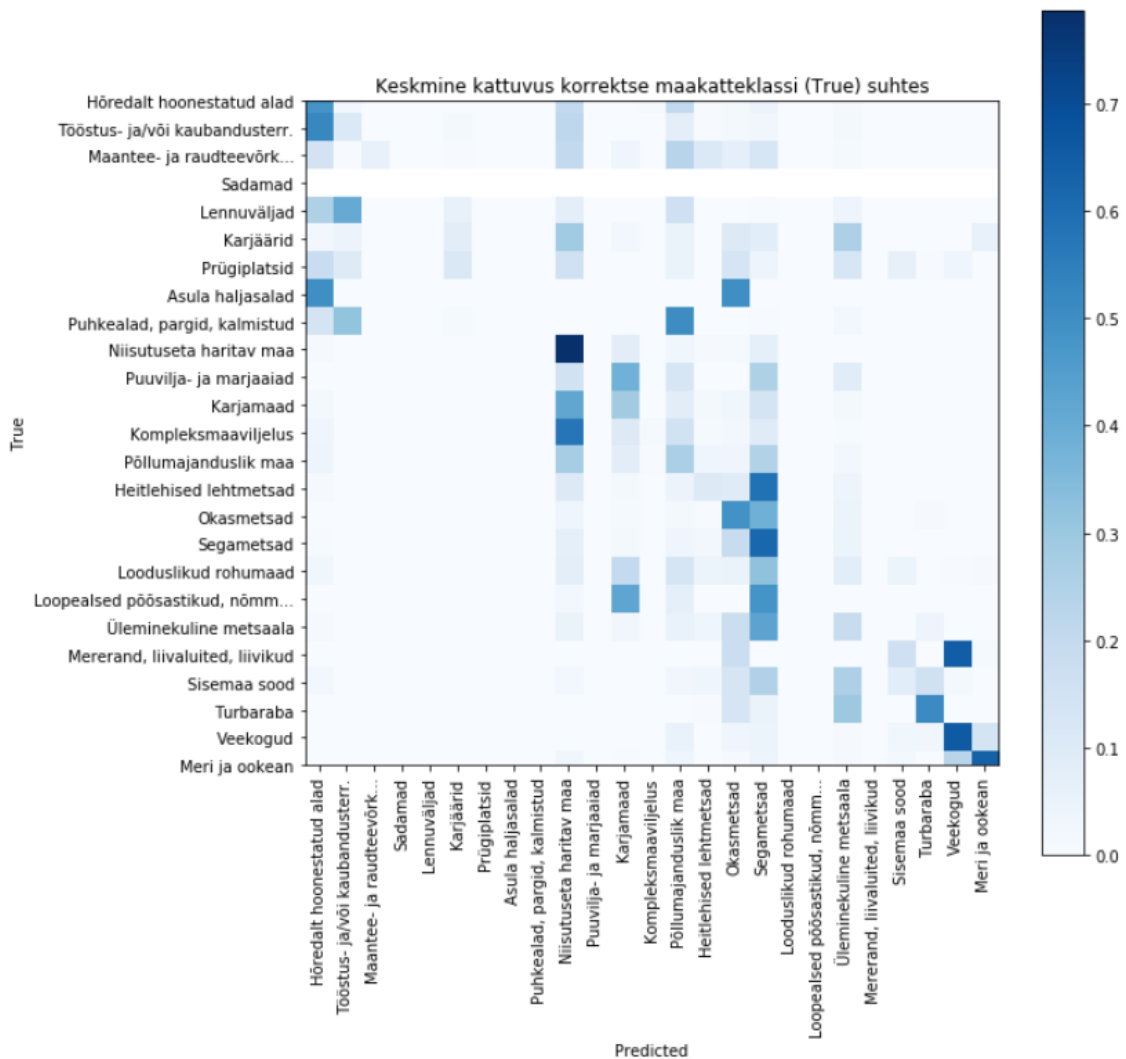
Maakatte tüüp	IoU skoor
Põllumajanduslik maa (< 75%) loodusliku taimkatte osalusega	0.264
Üleminekuline metsaala	0.185
Tööstus- ja/või kaubandusterritooriumid	0.108
Heitlehised lehtmetsad	0.101
Sisemaa sood	0.086
Karjäärid	0.081
Maantee- ja raudteevõrk ja piirnev ala	0.056
Kompleksmaaviljelus (haritavat maad > 75%)	0.011
Prügiplatsid	0
Asula haljasalad	0
Looduslikud rohumaad	0
Loopealsed põõsastikud, nõmmraba ja nõmm	0
Mererand, liivaluited, liivikud	0
Puuvilja- ja marjaaiad	0
Lennuväljad	0
Puhkealad, pargid, kalmistud	0
Sadamad	-

Näited segmenteerimismudeli tulemustest Joonisel 27 kõrvutatavad segmenteerimismudeli tulemused (neljandas tulbas) sisendandmeteks kasutatud maakasutuskaardiga (teises tulbas). Välja toodud näidetes on võimalik ära tunda sarnaseid kontuure kahe kaardi vahel ja on võimalik näha masinõppe mudeli poolt tehtud eralduste kõrgemat täpsust.

Eksimismaatriks Joonisel 28 näitab, et mudel teeb selgelt eksimusi klassifitseerides näiteks mereranna alad veekogudeks. Selle üheks põhjuseks saab lugeda maakasutuskaardi ebatäpsust mere ja ranna piirjoone jälgimisel. Silmapaistvad on veel ka näiteks lennuväljade klassifitseerimine tööstusalade sekka, asula haljasalade arvamine okasmetsaks / hõredalt hoonestatud alaks ja karjamaa haritavaks maaks pidamine. Taolised eristused võivad olla väga keerukad ka inimsilmale, kui vaadelda 120x120 piksli suuruse pildina ja ilma ümbritsevat konteksti nägemata.



Joonis 27. Segmenteerimismudeli tulemuste näide, maakasutuse 3. tase.



Joonis 28. Segmenteerimise eksimismatriks maakasutuse 3. taseme klasside lõikes.

4.4 Tulemuste analüüs

Saadud tulemustest tuli esmalt välja kahe kasutatava andmekogu keerukuse kasv klassifikatsioonitaseme suurenemisel. Esimesel tasemel on mudelitel vaja teha tunduvalt vähem eristusi ja ühel pildil on vähem erinevaid klasse. Arusaadavalt hierarhilise klassifikatsioonisüsteemi puhul see keerukus kasvab ja seda on näha ka andmete kardinaalsuse ja tiheduse näitajates.

