

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Ökoloogia ja maateaduste instituut
Geograafia osakond

Bakalaureusetöö geograafias (12 EAP)

**Suve- ja talvekultuuride tuvastamine Sentinel-2 andmete ja
masinõppe abil**

Emilie Susi

Juhendaja: Prof. Evelyn Uuemaa

Tartu 2025

Infoleht

Suve- ja talvekultuuride tuvastamine Sentinel-2 andmete ja masinõppe abil_SusiEmilie

Töö eesmärk oli hinnata Sentinel-2 andmete ja otsustusmetsa algoritmi sobivust suvi- ja taliviljade eristamiseks varakevadel. Uurimispiirkonnaks oli Tatra org Lõuna-Eestis. Kasutati Google Earth Engine'i keskkonda, kus viidi läbi andmetöötlus, treening ja klassifitseerimine. Mudeli koostamiseks kasutati spektraalindekseid ning välitöödel kogutud proovipunkte.

Treeningandmetel saavutati kuni 86% täpsus, punktipõhisel valideerimisel 78% ja pikslipõhisel hindamisel 67%. Peamised eksimused esinesid taliviljade ja küntud maa vahel, mis viitab spektraalsele kattuvusele kevadel. Visuaalne analüüs näitas, et vigu leidis põllumassiivide kesk- ja servaaladel, kuid esines ka täpselt klassifitseeritud põlde.

Järelduseks kinnitab töö, et Sentinel-2 andmed võimaldavad usaldusväärset klassifitseerimist kevadistes tingimustes. Tulemusi parandaks mitme sensori andmete kasutamine ning aegridade kaasamine.

Märksõnad: Sentinel-2, kaugseire, taliviljad, klassifitseerimine, masinõpe

CERCS kood: T181 - Kaugseire

Abstract

The study evaluates the use of Sentinel-2 imagery and a decision tree algorithm to distinguish winter crops, summer crops (ploughed land), and grasslands in early spring in the Tatra Valley, Estonia. Data processing and classification were carried out in Google Earth Engine using spectral indices and field-verified training points.

The model achieved 86% accuracy on training data, 78% on point-based validation, and 67% in pixel-based assessment. Most confusion occurred between winter crops and ploughed fields due to spectral similarity. Visual analysis revealed misclassifications mainly in field centers and edges.

The study concludes that Sentinel-2 data can reliably support early-season crop classification. Further accuracy could be achieved by integrating multi-sensor data and time series analysis.

Keywords: Sentinel-2, remote sensing, winter crops, classification, machine learning

CERCS code: T181 - Remote sensing

Sisukord

1. Sissejuhatus.....	3
2. Kirjanduse ülevaade.....	4
2.1 Kaugseire areng ja selle roll põllumajanduses.....	4
2.2 Spektraalindeksite ja masinõppe kasutamine maakatte hindamisel.....	5
2.3 Optiliste kaugseireandmete eelised ja piirangud taimkatte ning künni hindamisel.....	7
2.4 Sentinel-2 kanalid.....	8
2.5 Varasemad uurimused Sentinel-2 abil suve- ja talvekultuuride tuvastamise teemal.....	9
3. Andmed ja metoodika.....	11
3.1 Uurimisala.....	11
3.2 Andmed.....	12
3.3 Klasside valik.....	12
3.4 Välitööd.....	12
3.5 Välitöö tulemused.....	13
3.6 Andmeanalüüs.....	14
4. Tulemused.....	18
4.1 Punktipõhise täpsuse hindamine.....	18
4.2 Sentinel-2 satelliitandmetel põhinev klassifikatsioon.....	18
4.2 Lauspinnaline täpsuse hindamine.....	19
5. Arutelu.....	22
6. Kokkuvõte.....	24
7. Summary.....	26
8. Tänuavaldus.....	28
9. Kirjanduse loetelu.....	29
10. Lisad.....	35
Lisa 1. Lihtlitsents.....	35

1. Sissejuhatus

Kaugseireandmete kasutamine põllumajandusmaastike seires on muutumas üha olulisemaks, sest need võimaldavad ulatuslikku, ajaliselt tihedat ja objektiivset teavet taimkatte seisundi ning põllumajandustegevuse kohta (Zarco-Tejada et al., 2014). Eriti on kasvanud huvi satelliidipõhiste andmete rakendamise vastu põllukultuuride klassifitseerimisel ja külvitsüklike tuvastamisel (Khan et al., 2023). Euroopa Kosmoseagentuuri Sentinel-2 missiooni satelliidid pakuvad kõrglahutusega multispektraalandmeid, mille korduvkülastusaja ja spektraalkanalite ulatuse tõttu on need eriti sobilikud põllumajandusliku maakasutuse analüüsiks (Drusch et al., 2012).

Taliviljad ja suviviljad erinevad oma kasvutsüklike poolest: taliviljad külvatakse sügisel, talvituvad ja alustavad kasvu varakevadel, samas kui suviviljad külvatakse kevadel ning nende aktiivne kasv algab kevade teises pooles (Newete et al., 2024). Need fenoloogilised erinevused avalduvad taimede spektraalses käitumises ning võimaldavad neid satelliitandmete põhjal eristada, kasutades sobivaid ajavahemikke ja masinõppemeetodeid (Inglada et al., 2016). Varasemad uurimused on näidanud, et masinõppe algoritmid, näiteks otsustusmets (ingl. k *Random Forest*), suudavad edukalt eristada erinevaid põllukultuure isegi keerukates tingimustes, kui kasutada piisavas koguses spektraalse ja ajas muutuva info baasil treenitud mudeleid (Belgiu & Drăguț, 2016).

Käesoleva töö eesmärk on hinnata Sentinel-2 satelliidi multispektraalsete andmete ning otsustusmetsa algoritmi abil taliviljade ja suviviljade eristamise täpsust Eesti oludes. Uurimus keskendub Lõuna-Eestis paiknevale Tatra orule, kus 2024/25 aastatel viidi läbi välitööd andmete valideerimiseks. Suvi- ja talivilja kaardistamine on vajalik mitmel põhjusel. See aitab paremini planeerida viljavaheldust ja umbrohutõrjet, kuid on ka oluline kasvuhoonegaaside emissiooni hindamise ja valglast toitainete väljakande hindamisel (Behnke & Villamil, 2019).

2. Kirjanduse ülevaade

2.1 Kaugseire areng ja selle roll põllumajanduses

Kaugseire (ingl. k remote sensing) on tehnoloogia, mis võimaldab maapinda ja selle omadusi hinnata distantsilt, kasutades satelliite ja õhusõidukeid varustatuna erinevate sensoritega (Nasa, i.a.). Põllumajanduses on kaugseire võimaldanud automatiseeritud ja ulatuslikku maapinna jälgimist, mis on varasemalt olnud käsitsi mõõtmiste tõttu keeruline ning kulukas (Wang, 2024). Alates 1970. aastatest, kui USA maavaatluse satelliitide programm Landsat alguse sai, on kaugseire pidevalt arenenud, tuues kaasa parema ruumilise ja spektraalse lahutusvõime ning sagedasema andmete kättesaadavuse (USGS, 2025). Tänapäeval on kättesaadavad mitmesugused vabalt kasutatavad andmed (nt ESA Sentinel-1 ja Sentinel-2 satelliitandmed, NASA Landsat 8 andmed), mis võimaldavad analüüsida maapinna seisundit igal aastaajal ja ka muutuste dünaamikat.

Kaugseire võimaldab ülevaatlikku põllumajandusmaa haldamist, sealhulgas hinnata mulla niiskusesisaldust, taimestiku katvust, taimekasvu dünaamikat ning eristada töödeldud ja töötlemata alasid (Karmakar, 2024). Oluline on märkida, et erinevad satelliidid annavad infot erinevates spektraalsetes vahemikes, võimaldades nii optilist vaatlust (näiteks Sentinel-2 multispektraalsed andmed) kui ka mikrolainete vaatlust (Sentinel-1 radarandmed) (ESA, 2019).

Kaugseire kasutuselevõtt on avanud uusi võimalusi täppispõllumajanduses, võimaldades kiiresti tuvastada näiteks kündmise tagajärgi või taimkatteta muldade piirkondi, mis on olulised mulla erosiooni hindamiseks ja põllumajandusliku maaharimise planeerimiseks (Liu et al., 2022). Kaugseire on samuti aidanud tõhustada keskkonnahoidlikku põllumajandust, pakkudes pidevalt uuenevat infot maapinna seisundi kohta ja toetades otsuste tegemist.