Mudelite treenimise protsess oli kõigi kolme täpsusastme juures sarnane. Treenimise protsessi klassifikatsioonimudeli puhul alustati eeltreenimata ResNet50 struktuuriga mudelitega. Mudeleid treeniti vahemikus 10-15 epohhi, mis oli piisav, et mudelid viia stabiilsesse positsiooni, millest edasine oli areng väike.

Segmentatsioonimudeli treenimise puhul rakendati esimeses protsessis eeltreenitud mudeleid. Esmalt võeti U-Net mudeli kodeerija pooleks klassifikatsioonimudel ning treenimine toimus ainult dekodeerija mudeli poolel. Peale esimest sammu avati kogu mudel treenimiseks järgnevate epohhide vältel.

Loodud klassifikatsioonimudelid näitasid kõrgeid tulemusi kõigil kolmel tasemel, kuid selgelt on näha, et klasside arvu kasvades kasvab ka nõrgemate klasside osakaal. Metsade, põllumaade ja veekogude puhul olid tulemused kõige paremad üle kõigi kolme taseme. Väga hästi oli mudel isegi võimeline tuvastama erinevaid metsatüüpe.

Kui vaadelda klassifitseerimismudeli suuremaid eksimusi individuaalsete piltide tasemel, siis avalduvad ebatäpsused BigEarthNet andmekogus. Joonisel 29 on näha näiteid piltidest, mille pealkirjad on töö autori arvates ebatäpselt määratletud. Sellise ebatäpsuse üheks võimalikuks põhjuseks on CORINE maakasutuse kaardi madalam täpsus (100m) võrreldes Sentinel-2 piltide täpsusega (10m). Maakasutuskaardi madalama täpsuse tõttu on selgelt näha ka maakasutuse alade piirjoonte ebatäpsuseid, mis loovad masinõppe mudeli treenimiseks vigadega treeningandmeid. Näiteid sellistest ebatäpsustest on toodud Joonisel 30.

Mudelite täpsuste võrdluses üle kolme CORINE taseme on näha, et täpsus on kõrgeim esimese taseme puhul ja madalaim kolmanda taseme puhul. See on ühtlasi ka oodatud tulemus, sest sellises järjekorras kasvab ka liigituste arv (vastavalt 5, 15 ja 43 maapinna klassi taseme kohta). Kui esimesel tasemel on masinõppe mudelil vaja eristada tehislik maakate metsadest ja veekogudest, siis kolmandal tasemel on vaja teha palju täpsemaid eristusi näiteks lehtmetsade ja segametsade vahel.

Predicted: Agricultural areas
Actual: -- Artificial surfaces
Loss: 6.426020622253418
Probability: 0.890999972820282



Predicted: Broad-leaved forest
Actual: -- Continuous urban fabric
Loss: 7.653176784515381
Probability: 0.6650000214576721



Predicted: Coniferous forest
Actual: -- Continuous urban fabric
Loss: 6.864468574523926
Probability: 0.7829999923706055



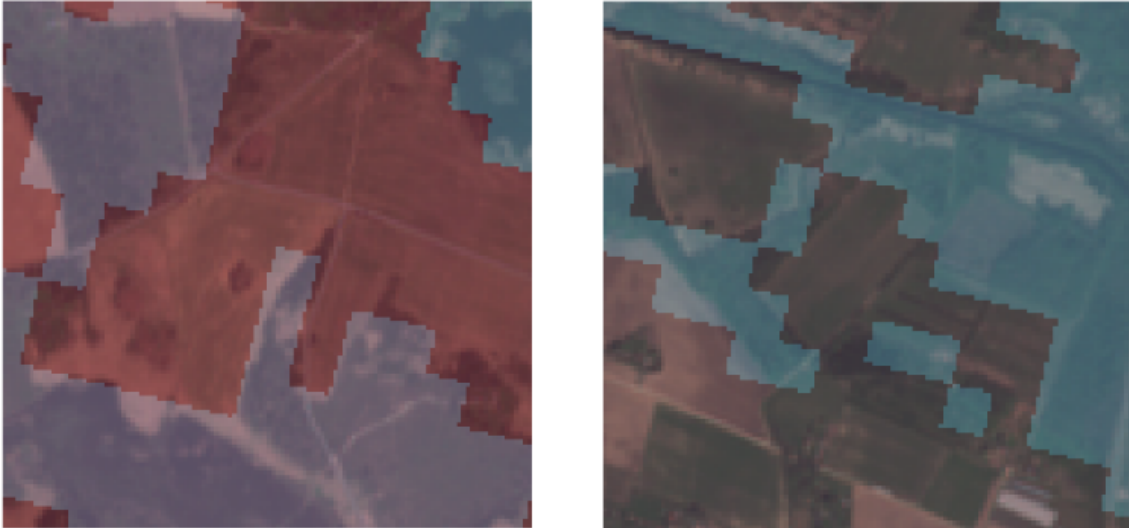
Predicted: Agro-forestry areas
Actual: -- Continuous urban fabric
Loss: 6.637006759643555
Probability: 0.5350000262260437



Joonis 29. Ebatäpsused BigEarthNet andmekogus.

Iga mudeli treenimise puhul jõudis mudel stabiilsesse positsiooni, millest seatud parameetritega enam suuremat edasiarengut ei olnud võimalik saavutada. Eeldatavalt on tulemust võimalik parandada veel enam läbi hüperparameetrite muutmise ja pikema treenimise, kuid antud probleemi analüüsimiseks oli saavutatud tulemus piisav.

Loodud masinõppe mudelid näitasid küll väga head tulemust suuremate maakasutuse klasside puhul (näiteks metsad, põllumaad ja veekogud), kuid tulemus nõrgenes mitmete väiksemate klasside puhul. Selle taga on ühelt poolt kindlasti see, et andmekogud ei ole tasakaalus. Tulemused aga näitavad, et see ei ole alati seoses. Teiseks oluliseks teguriks on kindlasti ka klasside vaheline visuaalne eristatavus. Mida rohkemate klasside peale liikuda, seda rohkem eristusi on mudelil vaja teha ja seda väiksemaks muutub visuaalne eristatavus



Joonis 30. Ebatäpsused maakasutuskaardil.

klasside vahel. Kolmandaks oluliseks teguriks on kindlasti veel ka andmekogude enda ebatäpsused. CORINE maakasutuskaardi koostamisel on nõutavaks täpsuseks 85% ja sellest tulenevad nähtavad ebatäpsused on selgesti eristatavad kui kaart kokku viia palju teravama satelliidifotoga. Kui aga masinõppe mudel õpib tundma satelliidifotot ja sellel olevat maapinda, siis võib nähtav mudeli viga tekkida ka just sellest andmekogu enda ebatäpsusest, mille vastu mudelit hinnatakse.

Näiteid segmenteerimismudeli tulemustest kolmel klassifikatsioonitasemel on välja toodud Lisas 4. Nende piltide peal võib mõnel juhul selgelt ka näha masinõppe mudeli loodud selgemaid piirjooni klasside vahel, mis vaatluses jooksevad kokku satelliidifoto peal olevate piirjoontega.

Töös läbitud protsess ja mudeli arhitektuurid võivad väga hästi sobida näiteks metsakatet jälgiva automatiseeritud süsteemi loomiseks, mis võimaldaks senisest kõrgema täpsuse ja väikse ajalise intervalliga saada ülevaadet metsakatte kohta ja kaardistada selle muutuseid. Metsade jälgimine on ühtlasi ka üks ÜRO säästva arengu eesmärkidest, mida globaalsel tasemel jälgitakse. Soovituslikke samme selles suunas on toodud ka järgnevas peatükis.

Töö esimese etapina loodud ResNet klassifikatsioonimudelid on ühtlasi ka hea alguspunkt teiste satelliidifotodel põhinevate masinõppe mudelite loomiseks, näiteks pildituvastuse vallas. Klassifikatsioonimudel on võimalik võtta uue lahenduse puhul mudeli treenimise alguspunktiks, kuna see on juba eelnevalt satelliidifotode lugemise peale treenitud. Sõltuvalt lahendatavast probleemist võib seega olla piisav ka ainult mudeli viimaste kihtide välja vahetamine ja nende üle treenimine, jättes mudeli esimese poole kihid puutumata.

4.5 Teemad edasiseks uurimistööks

Magistritöö käigus avaldusid mitmed võimalused loodava lahenduse edasi arendamiseks ja mudelite täpsuse tõstmiseks. Selles alapeatükis on välja toodud võimalikke suundasid edasiseks uurimistööks.

- Täpsem aluskaart treenimiseks. CORINE maakasutuse kaardi täpsus jääb siinkohal peamiseks kitsaskohaks. Võimalik on ära kasutada eelteenitud mudeleid, tänu millele ei ole vaja suuremahulist andmekogu täpsema segmenteerimismudeli loomiseks.
- Käesolevas töös kasutati andmekogudes vaid Sentinel-2 R,G ja B kanaleid, kuid kõikide kanalite kasutamine annab masinõppe mudelile lisainformatsiooni ja võimaldab tõsta mudeli täpsust veelgi.
- Lisaks Sentinel-2 andmetele on üheks potentsiaalseks edasiarenduse suunaks täiendavate andmekogude kasutamine, näiteks Sentinel-1 radarsatelliidi andmed.
- Satelliidifotode üheks oluliseks omaduseks on nende ajaline mõõde, võimalus võrrelda tulemusi läbi aja. Selle abil on võimalik kinnistada mudeli pakutud maakasutuskaardi jaotuseid ja hakata välja tooma maakasutuse muutuseid.
- BigEarthNet andmestik on koostatud automatiseeritud viisil, kombineerides Sentinel-2 ja CORINE andmestikud. Magistritöö tulemustes oli näha, et selles andmestikus esineb puudusi. Masinõppe abil on võimalik muuta andmestikku täpsemaks, vajadusel pildi pealkirju korrigeerides.
- Töö tulemustest on näha osade klasside puhul suuremat täpsust, näiteks metsade tuvastamisel. Võimalikuks edasiarenduseks on katsetada vaid ühe klassi keskseid mudeleid (näiteks binaarne segmenteerimisülesanne kahe klassiga: Mets, Muu).
- Võimalus luua mudel, mis kasutaks ära CORINE hierarhilist klassifikatsiooni struktuuri. Esimese taseme täpsem mudel annab väärtusliku sisendi järgmiste tasemete jaotusesse. Näiteks kui mudel esimeses tasemel suudab täpselt tuvastada metsa, siis kolmandal tasemel saaks mudel keskenduda vaid korrektse metsatüübi valimisele, mitte otsustama võrdselt kõikide klasside vahel.
- Magistritöö kirjutamise ajal väljastati ka CORINE ülemaailmne maakasutuse kaart, mis on koostatud masinõppe abil ja kasutab lihtsustatud klassifikatsioonijaotust võrreldes töös kasutatud 43-klassilise jaotusega. Võimalus on kasutada seda uut andmestikku ja klassifikatsiooni jaotust, ning analüüsida võimalusi selle madala, 100m resolutsiooni tõstmiseks.

Kokkuvõte

Käesoleva magistritöö eesmärgiks oli masinõppe meetoditel põhinevate algoritmide loomine satelliidifotode abil maapinna klassifitseerimiseks ja loodud lahenduse analüüsimine maakasutuskaartide täpsuse tõstmisel ja maapinnal toimuvate muutuste jälgimisel.

Masinõppe, tehisintellekt ja andmeteadus on kosmosetehnoloogia sektori kümne kõige rohkem rahastatud valdkonna seas ja hoogsalt kasvav kosmosetööstus kaasas 2019. aasta jooksul prognoositavalt 2 miljardit dollarit riskikapitali rahastust. Tehnoloogilised uuendused kinnistavad üha rohkem ennast ka keskkonnakaitse vallas. ÜRO säästva arengu eesmärgid on seadnud konkreetsete sihid jätkusuutliku elukeskkonna loomiseks ja maapinnal toimuvate muudatuste jälgimiseks, milleks satelliidiandmetel ja masinõppel on suur potentsiaal.

Töö teoreetilises osas tutvustati maapinna vaatluse tausta ja sellega seonduvaid andmekogusid, lisaks loodi ülevaade töös kasutatavatest masinõppe meetoditest.

Töö käigus loodi konvolutsioonilistel meetoditel põhinevad masinõppe mudelid satelliidifotode klassifitseerimiseks ja piksli tasemel segmenteerimiseks kasutades CORINE maakasutuse kategooriaid. Töös kasutati klassifikatsiooniülesande lahendamiseks BigEarthNet andmekogu, mis on esimene suuremahuline satelliidifotodest koosnev masinõppe andmekogu. Lisaks loodi töö käigus ka teine andmekogu segmenteerimisülesande lahendamiseks Sentinel-2 satelliidifotode ja CORINE maakasutuskaartide põhjal.