Samas esineb kaugseire kasutamisel ka piiranguid, näiteks ilmastikutingimused, andmete ajastamise piiratus ning erinevate sensorite omaduste mõjud analüüsi täpsusele (Abdelmajeed, 2024). Seetõttu on oluline kombineerida mitut andmeallikat ja arendada uusi andmetöötlusmeetodeid, et saavutada maksimaalselt usaldusväärsed ja praktilised tulemused (Liu et al., 2022; Mzid et al., 2021).

2.2 Spektraalindeksite ja masinõppe kasutamine maakatte hindamisel

Maakattetüüpide eristamine kaugseire abil on osutunud keerukaks aga oluliseks ülesandeks põllumajandusliku maakasutuse jälgimisel. Tõhus analüüs eeldab sobivate spektraalindeksite kasutamist ning sobivate andmetöötlusmeetodite rakendamist, et tuvastada mulla katvust, taimestiku seisundit ja põllumajanduspraktikate mõju.

Üks asjakohane ja laialdaselt kasutatav spektraalindeks kündmise ja palja mulla tuvastamisel on *Bare Soil Index* (BSI), mis kasutab sinise, punase, lähiiinfrapuna (NIR) ja lühilaine infrapuna (SWIR) lainepikkusi (Nguyen et al., 2021). BSI on eriti tundlik pinnakatte muutuste suhtes, kuna suudab tõhusalt eristada taimkatteta mulda taimkattega aladest isegi siis kui vegetatsioon on osaliselt lagunenu või hõre (Zha et al., 2003). Indeksit arvutatakse järgmiselt:

$$BSI = \frac{(SWIR+Red)-(NIR+Blue)}{(SWIR+Red)+(NIR+Blue)}$$

BSI väärtused on kõrgemad piirkondades kus pinnas on taimkatteta ja madalamad seal kus domineerib taimkate, mistõttu on see indeks sobilik kündmise, mis toob kaasa märgatava taimkatte eemaldumise ja palja mulla eksponeerimise, tuvastamiseks (Nguyen et al., 2021).

Lisaks BSI-le kasutatakse taimkatte ja põllumajandustegevuse jälgimisel ka mitmeid teisi spektraalindekseid, millest üks levinumaid on Normaliseeritud Vegetatsiooniindeks (NDVI). Sentinel-2 kontekstis on NDVI arvutuskäik järgnev:

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)}$$

NDVI on tundlik taime roheline biomassiga seotud klorofüllü sisalduse suhtes ning seda kasutatakse laialdaselt taimkatte olemasolu ja tiheduse määramiseks (Rouse et al., 1974). Selle võimalikud väärtused varieeruvad vahemikus -1 kuni +1 (Tucker, 1979). Kõrgemad väärtused viitavad intensiivsemale taimkattele ning madalamad väärtused vähesele või puudulikule taimkattele, nagu paljas muld või künd (Tucker, 1979).

NDVI-l, nagu ka teistel indeksitel, on piirangud, eriti intensiivse taimkatte korral, kuna indeksi väärtus kipub küllastuma – s.t muutub vähem tundlikuks biomassi lisandumise suhtes alates teatud punktist (Sellers, 1985).

Lisaks on kasutusel ka normaliseeritud veekoguindeks (MNDWI) ja normaliseeritud linnamaastikuindeks (NDBI), mille vastavad arvutuslikud valemid on järgnevad:

$$MNDWI = \frac{GREEN-SWIR}{GREEN+SWIR} \text{ (Space4Water Portal, 2020)}$$

$$NDBI = \frac{SWIR-NIR}{SWIR+NIR} \text{ (Kafy, Khan, Islam, \& Sarker, 2020)}$$

Talivilja, mille kasvuperiood algab juba sügisel, puhul võib sügisene BSI näitaja olla madal tänu taimkatte olemasolule, samas kui hilissügisel või kevadel toimunud kündmine põhjustab järsu BSI väärtuse tõusu. Sellised erinevused võimaldavad kasutada BSI väärtuste ajas muutumise analüüsi, et eristada sügisel külvatud talivilju sügisel küntud põldudest (Kross et al., 2015). Parimaks ajaperioodiks taimkatteindeksite võrdlemiseks peetaksegi talivilja puhul varakevadet (õitsemise eelset perioodi), sest enne õitsemist on taimik piisavalt arenenud, et eristada biomassi erinevusi, kuid taimekatte pole veel nii tihe, et vegetatsiooniindeksid (nagu NDVI) küllastuksid (Wang et al., 2022).

BSI andmetöötlusel kasutatakse sageli statistilisi ja masinõppe mudeleid, nagu otsustusmets või tugivektorimeetod (SVM), mis võimaldavad klassifitseerida maapinna katvust erinevate väärtuste ja muutumistrendide põhjal (Belgiu & Drăguț, 2016). Masinõppemudelid võimaldavad kombineerida BSI väärtusi koos teiste ajas muutuvate parameetritega, näiteks NDVI-ga, suurendades nii klassifitseerimise täpsust (Zayani et al., 2023).

Lisaks on oluline ajaliselt tihe andmestik, mis võimaldab jälgida mulla ja taimestiku seisundi muutusi erinevatel ajaperioodidel. Sentinel-2 kõrge ajalis-ruumiline resolutsioon on siinjuures osutunud eriti väärtuslikuks, pakkudes piisavat sagedust, et registreerida lühiajalisi muutusi, näiteks kündmisjärgset mullakatte muutumist (Drusch et al., 2012). Samas tuleb märkida, et aegridade töötlemine on väga mahukas ja nõuab oluliselt arvutusressursse ning efektiivseid algoritme, et toime tulla suure andmemahutude ja sagedaste mõõtmistega (Ma et al., 2015).

Viimastel aastatel on järjest enam kasutusele võetud masinõppe algoritme, mis võimaldavad tuvastada keerukaid mustreid ja seoseid maapinna seisundi ning selle kasutusviiside vahel. Sellised meetodid nõuavad sageli kvaliteetset treeningandmestikku, mida luuakse välitööde käigus.

2.3 Optiliste kaugseireandmete eelised ja piirangud taimkatte ning künni hindamisel

Optilised kaugseireandmed pakuvad põllumajanduses ja keskkonnaseires mitmeid eeliseid võrreldes traditsiooniliste meetoditega, nagu maapealsed vaatlus- ja mõõtmistehnikad. Üheks oluliseks tugevuseks on optiliste sensorite võime eristada erinevaid materjale ja objekte nende spektraalsuse alusel (Zheng et al., 2014). Materjalide kiirguse neelduvus ja peegelduvus muutuvad sõltuvalt valguse lainepikkusest, mistõttu on võimalik tuvastada optilise kaugseirega mulla tüüpi, taimestiku seisundit ja ka põllumajandustegevusi (Zheng et al., 2014). Lisaks võimaldab kaugseire ulatuslikku ja korduvat andmekogumist, mis on oluline ajalise seisundi muutuste jälgimisel, näiteks põllukultuuride kasvuperioodid või maakasutuse muutused (Zheng et al., 2014).

Optilised andurid, eriti nähtava valguse ja lähiiinfrapuna (VIS–NIR) spektrivahemikus, on tõhusad taimekasvu jälgimisel, saagiklassifitseerimisel ja põllumajandustööde tuvastamisel (Khanal, KC, Fulton, Shearer & Ozkan, 2020). NIR- ja termilised sensorid võimaldavad avastada taimede seisundit või haiguseid juba enne, kui need visuaalselt avalduvad (Khanal et al., 2020).

Sentinel-2 missiooni üks eelistest on suur pinnakatvus, andmed on sageli tasuta või madalate kuludega võrreldes traditsiooniliste andmekogumismeetoditega, suurendades nende andmete kättesaadavust (Khanal et al., 2020). Lisaks võimaldab Sentinel-2 andmete kõrge ruumiline (10–60m) ja ajaline (viiepäevane) eraldusvõime detailset aegridade põhise analüüsi, ka väiksemate põllumassiivide puhul (Bégué et al., 2018). See omakorda võimaldab tuvastada lühiajalisi protsesse, nagu viljakoristus, kündmine ja põllumajandusjääkide käsitlemine (Bégué et al., 2018).

Võrreldes traditsiooniliste andmekogumise meetoditega võimaldab kaugseire tõhusamat ja ajaliselt tihedamat seiret laiaulatuslikel aladel, muutes selle sobivaks tööriistaks näiteks kündmise või saagikoristuse jälgimisel (Zheng et al., 2014). Seeläbi on kaugseire andmetel oluline roll keskkonnasäästliku põllumajanduse planeerimisel, võimaldades andmepõhist otsustamist (Zheng et al., 2014).

Siiski kaasnevad optiliste kaugseireandmete kasutamisega ka mitmed piirangud. Näiteks on kündmise kaardistamine keerukas ajastuse erinevuste (küundmine vs külvamine) ja taimede varieeruvuse tõttu, mis võivad tekitada vigu andmetölgenduses (Zheng et al., 2014). Samuti

on optilised kaugseireandmed tundlikud ilmastikutingimuste suhtes - nt pilvisus ja lumekate võivad takistada satelliitpiltide kasutamist muutes need kasutuskõlbmatuks (Khanal et al., 2020).

Kuigi masinõppe meetodid võimaldavad luua keerukaid ja täpseid mudeleid, mis seovad kaugseireandmeid põllumajandust puudutavate aspektidega, sõltub nende mudelite täpsus treeningandmete kvaliteedist, mis ei pruugi piisavalt peegeldada keerulisi bioloogilisi ja keemilisi protsesse (Khanal et al., 2020). Lisaks tuleb arvestada, et väga kõrge resolutsiooniga piltide töötlemine on ressursimahukas ning nõuab suurt arvutusvõimsust, mistõttu on selliseid uuringuid sageli võimalik teostada vaid piiratud ulatuses või väikestel aladel (Bégué et al., 2018).

2.4 Sentinel-2 kanalid

Sentinel-2 pakub kõrglahutusega andmeid 13 spektraalkanalis, mis hõlmavad nähtava valguse, lähiinfrapuna (NIR) ja lühilaineinfrapuna (SWIR) alasid (European Space Agency, 2015). Varasemad uurimused on näidanud, et mõned spektraalkanalid osutuvad eriti oluliseks põllumajanduslike maakatete, mullatüüpide ja kultuuride eristamisel (Immitzer, Vuolo, & Atzberger, 2016).

Tihti kasutatakse nähtava valguse alasid – sinist (B2 – 490 nm), rohelist (B3 – 560 nm) ja punast kanalit (B4 – 665 nm) – kuna need on aluseks vegetatsiooninäitajate, näiteks NDVI arvutamisel (B4 ja B8) (Clevers & Gitelson, 2013). Punase serva (red edge) kanalid (B5 – 705 nm, B6 – 740 nm, B7 – 783 nm) on samuti olulised taimkatte seisundi hindamisel, eriti varajase terviseprobleemi või kultuuride vahetumise tuvastamisel, kuna need on tundlikud klorofüllü kontsentratsiooni muutustele (Clevers & Gitelson, 2013).

Sentinel-2 lähiinfrapuna kanal (B8 – 842 nm) on oluline taimkatte biomassi ja tiheduse hindamiseks, kuna taimestik peegeldab selles spektraalpiirkonnas kõrgel tasemel kiirgust, võimaldades selgelt eristada taimkatet mullast ja teistest pindadest (European Space Agency, 2015).

SWIR-kanalid (B11 – 1610 nm ja B12 – 2190 nm) võimaldavad eristada mullatüüpe ja niiskuse taset ning on eriti kasulikud mullapindade tuvastamisel, näiteks kündmise või palja mulla korral (Bousbih et al., 2019). Need kanalid aitavad eristada märgi ja kuivi alasid ning tuvastada erinevaid mullastruktuure (Bousbih et al., 2019).

Sentinel-2 kanalid erinevad ka ruumilise lahutusvõime poolest: osad (nt B2, B3, B4 ja B8) on 10 m resolutsiooniga, red edge ja mõned SWIR kanalid (nt B5, B6, B7, B8a, B11, B12) on 20 m ning atmosfääri korrigeerimiseks kasutatavad kanalid (B1, B9, B10) on 60 m resolutsiooniga (European Space Agency, 2015).

2.5 Varasemad uurimused Sentinel-2 abil suve- ja talvekultuuride tuvastamise teemal

Varasemates uurimustes on Sentinel-2 andmete võimekust edukalt rakendatud suve- ja talvekultuuride tuvastamiseks ning kasvatustsüklite jälgimiseks erinevates tingimustes. Talvekultuuride (nt talinisu ja talioder) kaardistamiseks on kasutatud erinevaid lähenemisi, sealhulgas hierarhilist ja otseekstraktsiooni-põhist klassifitseerimist, samuti pikslipõhist ja objektipõhist meetodikat (Xie & Niculescu, 2022). Uuring Prantsusmaal näitas, et hierarhiline klassifitseerimine saavutab parema täpsuse (üldine täpsus 93,2%), kusjuures pikslipõhine meetodika sobis paremini just talvekultuuride eristamiseks (Xie & Niculescu, 2022). Lisaks kasutati fenoloogiliste staadiumite tuvastamiseks Sentinel-1 SAR andmestikku, mis võimaldas kultuuride kasvufaaside jälgimist ka pilvise ilmaga (Xie & Niculescu, 2022).

Pan et al. (2021) uurisid talvekultuuride tuvastamist Hiinas, rakendades Sentinel-2 ja Landsat andmete põhjal koostatud pikslipõhiseid aegridasid, millelt arutati fenoloogilised indikaatorid, nagu kasvuperioodi algus ja lõpp, rohetumiskiirus ja kasvuperioodi pikkus. Tulemused näitasid, et fenoloogiapõhine lähenemine võimaldab talvekultuuride külvipindade täpset kaardistamist suurel territooriumil (üldine täpsus 94,56%) (Pan et al., 2021).

Varasemad tööd on näidanud ka seda, et mitme sensori andmete (nt Sentinel-2, Landsat ja SPOT-4) ühendamise fenoloogiliste ja struktuursete biofüüsikaliste näitajate alusel võib täiustada kultuuride klassifitseerimist (Waldner et al., 2015). Biokeemiliste parameetrite kasutamine võimaldas saavutada täpsused vahemikus 73–77% (Waldner et al., 2015).

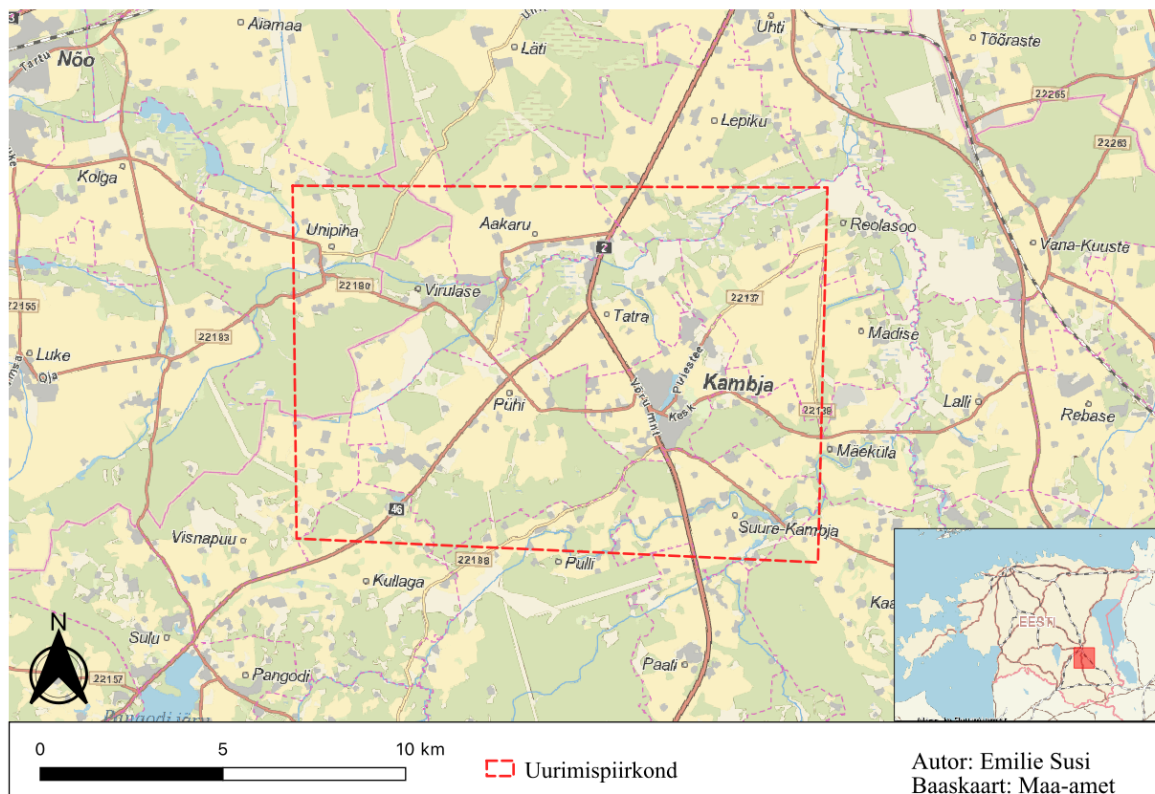
Suvekultuuride kaardistamisele keskendunud uurimus Itaalias kasutas varase suve ajal optiliste ja SAR-andmete (Landsat 8) kombinatsiooni, rakendades klassifitseerimispuude algoritmi (Villa et al., 2015). Tulemused kinnitasid, et SAR-andmete kaasamine suurendab klassifikatsioonitäpsust ja algoritmi ülekantavust erinevatele aastatele, saavutades täpsuse üle 86% (Villa et al., 2015).