Töö tulemustest järeldub, et konvolutsioonilistel närvivõrkudel põhineva automatiseeritud maakasutuskaardi täpsuse suurendamiseks on oluline valida väiksem ja teineteisest selgelt eralduv maakasutuse klasside nimekiri. Töö käigus loodud väiksema klasside arvuga mudelid näitasid märgatavalt paremaid tulemusi ja väikest eksimuste osakaalu. Lisaks toodi töös välja meetodid mudeli tulemuste analüüsimiseks ja hindamiseks, kuna maakasutuse pindalad on tugevalt tasakaalust väljas, tuleb vaadelda iga klassi tulemusi eraldiseisvalt.

Tulemused näitasid, et U-Net arhitektuuril põhinevat masinõppe mudelit on isegi võimalik rakendada olemasolevate maakasutuskaartide täpsuse tõstmiseks. Loodud mudel kasutas olemasolevat kaarti sisendina ja suutis oma väljundis näidata eri klasside vahel täpsemaid

piirjooni kui selle treenimiseks kasutatud andmekogus.

Rakendatud andmekogude puhul töid loodud masinõppe mudelid esile mitmeid puuduseid, mis avaldavad mõju mudelite täpsusele. BigEarthNet andmekogu puhul tuli välja valesti pealkirjastatud pilte ja CORINE maakasutuskaardi ebatäpsuse taga on selle madal resolutsioon ja alla 25ha suuruse pindalaga maa-alade välistamine.

Magistritöö püstitatud eesmärgid said täidetud ja loodud masinõppe mudelid tõestasid tehisnärvivõrkude võimekust probleemi lahendamisel. Lisaks tuli töö käigus välja ka mitmeid võimalusi käsitletud lahenduse parandamiseks, millest tulenevalt tõi töö autor välja teemasid edasiseks uurimistööks ja lahenduse täpsuse tõstmiseks.

Kasutatud kirjandus

- [1] *Sustainable Development Goal indicators website*. URL: <https://unstats.un.org/sdgs/>.
- [2] Statistikaamet. *Säästva arengu näitajad*. 2018. URL: https://www.stat.ee/valjaanne-2018_saastva-arengu-naitajad.
- [3] *COPERNICUS IN SUPPORT OF THE UN SUSTAINABLE DEVELOPMENT GOALS*. Tech. rep. 2018.
- [4] Alessandro De Concini and Jaroslav Toth. *The future of the European space sector*. Tech. rep. 2019.
- [5] Ian Agar. *To the moon and back: The top VC investors in space tech* | *PitchBook*. 2019. URL: <https://pitchbook.com/news/articles/to-the-moon-and-back-the-top-vc-investors-in-space-tech>.
- [6] *SpaceX plans to put more than 40,000 satellites in space* | *New Scientist*. URL: <https://www.newscientist.com/article/2220346-spacex-plans-to-put-more-than-40000-satellites-in-space/>.
- [7] *Planet Monitoring* | *Planet*. URL: <https://www.planet.com/products/monitoring/>.
- [8] M. Drusch et al. “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services”. In: *Remote Sensing of Environment* (2012). ISSN: 00344257. DOI: 10.1016/j.rse.2011.11.026.
- [9] *MSI Instrument – Sentinel-2 MSI Technical Guide – Sentinel Online*. URL: <https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>.
- [10] *ICEYE Solutions*. URL: <https://www.iceye.com/solutions>.
- [11] George Büttner et al. *THE CORINE LAND COVER 2000 PROJECT*. Tech. rep. 2004. URL: <http://terrestrial.eionet.eu.int>.
- [12] Barbara Kosztra György Büttner and Gerard Hazeu Stephan Arnold. *Updated CLC illustrated nomenclature guidelines*. Tech. rep. 2019.
- [13] Valentina Sagris and Anneli Palo. “CORINE Land Cover 2018 projekt – fookuses on metsad”. In: *TARTU ÜLIKOOL TARTU OBSERVATOORIUM KAUGSEIRE EESTIS 2018 ARTIKLIKOGUMIK* (2018), pp. 59–68.

- [14] *CORINE LAND COVER 2012 FINAL VALIDATION REPORT*. Tech. rep. 2017. URL: <https://land.copernicus.eu/user-corner/technical-library/clc-2012-validation-report-1>.
- [15] Jesús Balado et al. “Automatic CORINE land cover classification from airborne LIDAR data”. In: *Procedia Computer Science*. Vol. 126. Elsevier B.V., 2018, pp. 186–194. DOI: 10.1016/j.procs.2018.07.222.
- [16] Claudia Paris, Lorenzo Bruzzone, and Diego Fernandez-Prieto. “A Novel Approach to the Unsupervised Update of Land-Cover Maps by Classification of Time Series of Multispectral Images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.7 (July 2019), pp. 4259–4277. ISSN: 15580644. DOI: 10.1109/TGRS.2018.2890404.
- [17] Marcel Buchhorn et al. *Copernicus Global Land Service: Land Cover 100m: Collection 2: epoch 2015*. 2019. DOI: 10.5281/zenodo.3243509. URL: <https://doi.org/10.5281/zenodo.3243509>.
- [18] Ethem Alpaydin. *Introduction to Machine Learning*. 2014. URL: <https://ieeexplore.ieee.org/book/6895440>.
- [19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [20] Vincent Dumoulin and Francesco Visin. “A guide to convolution arithmetic for deep learning”. In: (Mar. 2016). URL: <http://arxiv.org/abs/1603.07285>.
- [21] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2016-December. IEEE Computer Society, Dec. 2016, pp. 770–778. ISBN: 9781467388504. DOI: 10.1109/CVPR.2016.90.
- [22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 9351. Springer Verlag, 2015, pp. 234–241. ISBN: 9783319245737. DOI: 10.1007/978-3-319-24574-4_{_}28.
- [23] Chuanqi Tan et al. “A Survey on Deep Transfer Learning”. In: (Aug. 2018). URL: <http://arxiv.org/abs/1808.01974>.
- [24] Gabriel J. Brostow, Julien Fauqueur, and Roberto Cipolla. “Semantic object classes in video: A high-definition ground truth database”. In: *Pattern Recognition Letters* 30.2 (Jan. 2009), pp. 88–97. ISSN: 01678655. DOI: 10.1016/j.patrec.2008.04.005.