Need uurimused kinnitavad Sentinel-2 (ja teiste sensorite) sobivust suve- ja talvekultuuride tuvastamiseks tänu oma kõrgele ruumilisele ja ajalisele eraldusvõimele.

3. Andmed ja meetodika

3.1 Uurimisala

Uurimispiirkonnaks valiti ca 15×10 km suurune ala Tartumaal Tatra ürgorus, kus tehti ka paralleelselt andmete valideerimise tarvis välitöid (Joonis 1).



Joonis 1. Uurimispiirkonna kaart

Tatra ürgorg asub Kambja vallas, Tartu maakonnas (Maa- ja Ruumiamet, 2025). See moodustab idapoolse osa Voika-Tatra ürgorust Otepää kõrgustiku põhjaserval (Eesti Entsüklopeedia, 2011). Tatra ürgoru piirkonna maakasutus (joonis 1) on tugevalt põllumajanduse suunitlusega, mida kinnitab ulatuslik põllumaade levik kogu piirkonnas (Maa- ja Ruumiamet, 2025). Tihedamat asustust ja elamumaad leidub peamiselt Tatra külas ning selle ümbruses, kus esineb ka väiksemaid tootmis- ja sotsiaalmaa alasid (Maa- ja Ruumiamet, 2025).

3.2 Andmed

Satelliitandmetena kasutati Sentinel-2 missiooni Level-2A taseme pinnapeegelduse andmeid, millele on atmosfääri korrektsioon ja geomeetiline korrektsioon tehtud. Kvaliteetse andmestiku tagamiseks valiti andmetöötluseks ainult need kujutised, mille pilvisuse protsent oli alla 10%. Lisaks eemaldati kõik pilvede ja varjude mõjuga pikslid, kasutades Sentinel-2 *Cloud Score+* tööriista (Pasquarella et al., 2023). Uuritavaks ajavahemikuks valiti 2025. aasta aprillikuu, kuna sellel perioodil esinevad Eestis korraka veen kasvamise algjärgus olevad taliviljad ja taimkatteta (küntud või muul moel haritud) faasis olevad põllud, mis tagab spektraalse varieeruvuse klassifitseerimiseks sobival määral.

3.3 Klasside valik

Uuringu raames jagati põllumajanduslik maa kolme üldisesse klassi: rohumaa, küntud maa (tõenäoliselt suvekultuuri külvamise ettevalmistusetapina) ning talivili. Rohumaa hõlmab alasid kus ei toimu regulaarset harimist või mille optilised tunnused viitavad mitmeaastastele rohttaimede kooslustele (nt karjamaa või niit). Küntud maa tähistab alasid, kus visuaalselt esineb kevadel ja suvel tugevat muutlikkust (suvekultuuride kasvatamine, millele eelneb sageli kevadine künd). Talivili määrati aladele, millel oli rohekat vegetatsiooni juba varakevadel (aprill) ning mille kasvutsükkel viitas sügisel külvatud kultuurile (nt talirukis või taliraps).

3.4 Välitööd

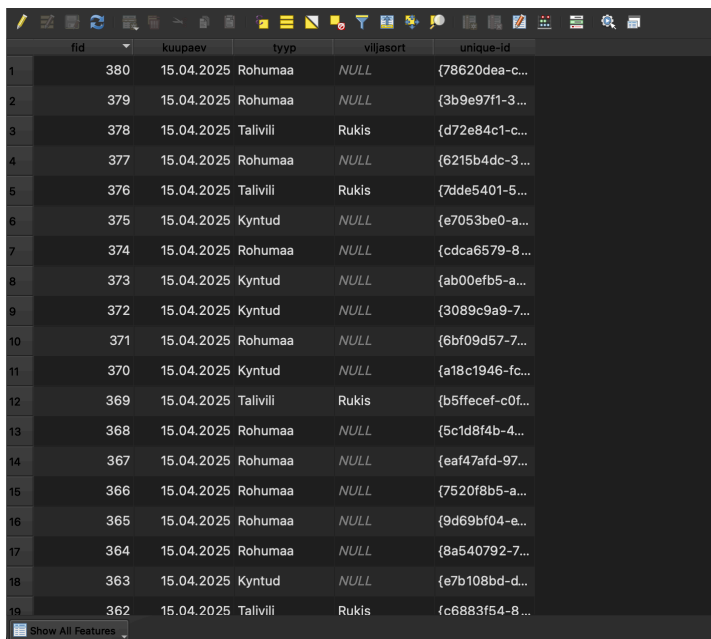
Välitööd viidi läbi 2024. aasta sügisel ja 2025. aasta kevadel Tatra ürgorus. Välitööde aluseks võeti PRIA põllumassiivide kaart, mille abil käidi läbi kõik avalikkusele ligipääsetavad üksused (PRIA, 2025). Uuringualal kasutati MerginMaps mobiilirakendust maakasutuspunktide kogumiseks. Kokku märgiti 373 punkti, mis kuulusid kolme klassi: taliviljad (nt talirukis ja taliraps), äsja küntud või haritud põllud ning rohumaad, sealhulgas püsirohumaad ja ajutised heinamaad. Talivilja puhul märgiti ära ka arvatav kultuur. Lisaks tehti hilisema valideerimisprotsessi tarvis igast põllumassivist ka pilt ning talivilja olemasolul 3 tükki - vastavast taimest lähedalt, kasvavast kultuurist mõõdukalt kauguselt ning põldu üldisemalt näitav kaugemalt tehtud pilt.

MerginMaps on mobiilirakendus, mis võimaldab ruumiliste andmete kogumist ja haldamist maastikul töötades. Rakendus on integreeritud avatud lähtekoodiga QGIS platvormiga ning

võimaldab kasutajatel kasutada ja sünkroniseerida QGIS projekte mobiilseadmes. See muudab võimalikuks andmekogumise ka piirkondades kus puudub püsiv internetiühendus (Hodkiewicz & Paul, 2023). MerginMaps toetab erinevaid ruumiandmete vorme ning võimaldab kogutud andmeid pilve kaudu sünkroniseerida (Aherin & Paulovic, 2025). See toetab koostööpõhist töövoogu ning aitab vältida andmete käsitsi sisestamisest tekkivaid vigu (Aherin & Paulovic, 2025). Rakendust kasutatakse nii teadusuurimustes kui ka keskkonnaseires ja linnaplaneerimises (Aherin & Paulovic, 2025).

3.5 Välitöö tulemused

Välitööde tulemusena koguti kahe perioodi (2024 sügis ja 2025 kevad) andmed maakattetüübi kohta 180 põllumassiivilt (joonis 2) ning vastavalt ka 892 pilti antud massiividest (joonis 3). Andmed koguti Merginmaps programmiga (joonis 4).

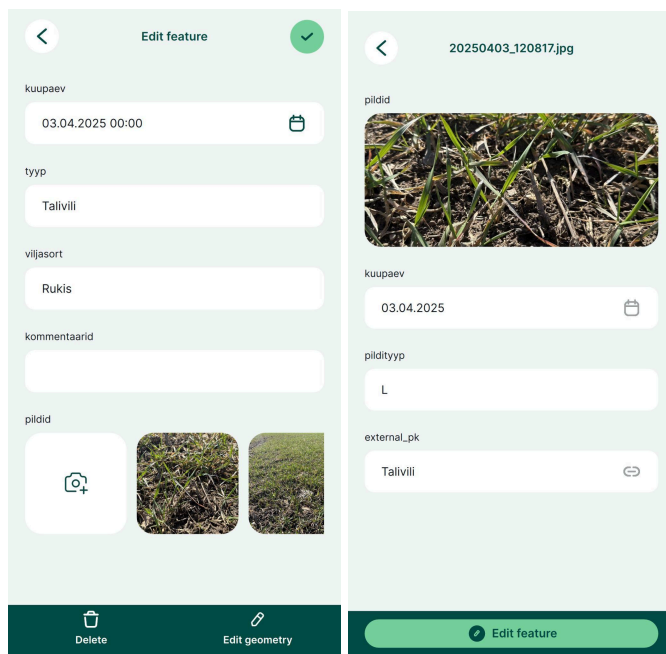


fid	kuupaev	tyyp	viljasort	unique-id
1	380	15.04.2025 Rohumaa	NULL	{78620dea-c...
2	379	15.04.2025 Rohumaa	NULL	{3b9e97f1-3...
3	378	15.04.2025 Talivili	Rukis	{d72e84c1-c...
4	377	15.04.2025 Rohumaa	NULL	{6215b4dc-3...
5	376	15.04.2025 Talivili	Rukis	{7dde5401-5...
6	375	15.04.2025 Kyntud	NULL	{e7053be0-a...
7	374	15.04.2025 Rohumaa	NULL	{cdca6579-8...
8	373	15.04.2025 Kyntud	NULL	{ab00efb5-a...
9	372	15.04.2025 Kyntud	NULL	{3089c9a9-7...
10	371	15.04.2025 Rohumaa	NULL	{6bf09d57-7...
11	370	15.04.2025 Kyntud	NULL	{a18c1946-fc...
12	369	15.04.2025 Talivili	Rukis	{b5ffecef-c0f...
13	368	15.04.2025 Rohumaa	NULL	{5c1d8f4b-4...
14	367	15.04.2025 Rohumaa	NULL	{eaf47afd-97...
15	366	15.04.2025 Rohumaa	NULL	{7520f8b5-a...
16	365	15.04.2025 Rohumaa	NULL	{9d69bf04-e...
17	364	15.04.2025 Rohumaa	NULL	{8a540792-7...
18	363	15.04.2025 Kyntud	NULL	{e7b108bd-d...
19	362	15.04.2025 Talivili	Rukis	{c6883f54-8...

Joonis 2. Vaatlusandmete atribuuttabel



Joonis 3. Näidis tehtud piltidest (allikas: erakogu)



Joonis 4. Väljavõte Merginmaps programmist

3.6 Andmeanalüüs

Andmetöötlus viidi läbi Google Earth Engine (GEE) pilvepõhisel platvormil, mis võimaldab efektiivselt töödelda ajaliselt ja ruumiliselt suuremahulisi satelliitandmeid (Gorelick et al., 2017).

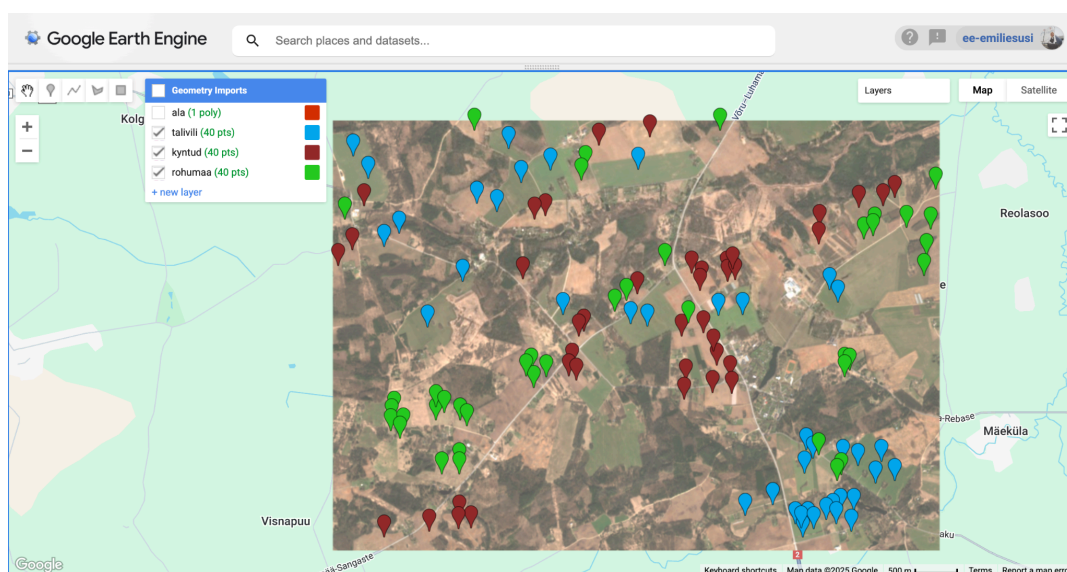
Andmetöötluse esimeses etapis loodi Sentinel-2 satelliitandmete põhjal aprillikuu mediaankomposiit, eesmärgiga vähendada pilvisusest ja juhuslikest anomaaliatest tingitud

müra. Komposiidi koostamisel kasutati pindala piires kõiki aprillikuu jooksul kogutud pilti, mille pilvisuse osakaal oli alla 10%. Kasutatavad spektraalribad ja indeksid arvutati sellest ajavahemikust moodustatud pildikogumist, mille alusel arvutati mediaanväärtused iga piksli kohta. Lõplikku komposiiti kaasati kokku 6 Sentinel-2 pilti, mis saadi Google Earth Engine'i kaudu tehtud päringu tulemusel. Mediaankomposiit moodustas aluse edasisteks spektraalindeksite arvutusteks ja klassifitseerimisanalüüsiks.

Seejärel arvutati eelmainitud valemite põhjal (ptk 2.3) indeksid, mis rõhutavad erinevaid omadusi: vegetatsiooni aktiivsust mõõtev normaliseeritud taimestiku indeks (NDVI), paljapinnaindeks taimkatteta mulla iseloomustamiseks (BSI), niiskustingimustele tundlik muudetud veeindeks (MNDWI) ning linnastumist ja tihendatud ala käsitlev indeks (NDBI). Indeksite lisamine võimaldas suurendada maakatte kategooriate eristamisvõimet. Need indeksid liideti koos spektraalribadega komposiitkujutiseks, mis edastas maksimaalse spektraalse info mudelile.

Suve- ja taliviljade eristamiseks valiti masinõppel põhinev otsustusmetsa algoritm, mis sobib keerukamate ja kattuvate klasside eraldamiseks ning ei vaja eeldust andmete jaotuse kohta (Tugac, Simsek, & Torunlar, 2024).

Usaldusväärse mudeli treenimiseks loodi välitööde käigus kogutud andmetel põhinev treeningandmestik. Välitööde andmeid kasutati GEE platvormil piksli põhisel otsustusmetsa treeningandmetega (joonis 5).



Joonis 5. Väliandmetel põhinevad treeningpunktid GEE keskkonnas.

Igale klassile (talivili, küntud, rohumaa) määrati 40 treeningpunkti (joonis 5). Klassifitseerimisel kasutati otsustusmetsa algoritmi, mille parameetrina määrati otsustusmetsa laiuseks 50 puud. Klassifikaatorile anti sisendiks kõik Sentinel-2 kanalid (B1-B12) ja eelnevalt arvutatud indeksid NDVI, BSI, NDBI ning MNDWI. Treenitud mudel rakendati kogu uurimispiirkonnale, mille tulemusena loodi pikslipõhine klassifitseeritud raster, milles iga piksel määrati ühte kolmest klassist (talivili, küntud, rohumaa) vastavalt sarnasusele treeningandmetega.

Järgnevalt tehti lausaline kontroll klassifitseeritud põldude kvaliteedi hindamiseks. Selle tarvis rakendati esimese sammuna Sieve tööriista, mille eesmärk oli eemaldada väiksemaid müraribasid ja üksikuid väärnimärgistatud pikseleid, mis tekkisid soola-pipra efektina (klassifitseeritud pildil juhuslikult hajutatud üksikuid valesid pikseleid) klassifikatsiooni tulemustes. Piksli klastrid lävendiks määrati 3, et vältida väiksemate põldude andmete kadumist. Sieve tööriista kasutamise eesmärk oli puhastada rasterkiht pisivigadest parandades üldist klassifikatsiooni ühtlust ja usaldusväärsust. Viimaks välistati muud maakattetüübid rasterkihilt, kasutades põldude kihti maskina. Eemaldatud ala eraldi ei klassifitseeritud, vaid lihtsalt eemaldati.

Seejärel väljastati puhastatud andmed rastervormingus edasiseks analüüsiks ja võrdlemiseks välitööde andmetega (joonis 6).

Failid viidi QGIS platvormile kus neid võrreldi välitöödel kogutud andmetega. Andmete tõepärasuse huvides filtreeriti välja üksikud põllumassiivid, millel kasvas rohkem kui üks kultuur või esines rohkem kui üks maakattetüüp. Lisaks jäeti valideerimisandmestikust välja sügisperioodil kogutud andmed, mida kasutati üksnes lisainfona.

Klassifitseeritud andmestikuga võrdluse jaoks oli tarvis esmalt välitööde punktikiht ja põllumassiivide kiht ühendada, et saada massiivid, millel on vastav valideeritud maakatteklass. Selle jaoks kasutati *spatial join* tööriista. Kui välitööde punktikiht ja põllumassiivide polügoonikiht olid ühendatud, siis määrati igale maakattetüübile vastav klass ning viidi läbi võrdlus.