- [25] Swarnendu Ghosh et al. “Understanding Deep Learning Techniques for Image Segmentation”. In: (July 2019). URL: <http://arxiv.org/abs/1907.06119>.
- [26] Gencer Sumbul et al. “BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding”. In: (Feb. 2019). URL: <http://arxiv.org/abs/1902.06148>.
- [27] *ESTHub riiklik satelliidiandmete keskus*. URL: <https://ehdatahub.maaamet.ee/dhus/#/home>.
- [28] *CLC 2018 — Copernicus Land Monitoring Service*. URL: <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>.
- [29] *bigearthnet | TensorFlow Datasets*. URL: <https://www.tensorflow.org/datasets/catalog/bigearthnet>.
- [30] *BigEarthNet*. URL: <http://bigearth.net/>.
- [31] Leslie N. Smith. “Cyclical Learning Rates for Training Neural Networks”. In: (June 2015). URL: <http://arxiv.org/abs/1506.01186>.
- [32] Leslie N. Smith. “A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay”. In: (Mar. 2018). URL: <http://arxiv.org/abs/1803.09820>.
- [33] Flavia Cristina Bernardini et al. *Learning and Nonlinear Models-Revista da Sociedade Brasileira de Redes Neurais CARDINALITY AND DENSITY MEASURES AND THEIR INFLUENCE TO MULTI-LABEL LEARNING METHODS*. Tech. rep. 2009. URL: <http://mulan.sourceforge.net..>
- [34] Flavia Cristina Bernardini, Rodrigo Barbosa Da Silva, and Edwin Mitacc Meza. “Analyzing the Influence of Cardinality and Density Characteristics on Multi-Label Learning”. In: *ENIAC*. 2013. URL: <http://mulan.sourceforge.net..>
- [35] David Martin Ward Powers. “EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION”. In: *Journal of Machine Learning Technologies* 2.1 (2011), pp. 37–63. DOI: 10.9735/2229-3981. URL: <http://www.bioinfo.in/contents.php?id=51>.
- [36] Sorour Mohajerani and Parvaneh Saeedi. “Cloud-Net: An end-to-end Cloud Detection Algorithm for Landsat 8 Imagery”. In: (Jan. 2019). URL: <http://arxiv.org/abs/1901.10077>.

Summary

The goal of this master's thesis was to create machine learning models for classification and segmentation of satellite imagery with the aim of improving existing land use maps and land use change detection.

Currently, machine learning, AI and data science are among the top 10 funded fields in space technology and the actively growing space industry incorporated an estimated 2 billion USD of venture capital in 2019. Advanced technologies are also taking an important role in supporting environmental causes. The UN Sustainable Development Goals have set out strict aims for environment sustainability and monitoring changes on our planet, for which remote earth observation and machine learning can be applied.

In the theoretical section of this work the concepts of earth observation were discussed, together with an overview of the used datasets and machine learning methods.

A set of classification and segmentation models were created for satellite image classification and pixel level segmentation according to a three-level land use category system used by the CORINE programme. A novel BigEarthNet dataset was used, which is the first public satellite imagery dataset for machine learning application. Also, a new dataset was composed for training segmentation models, combining a Sentinel-2 image of Estonia with the CORINE land use map.

The results of this thesis show the possibilities of using a convolutional neural network for this sort of task and underline the need for changing the land use categories for achieving better accuracy. Also, as land use is severely unbalanced, there is a need for class based analysis and accuracy measurement, which was used in this work.

As an important additional result the U-Net models showed a capability of improving on the existing low resolution of land cover maps. The models used existing map data as input and managed to offer up better land cover mappings than the data they were trained on.

The use of BigEarthNet and CORINE land use datasets for machine learning highlighted some noise in the data, which affects the results. In BigEarthNet some images seem to

be mislabelled, with land cover maps a limitation is its 100m resolution and the fact that distinctive areas under 25ha are not included on the map.

The goals of this master's thesis were achieved and a total of six convolutional neural networks were created in order to analyse land use classification and segmentation on three classification levels set by CORINE land use mapping. In addition, several ideas for the improvement of the results and further research were proposed.

Lisa 1 - BigEarthNet klassid ja piltide arv

Tüübi kood	Maakatte tüüp	Piltide arv
313	Mixed forest	217 119
312	Coniferous forest	211 703
211	Non-irrigated arable land	196 695
324	Transitional woodland/shrub	173 506
311	Broad-leaved forest	150 944
243	Land principally occupied by agriculture, with significant areas of natural vegetation	147 095
242	Complex cultivation patterns	107 786
231	Pastures	103 554
512	Water bodies	83 811
523	Sea and ocean	81 612
112	Discontinuous urban fabric	69 872
244	Agro-forestry areas	30 674
412	Peatbogs	23 207
212	Permanently irrigated land	13 589
121	Industrial or commercial units	12 895
321	Natural grassland	12 835
223	Olive groves	12 538
323	Sclerophyllous vegetation	11 241
111	Continuous urban fabric	10 784
511	Water courses	10 572
221	Vineyards	9 567
241	Annual crops associated with permanent crops	7 022
411	Inland marshes	6 236
322	Moors and heathland	5 890
142	Sport and leisure facilities	5 353
222	Fruit trees and berry plantations	4 754
131	Mineral extraction sites	4 618
213	Rice fields	3 793

Jätkub...

Tabel 9 – Jät kub...

Tüübi kood	Maakatte tüüp	Piltide arv
122	Road and rail networks and associated land	3 384
332	Bare rock	3 277
141	Green urban areas	1 786
331	Beaches, dunes, sands	1 578
333	Sparsely vegetated areas	1 563
421	Salt marshes	1 562
521	Coastal lagoons	1 498
133	Construction sites	1 174
522	Estuaries	1 086
423	Intertidal flats	1 003
124	Airports	979
132	Dump sites	959
123	Port areas	509
422	Salines	424
334	Burnt areas	328

Lisa 2 - Klassifikatsioonimudelite eksimismaatriksid (2. ja 3. tase)

Põllumaa

		Pakutud	
		Ei	Jah
Korrektne	Ei	60 064	5 048
	Jah	5 893	32 862

Kunstlikud, mittepõllumaj..

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 517	51
	Jah	1 153	146

Avatud ruumid, vähene taimeistik

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 487	181
	Jah	675	524

Metsad

		Pakutud	
		Ei	Jah
Korrektne	Ei	35 299	4 204
	Jah	4 047	60 317

Heterogeensed põllumaj..

		Pakutud	
		Ei	Jah
Korrektne	Ei	54 404	6 792
	Jah	9 778	32 893

Püsikultuurid

		Pakutud	
		Ei	Jah
Korrektne	Ei	98 183	716
	Jah	2 840	2 128

Tööstus -, äri - ja transpordialad

		Pakutud	
		Ei	Jah
Korrektne	Ei	100 307	510
	Jah	1 699	1 351

Siseveed

		Pakutud	
		Ei	Jah
Korrektne	Ei	89 175	1 150
	Jah	2 904	10 638

Asustatud alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	86 572	2 177
	Jah	4 196	10 922

Sisemaal olevad märgalad

		Pakutud	
		Ei	Jah
Korrektne	Ei	99 065	378
	Jah	2 444	1 980

Mereveed

		Pakutud	
		Ei	Jah
Korrektne	Ei	88 577	207
	Jah	323	14 760

Võsa / rohttaimestiku ühendused

		Pakutud	
		Ei	Jah
Korrektne	Ei	65 893	6 154
	Jah	12 941	18 879

Mere märgalad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 376	36
	Jah	263	192

Kaevandused, prügilad, ehitusp..

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 627	43
	Jah	1 009	188

Karjamaad

		Pakutud	
		Ei	Jah
Korrektne	Ei	81 283	2 717
	Jah	6 498	13 369

Klassifikatsioonimudeli eksimismatriksid 2. tasemel.

Agrometsandus

		Pakutud	
		Ei	Jah
Korrektne	Ei	96 681	1 121
	Jah	1 428	4 637

Lennuväljad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 694	1
	Jah	171	1

Soolakud

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 765	12
	Jah	51	39

Püsikult. seotud 1a põllukult.

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 233	226
	Jah	1 045	363

Kivipaljand

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 214	104
	Jah	263	286

Sklerofiilne taimestik

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 468	181
	Jah	1 368	850

Mererand, liivaluited, liivikud

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 463	45
	Jah	153	206

Heitlehised lehtmetsad

		Pakutud	
		Ei	Jah
Korrektne	Ei	70 620	4 994
	Jah	8 036	20 217

Hõreda taimkattega alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 606	13
	Jah	223	25

Põlengualad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 795	0
	Jah	72	0

Rannikulaguunid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 531	34
	Jah	85	217

Üleminekuline metsaala

		Pakutud	
		Ei	Jah
Korrektne	Ei	68 466	5 922
	Jah	13 281	16 198

Kompleksmaaviljelus

		Pakutud	
		Ei	Jah
Korrektne	Ei	78 519	4 400
	Jah	7 588	13 360

Okasmetsad

		Pakutud	
		Ei	Jah
Korrektne	Ei	66 368	4 734
	Jah	4 670	28 095

Veekogud

		Pakutud	
		Ei	Jah
Korrektne	Ei	91 167	994
	Jah	2 770	8 936

Klassifikatsioonimudeli eksimismatriksid 3. tasemel (1/3).

Ehitusplatsid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 620	0
	Jah	247	0

Tiheda hoonestusega alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 225	387
	Jah	743	1 512

Maantee- ja raudteevõrk

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 189	26
	Jah	613	39

Hõredalt hoonestatud alad

		Pakutud	
		Ei	Jah
Korrektne	Ei	88 763	1 866
	Jah	4 047	9 191

Prügiplatsid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 712	2
	Jah	143	10

Rannasoolakud

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 558	23
	Jah	193	93

Estuaarid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 600	44
	Jah	122	101

Puuvilja- ja marjaaiad

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 905	40
	Jah	772	150

Meri ja ookean

		Pakutud	
		Ei	Jah
Korrektne	Ei	89 078	186
	Jah	317	14 286

Asula haljasalad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 533	15
	Jah	296	23

Tööstus- ja/või kaubandusterr.

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 089	368
	Jah	1 461	949

Puhkealad, pargid, kalmistud

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 839	21
	Jah	938	69

Sisemaa sood

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 767	21
	Jah	974	105

Mõõnapaljandid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 660	14
	Jah	130	63

Veiniistandused

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 836	142
	Jah	1 419	470

Klassifikatsioonimudeli eksimismatriksid 3. tasemel (3/3).

Põllumajanduslik maa

		Pakutud	
		Ei	Jah
Korrektne	Ei	74 044	3 566
	Jah	13 229	13 028

Karjäärid

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 016	32
	Jah	708	111

Vooluveed

		Pakutud	
		Ei	Jah
Korrektne	Jah	101 627	302
		679	1 259

Segametsad

		Pakutud	
		Ei	Jah
Korrektne	Ei	61 644	6 843
	Jah	6 427	28 953

Loopealsed põõsastikud, nõmm

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 754	63
	Jah	838	212

Riisipõllud

		Pakutud	
		Ei	Jah
Korrektne	Ei	102 980	162
	Jah	373	352

Looduslikud rohumaad

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 419	222
	Jah	1 532	694

Niisutuseta haritav maa

		Pakutud	
		Ei	Jah
Korrektne	Ei	62 574	4 550
	Jah	6 623	30 120

Sadamad

		Pakutud	
		Ei	Jah
Korrektne	Ei	103 767	12
	Jah	62	26

Oliiviistandused

		Pakutud	
		Ei	Jah
Korrektne	Ei	101 110	268
	Jah	1 767	722

Karjamaad

		Pakutud	
		Ei	Jah
Korrektne	Ei	81 451	2 549
	Jah	6 782	13 085

Pidevalt niisutatud maa

		Pakutud	
		Ei	Jah
Korrektne	Ei	100 773	411
	Jah	1 332	1 351

Turbaraba

		Pakutud	
		Ei	Jah
Korrektne	Ei	100 151	353
	Jah	1 556	1 807

Klassifikatsioonimudeli eksimismatriksid 3. tasemel (3/3).