Võrdluseks kasutati QGIS platvormil *Zonal Statistics* ja *Group Stats* tööriistu. Kõigepealt rakendati tööriista *Zonal Statistics*, et liita klassifitseeritud rasteri väärtused igale välitöödel

määratud põllupolügoonile. Selle tulemusel lisandus igale polügoonile väli, mis näitas vastava ala domineerivat klassifitseeritud väärtust (*cls_majority*).

Seejärel loodi uus väli *match*, mille väärtus näitas kas klassifitseeritud väärtus (*cls_majority*) kattus tegeliku väärtusega (klass). Selleks kasutati käsklust:

```
CASE
    WHEN "tyyp" = "cls_majority" THEN 'õige'
    ELSE 'vale'
END.
```

Välja väärtused olid “õige” juhul, kui klassid kattusid, ja “vale”, kui mitte. Väli *match* võimaldas *Field Statistics* tööriistaga loendada õigete ja valede vastete arvu ning arvutada üldise täpsuse valemiga:

Täpsusprotsent = (õigete vastete arv / kõikide polügoonide arv) * 100

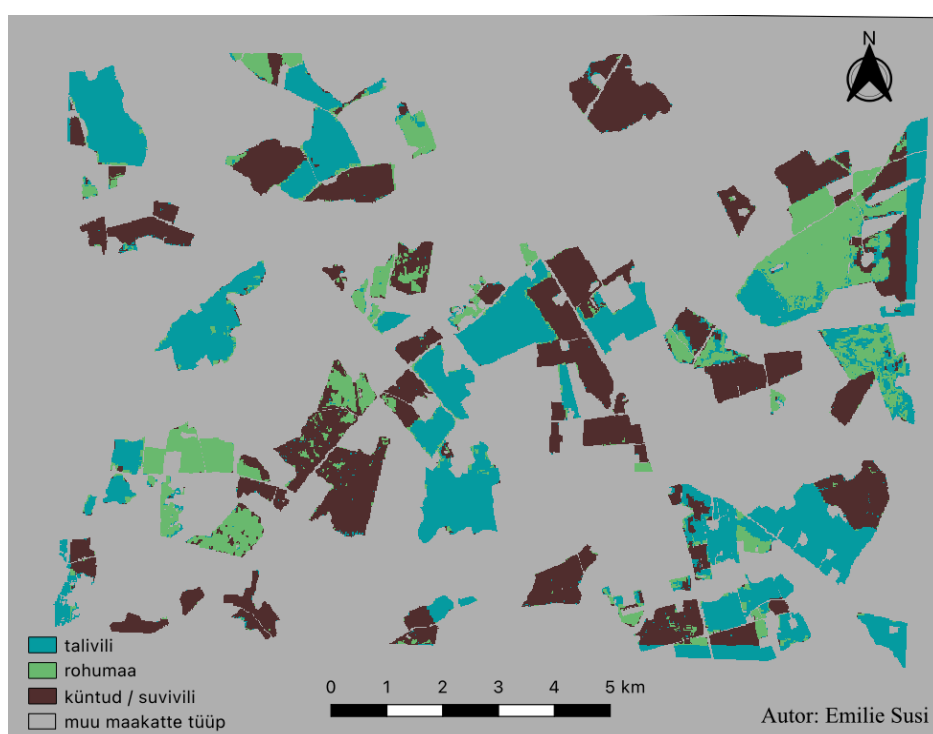
Lisaks viidi läbi täpsem analüüs, kasutades *Group Stats* moodulit. Selle abil koostati veamatriks kus ridadele määrati väli “klass” (tõene klass), veergudele väli “cls_majority” (klassifitseeritud väärtus), väärtusteks kasutati polügoonide unikaalset identifikaatorvälja “xy_id”, ja agregeerimise meetodiks *Count*.

Tulemuseks saadud matriks näitas, kui palju objekte kuulus igasse tõelise ja klassifitseeritud klassi kombinatsiooni. Diagonaalelemendid näitasid õigesti klassifitseeritud juhtumite arvu, kõrvalkaldeid tähistasid valesid klassifikatsioone.

4. Tulemused

4.2 Sentinel-2 satelliitandmetel põhinev klassifikatsioon

Treeningandmete kasutamine klassifikatsiooniks Google Earth Engine'i (GEE) keskkonnas kasutades Sentinel-2 andmeid ning rakendades otsustusmetsa algoritmi, andis häid tulemusi. Punktipõhise klassifikatsioonimudeli üldine täpsus (*overall accuracy*) ulatus 86%-ni, mis kinnitas mudeli head vastavust tegelikele klassidele ja ennustatud tulemustele. See viitab mudeli heale võimekusele andmeid üldistada ning annab potentsiaali ka mudeli laiemale kasutusele piirkondades väljaspool treeningala.



Joonis 6. Klassifitseeritud maakattetüüpidega raster

4.1 Punktipõhise täpsuse hindamine

Punktipõhine täpsuse hindamine toimus välitöödel kogutud proovipunktide ja klassifitseeritud tulemuste võrdlemisel. Selle meetodi abil hinnati mudeli suutlikkust õigesti klassifitseerida maapinna seisundit kindlates kontrollpunktides.

Punktipõhise hindamise tulemusena saadi üldiseks täpsuseks 78%, mis arvutati valemi kaudu:

$$TÄPSUS (\%) = \frac{\text{õigesti hinnatud üksused}}{\text{kõik hinnatud üksused}} \times 100$$

Lisaks koostati punktipõhise hindamise veemaatriks (tabel 1), kus on välja toodud klassidevahelised eksimused - valideeritud klassid ehk vt_klass vastavad veergudele, klassifitseeritud klassid ehk cls_klass ridadele.

Tabel 1. Punktivõrdluse veemaatriks

vt_klass / cls_klass	0 (talivili)	1 (küntud)	2 (rohuma)
0 (talivili)	36	5	4
1 (küntud)	3	49	0
2 (rohuma)	8	12	33

Veemaatriksist on näha, et taliviljade klassist eksiti peamiselt küntud maa kasuks (5 eksimust), samuti juhtus vastupidine eksimus (3 eksimust). See on seletatav nende kultuuride spektraalsete omaduste kattuvusega, eriti kevadistel kasvufaasidel, kus küntud maa ja varajane talivili on spektraalselt sarnased. Samuti kattusid spektraalselt künnipõllud ja rohumaad, põhjustades nende klasside vahel 12 eksimust. Need kattuvused võivad tuleneda pinnase niiskuse, värvuse ja tekstuuri spektraalsetest profiilidest. Punktipõhine analüüs võimaldab detailselt hinnata mudeli toimivust kindlates maapinna olukordades ja on oluline mudeli parendamisel.

4.2 Lauspinnaline täpsuse hindamine

Lauspinnalise kontrolli meetod hindab klassifikatsiooni täpsust kogu uurimisalal, kasutades rasterandmete pikslipõhist võrdlust tõeliste klassidega. Alljärgnevas veemaatriksis on esitatud lauspinnaline võrdlus tõeliste ja prognoositud klasside vahel pikslite arvuna (tabel 2).

Tabel 2. Lausalise kontrolli veemaatriks

tõeline/prognoositud	0 (talivili)	1 (küntud)	2 (rohuma)
0 (talivili)	414	78	107
1 (küntud)	42	323	21
2 (rohuma)	49	134	145