Lisa 3 - Klassifikatsioonimudeli näitajad klasside põhiselt, 3. tase

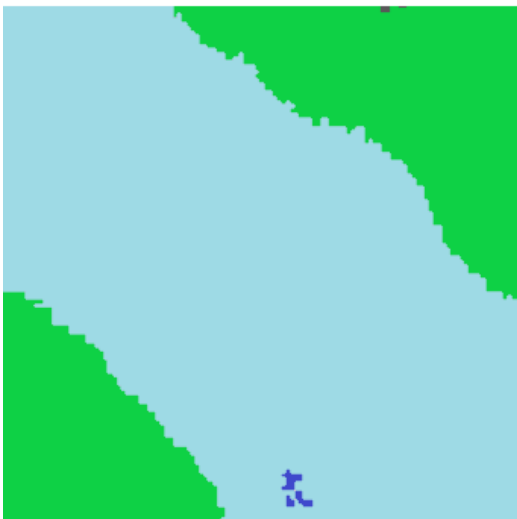
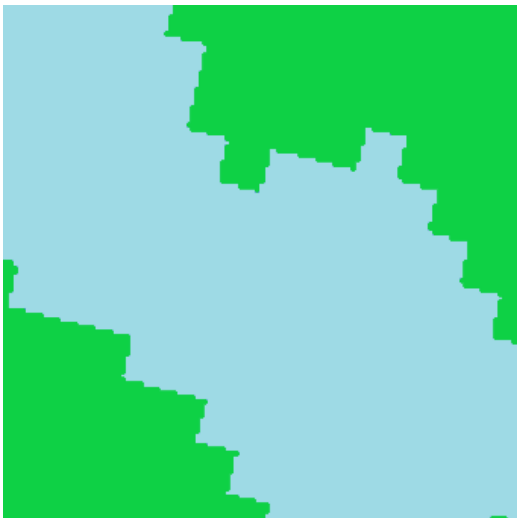
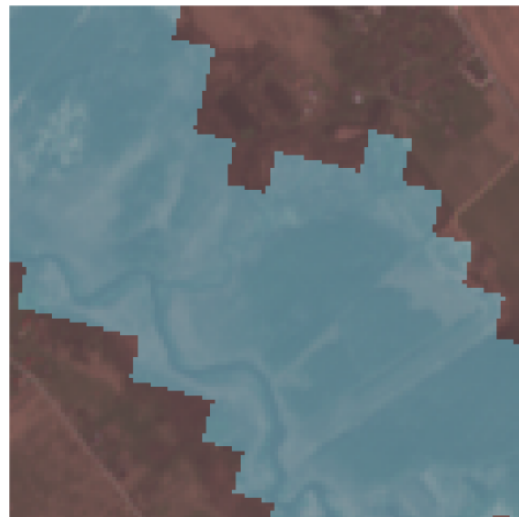
Klass	Täpsus	Saagis	F ₁	Pilte
Agrometsandus	0.81	0.76	0.78	6 065
Lennuväljad	0.50	0.01	0.01	172
Püsiluhtudega seotud 1a põllukultuurid	0.62	0.26	0.36	1 408
Kivipaljand	0.73	0.52	0.61	549
Mererand, liivaluited, liivikud	0.82	0.57	0.68	359
Heitlehised lehtmetsad	0.80	0.72	0.76	28 253
Põlengualad	0.00	0.00	0.00	72
Rannikulaguunid	0.86	0.72	0.78	302
Kompleksmaaviljelus (haritavat maad > 75%)	0.75	0.64	0.69	20 948
Okasmetsad	0.86	0.86	0.86	32 765
Ehitusplatsid	0.00	0.00	0.00	247
Tiheda hoonestusega alad	0.80	0.67	0.73	2 255
Hõredalt hoonestatud alad	0.83	0.69	0.76	13 238
Prügiplatsid	0.83	0.07	0.12	153
Estuaarid	0.70	0.45	0.55	223
Puuvilja- ja marjaaiad	0.79	0.16	0.27	922
Asula haljasalad	0.61	0.07	0.13	319
Tööstus- ja/või kaubandusterritooriumid	0.72	0.39	0.51	2 410
Sisemaa sood	0.83	0.10	0.17	1 079
Mõõnapaljandikud	0.82	0.33	0.47	193
Põllumajanduslik maa (< 75%) loodusliku taimkatte osalusega	0.79	0.50	0.61	26 257
Karjäärid	0.78	0.14	0.23	819
Segametsad	0.81	0.82	0.81	35 380
Loopealsed põõsastikud, nõmmraba ja nõmm	0.77	0.20	0.32	1 050
Looduslikud rohumaad	0.76	0.31	0.44	2 226
Niisutuseta haritav maa	0.87	0.82	0.84	36 743

Jätub...

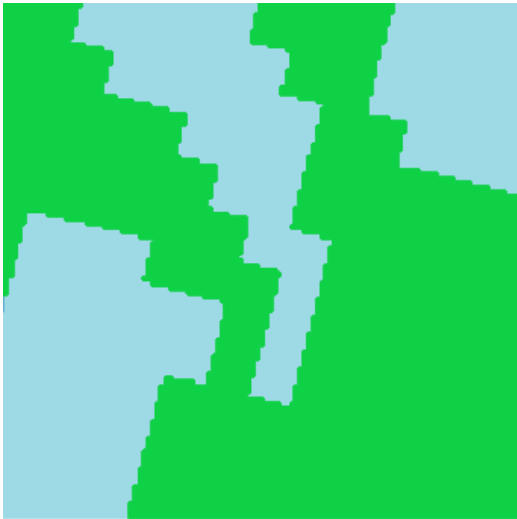
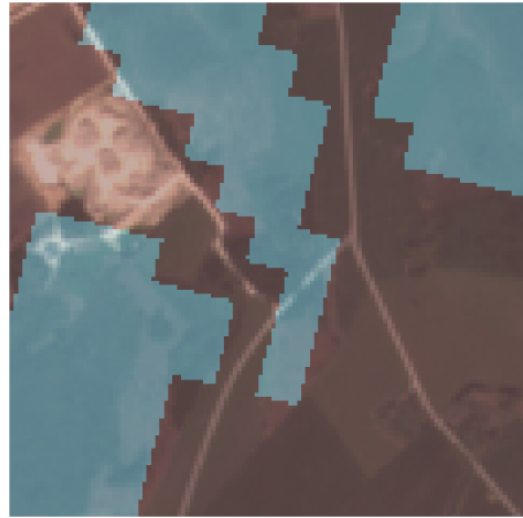
Tabel 10 – Jät kub...