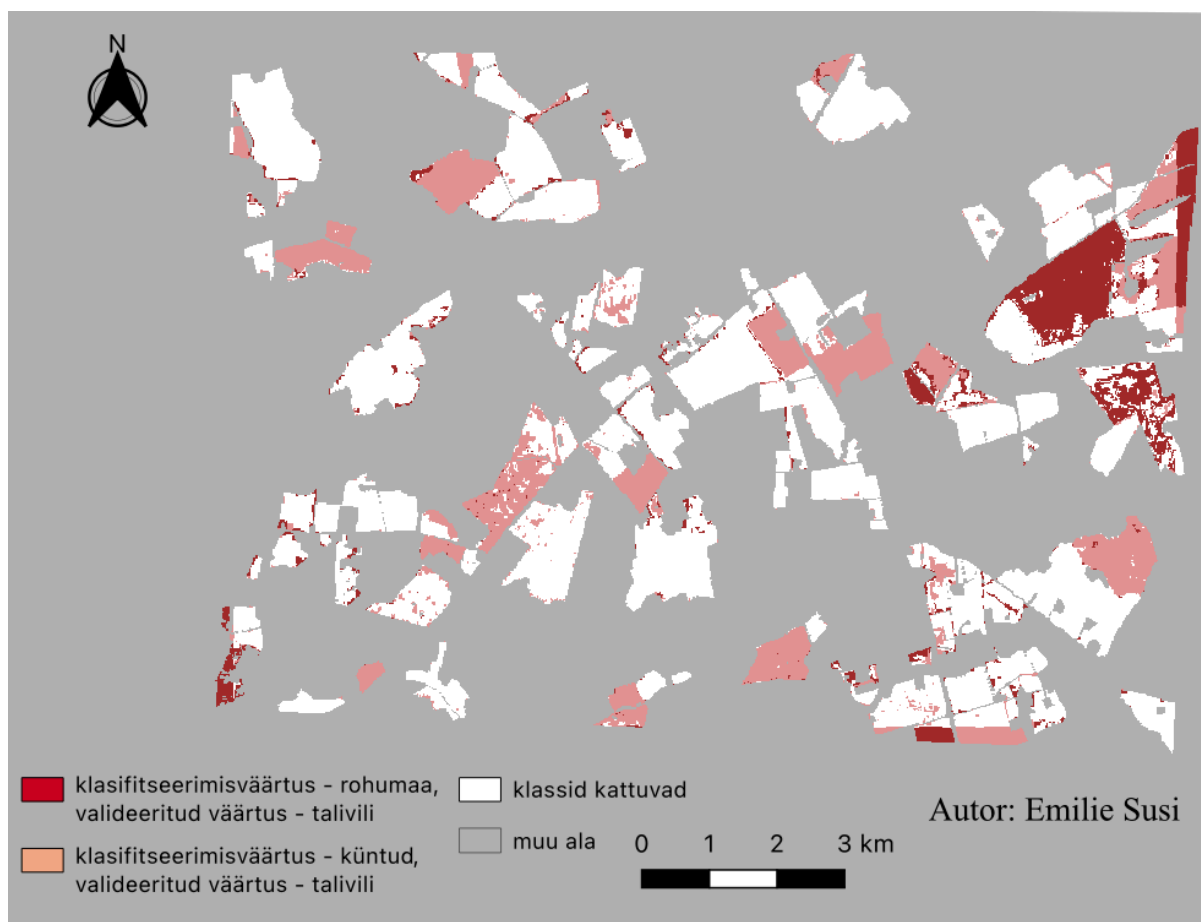
Õigesti klassifitseeritud pikslite hulgas tuvastati talivilja klassis 414 pikslit, küntud maa korral 323 ning rohumaa puhul 145 pikslit. Küll aga aeti segi taliviljade pikslid sageli küntud maa (78 pikslit) ja rohumaaga (107 pikslit). Küntud maa klassis oli eksimusi talivilja suunas (42 pikslit) ja rohumaaga (21 pikslit). Rohumaa klassis oli suuri kattumisi küntud maaga (134 pikslit), vähem taliviljaga (49 pikslit).

Selline tulemus võib tuleneda maakattetüüpide spektraalsest sarnasusest. Näiteks võib varases kasvufaasis taliviljast paista läbi mullapind, mistõttu selle pealt kogutud andmed meenutavad osaliselt küntud maad. Samuti võib rohumaa sarnaneda taliviljaga, eriti kui rohumaa taimestik on madal ja visuaalselt sarnane noore taliviljaga.

Teiseks võimalikuks põhjuseks on treeningandmete vähesus. Kui treeningandmed ei kata piisavalt klassidesisest varieeruvust, näiteks rohumaa erinevaid taimeliike või talivilja erinevaid kasvufaase, siis ei suuda algoritm neid klasse piisavalt täpselt eristada.

Lauspinnalise klassifikatsiooni üldine täpsus oli 67% ehk veidi rohkem kui $\frac{2}{3}$ pikslitest klassifitseeriti tõeselt.

Lisaks toodi välja vigaste klassifitseeritud väärtuste paiknevus põllumassiividel (joonis 7). Mainimata variatsioone klassifitseerimiskihi ja valideerimistulemuste rasterkihi vahel visuaalselt ei esinenud.



Joonis 7. Klassifitseeritud väärtuse kattuvus tegeliku väärtusega.

Visuaalsel analüüsil asub kõige rohkem väärtusega pikseid põllumassiivide keskosas ning nende hulgas enim klassifikatsiooni järgi küntud pinnasega, kuid valideerimisandmestiku kohaselt taliviljaga kaetud maad. Veel on väärtusega määratletud klassifikatsiooni rasterkihi kohaselt rohumaad, mis taaskord on valideerimisandmestiku järgi talivili. See võib tähendada, et ühel põllumassiivil kasvav taimestik on erinevates kasvujärgudes või taimkate on madal (muld paistab läbi), mistõttu nende klassifitseerimine on raskendatud. Samuti võib põhjus olla selles, et treeningandmestikus ei olnud piisavalt esindatud erinevaid kasvufaase või põldude variatsiooni, mistõttu mudel eksis nende eristamisega.

Lisaks on näha, et vead ei esine ainult massiivide keskosas, vaid ka servades. Seda põhjustab tõenäoliselt põllumassiivi katvast maakattetüübist vahetus muule tüübile, mis tekitab optilist segunemist kahe klassi vahel.

Samas on ka massiive, mis on üheselt klassifitseeritud. Need viitavad piirkondadele, kus mudel töötab ootuspäraselt. See tähendab, et klassifikaator suudab edukalt määrata massiivi kui ühetaolise üksuse.

Kaardil nähtav varieeruvus põllumassiivide siseselt näitab aga, et klassifitseerimismudel ei suuda alati üheselt määrata kogu massiivi katvat klassi. Ehk üks massiiv võib olla piirkonniti valesti klassifitseeritud, mis näitab mudeli piiratust ruumilise mustri mõistmisel. Selline puudujääk võib olla põhjustatud näiteks põldude looduslikkust varieeruvusest (erinevad kasvujärgud taimedel). Tulemused viitavad vajadusele täiustada mudelit, näiteks ruumiliselt teadlikuma algoritmi abil, mis oskab arvestada ka ruumilise aspektiga (nt klassifitseerimismudel, mis arvestab ka naaberpikslite väärtusi). Teine võimalus oleks täiendada treeningandmestikku nii, et see hõlmaks endas rohkem variatsiooni põldude ja taimekasvujärgude vahel.

5. Arutelu

Uurimuses saavutatud tulemused kinnitavad, et Sentinel-2 andmetel põhinev klassifitseerimine võimaldab edukalt eristada kevadisel perioodil suve- ja talvekultuure, mis vastab varasemate uurimuste järeldustele. Antud töös oli üldine klassifitseerimise täpsus (84,3%), mis jääb küll mõnevõrra alla osade varasemate uurimuste tulemusele (nt Xie & Niculescu, 2022 ja Pan et al., 2021), kuid on siiski piisavalt kõrge, et hinnata meetoodika rakendatavust Eesti kontekstis ja konkreetsetes tingimustes.

Xie ja Niculescu (2022) kasutasid hierarhilist klassifitseerimist ning kaasasid lisaks Sentinel-1 SAR-andmed, saavutades taliviljade kaardistamisel üldise täpsuse 93,2%. Antud töös kasutati vaid optilisi Sentinel-2 kujutisi aprillikuust, mil taliviljad on spektraalselt juba eristatavad, kuid suvekultuurid ei ole veel maapinnapealset taimestikku kasvatanud. Antud uurimuse tulemuste ruumiline ja temaatiline kooskõla välitöödega tõendab, et sobiv ajavalik ja hästi ettevalmistatud treeningandmestik võivad osaliselt kompenseerida mitme andmeallika puudumist.

Pan et al. (2021) tõid esile fenoloogiliste parameetrite olulisuse talvekultuuride eristamisel, kasutades terve kasvuperioodi ulatuses kogutud aegridasid. Selles töös piirduti aga ühe kuu (aprilli) komposiitkujutise kasutamisega, sest aegridade mahukus on kasutatud andmetega võrreldes väga suur.

Waldner et al. (2015) näitasid, et mitme sensori ja biofüüsikalise info kaasamine parandab klassifikatsiooni hooaja lõpus. Käesolev uurimus piirdus ainult Sentinel-2 andmetega, ent näitas siiski, et optiliste kanalite ja vegetatsiooniindeksite kombineerimine otsustusmetsa algoritmiga võimaldab usaldusväärset tulemuste saavutamist ka kevadise komposiitpildi põhjal. Tulemuste kvaliteeti suurendas ka see, et kasutati ainult madala pilvisusega kujutisi ning eemaldati pilvede ja varjude mõju CloudScore+ algoritmiga (Pasquarella et al., 2023).