Klass	Täpsus	Saagis	F₁	Pilte
Oliiviistandused	0.73	0.29	0.42	2 489
Karjamaad	0.84	0.66	0.74	19 867
Turbaraba	0.84	0.54	0.65	3 363
Püsivalt niisutatud maa	0.77	0.50	0.61	2 683
Sadamad	0.68	0.30	0.41	88
Riisipõllud	0.68	0.49	0.57	725
Maantee- ja raudteevõrk ja piirnev ala	0.60	0.06	0.11	652
Soolakud	0.76	0.43	0.55	90
Rannasoolakud	0.80	0.33	0.46	286
Sklerofiilne taimestik	0.82	0.38	0.52	2 218
Meri ja ookean	0.99	0.98	0.98	14 603
Hõreda taimkattega alad	0.66	0.10	0.17	248
Puhkealad, pargid, kalmistud	0.77	0.07	0.13	1 007
Üleminekuline metsaala	0.73	0.55	0.63	29 479
Veiniistandused	0.77	0.25	0.38	1 889
Veekogud	0.90	0.76	0.83	11 706
Vooluveed	0.81	0.65	0.72	1 938

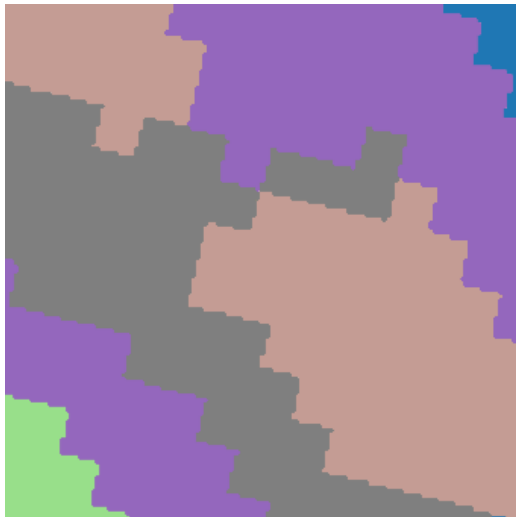
Lisa 4 - Näiteid saadud tulemustest



Näide 1, esimese taseme klassifikatsioon. Üleval - satelliidifoto ja ülekate maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused



Näide 2, esimese taseme klassifikatsioon. Üleval - satelliidifoto ja ülekate maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused

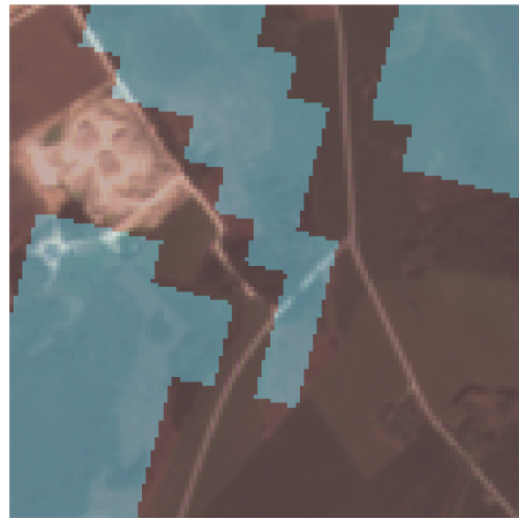


Põllumaa	Karjamaad	Heterogeensed põllumajandus piirkonnad
Metsad	Võsa ja / või rohhtaimestiku ühendused	

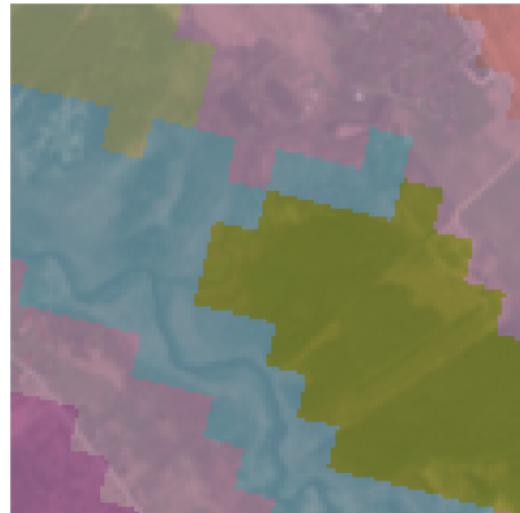


Põllumaa	Karjamaad	Heterogeensed põllumajandus piirkonnad
Metsad	Võsa ja / või rohhtaimestiku ühendused	Sisemaal olevad märgalad

Näide 3, teise taseme klassifikatsioon. Üleval - satelliidifoto ja ülekatte maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused



Näide 4, teise taseme klassifikatsioon. Üleval - satelliidifoto ja ülekatte maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused

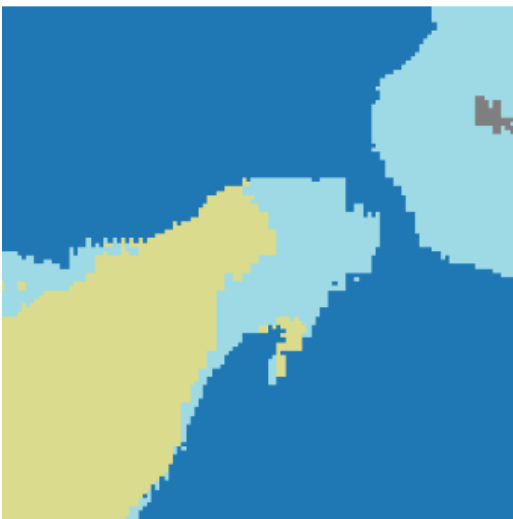
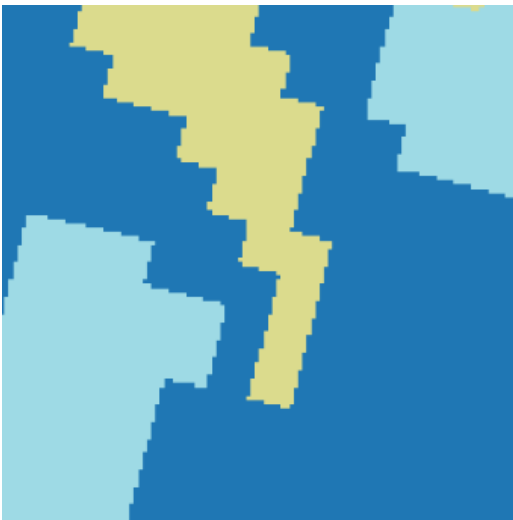


Niisutuseta haritav maa	Karjamaad	Kompleksmaaviljelus
Okasmetsad	Segametsad	Üleminekuline metsaala



Hõredalt hoonestatud alad	Niisutuseta haritav maa	Karjamaad
Kompleksmaaviljelus	Põllumajanduslik maa	Okasmetsad
Segametsad	Üleminekuline metsaala	

Näide 5, kolmanda taseme klassifikatsioon. Üleval - satelliidifoto ja ülekate maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused



Näide 6, kolmanda taseme klassifikatsioon. Üleval - satelliidifoto ja ülekaite maakasutuskaardiga, keskel - maakasutuskaart ja klasside nimetused, all - masinõppe mudeli tulemused ja klasside nimetused