Villa et al. (2015) uurisid suvekultuure ja rõhutasid SAR-andmete lisamise tähtsust, kuna need võimaldavad tuvastada põldude struktuurilisi erinevusi isegi pilvise ilmaga. Antud töö ei kasutanud SAR-andmeid, mistõttu võib oletada, et klassifitseerimise täpsust saab lisaks suurendada, eriti suvekultuuride varajases kasvufaasis (s.o küntud maa tuvastamine), SAR-signaalide kaasamisega.

Lisaväärtusena sellele tööle oli ulatuslik välitöödel kogutud valideerimisandmestik (kokku 373 punkti ja 892 pilti), mille abil kontrolliti mudeli tulemust ruumiliselt täpselt kaardistatud põllumassiivide lõikes. See võimaldas klassifikatsiooni tulemusi hinnata mitte ainult üldise täpsuse alusel, vaid ka maakattetüüpide kaupa. Eriti hästi eristus talivili, mille täpsus oli kõrge tänu kevadisele vegetatsioonile, samas kui rohumaad ja küntud maa eristamine oli keerukam. Seda kinnitasid ka veamaatriksi tulemused.

Kokkuvõtvalt võib öelda, et kuigi uurimistöös ei saavutatud sama kõrget üldist täpsust kui mõnes varasemas suuremastaabilises uurimuses, on tulemused Eesti kontekstis väga lootustandvad. See töö kinnitab Sentinel-2 andmete rakendatavust väiksemahulises piirkondlikus analüüsis, kus täpselt määratletud ajavahemik ja kvaliteetsed treeningandmed suudavad tagada märkimisväärse tulemuse. Edasistes uuringutes võiks kaasata SAR-andmeid ja fenoloogilisi aegridasid, et suurendada klassifikatsiooni stabiilsust kogu kasvuperioodi ulatuses ning parandada keerulisemate klasside eristamist.

6. Kokkuvõte

Käesolev töö käsitles Sentinel-2 satelliitandmetepõhist suve- ja talvekultuuride klassifitseerimist varakevadelisel perioodil Eesti põllumajandusmaal, keskendudes kolmele põhilisele kategooriale: taliviljad, küntud maa ja rohumaa. Klassifikatsiooni eesmärk oli hinnata masinõppe meetodite, täpsemalt otsustusmetsa algoritmi, rakendamise sobivust ning mudeli täpsust Google Earth Engine'i keskkonnas.

Treeningandmete põhjal saadud mudel näitas üldist täpsust kuni 86%, mis viitab heale mudeli võimele andmeid üldistada ja klassifitseerida erinevaid maakasutuse tüüpe ka väljaspool treeningala. Punktipõhise täpsuse hindamisel, kus mudeli tulemusi võrreldi välitööde käigus kogutud proovipunktidega, saavutati 78% üldine täpsus. Veamaatriksi analüüs näitas, et eksimused tekkisid peamiselt taliviljade ja küntud maa vahel, mis on seletatav nende spektraalse kattuvusega kevadises kasvufaasis. Samuti kattusid küntud maa ja rohumaa spektraalselt, põhjustades täiendavaid klassifikatsiooni ebatäpsusi.

Lauspinnalise täpsuse hindamisel, mis tugines rasterandmete pikslipõhisele võrdlusele, saavutati üldine täpsus 67%. Suurimad eksimused ilmnisid talivilja ja küntud maa ning rohumaa ja küntud maa vahel, mille põhjuseks võib olla maakattetüüpide spektraalne sarnasus ja treeningandmestiku piiratus. Visuaalsel analüüsil selgus, et valesti klassifitseeritud pikslid esinevad sageli põllumassiivide keskosas, kus võib esineda varieeruvust taimede kasvujärkudes, samuti massiivide servades, kus kahe klassi vahel tekib segunemine. Ometi leidis ka ühtlaselt klassifitseeritud massiive, mis viitab mudeli võimekusele teatud tingimustes.

Arutelu põhjal võib järeldada, et Sentinel-2 andmed võimaldavad varakevadelisel perioodil usaldusväärset eristada talve- ja suvikultuure, eriti kui kasutada sobivaid spektraalindekseid ja kvaliteetset treeningandmestikku. Kuigi klassifitseerimise üldine täpsus (84,3%) jäi veidi alla mõnede varasemate uuringute tulemustele, kinnitab see töö meetoodika rakendatavust Eesti tingimustes. Täpsust piiras eelkõige spektraalne sarnasus rohumaa ja küntud maa vahel, samuti SAR-andmete ja fenoloogiliste aegridade puudumine. Tulemuste kvaliteeti toetas aga kevadise ajavaliku sobivus ja ulatuslik välitöödel kogutud valideerimisandmestik. Edasistes uuringutes soovitatakse laiendada uurimisala, testida mudelit erinevates piirkondades ja aastatel, kaasata mitme sensori andmeid (nt Sentinel-1 radar) ning kasutada aegridasid, et

parandada keerukamate maakattetüüpide eristamist ja tagada klassifikatsiooni suurem stabiilsus kogu kasvuperioodi vältel.

Töö tulemused kinnitavad, et Sentinel-2 andmete ja masinõppe meetodite rakendamine on perspektiivikas lähenemine Eesti põllumajanduse maakasutuse seireks ning pakub tugevat alust edasiseks teadus- ja rakendusuringuteks.

7. Summary

Identifying summer and winter crops with Sentinel-2 data

Author: Emilie Susi

This study addressed the classification of summer and winter crops based on Sentinel-2 satellite data during the early spring period on Estonian agricultural land, focusing on three main categories: winter crops, ploughed land, and grassland. The objective of the classification was to evaluate the suitability of applying machine learning methods, specifically the Random Forest algorithm, and the accuracy of the model within the Google Earth Engine environment.

The developed model achieved an overall accuracy of up to 86%, indicating a strong capability to generalize and classify different land use types even beyond the training area. In the point-based precision assessment, where the model's results were compared with reference points collected during fieldwork, an overall accuracy of 78% was achieved. The error matrix analysis showed that most misclassifications occurred between winter crops and ploughed land, which can be explained by their spectral similarities during the spring growth phase. Spectral overlap was also observed between ploughed land and grassland, causing additional classification inaccuracies.

In the area-based precision assessment, which relied on pixel-level comparison of raster data, an overall accuracy of 67% was achieved. The most significant misclassifications occurred between winter crops and ploughed land, as well as between grassland and ploughed land. Visual analysis revealed that misclassified pixels were often located within the interior zones of agricultural fields, likely due to within-field variation in crop development stages. Misclassifications also appeared along the edges of fields, likely caused by transitional zones between land cover types, leading to mixed spectral signals. Nevertheless, several homogeneous field units were classified accurately, indicating that the model performs reliably under certain conditions. The observed within-field variability in classification results suggests that the model struggled to consistently assign a single class to entire agricultural parcels. This limitation may stem from natural variation in vegetation development and insufficient representation of such variability in the training data.

Based on the discussion, it can be concluded that Sentinel-2 data allows for reliable differentiation of winter and summer crops during the early spring period, especially when appropriate spectral indices and high-quality training data are used. Although the overall classification accuracy (84.3%) was slightly lower than in some previous studies, this work confirms the applicability of the methodology in Estonian conditions. The main limitations were the spectral similarity between grassland and ploughed land, as well as the model's limited ability to account for spatial patterns within fields. Future improvements could involve the use of spatially aware classification algorithms that incorporate neighborhood information, or the expansion of training datasets to better capture within-class variability.

For future research, it is recommended to expand the study area, test the model in different regions and years, include multi-sensor data (e.g., Sentinel-1 radar), and use time series to improve the classification of more complex land cover types and ensure greater stability throughout the growing season.

The results of this study confirm that the application of Sentinel-2 data and machine learning methods is a promising approach for monitoring land use in Estonian agriculture and provides a strong foundation for further scientific and applied research.

8. Tänuavaldus

Täna oma juhendajat, professor Evelyn Uuemaad, järjepideva toetuse, sisuka tagasiside ning väärtuslike tehniliste teadmiste jagamise eest. Samuti soovin tänada Hanna Marta Tsäkkot väliandmete kogumisel osutatud abi eest.

9. Kirjanduse loetelu

Abdelmajeed, A. Y. A., & Juszczak, R. (2024). *Challenges and Limitations of Remote Sensing Applications in Northern Peatlands: Present and Future Prospects*. *Remote Sensing*, 16(3), 591. <https://doi.org/10.3390/rs16030591>

Aherin, R., & Paulovic, J. (2025). *How to conduct a field survey using QGIS and Mergin Maps: A complete step-by-step guide*. Mergin Maps. https://merginmaps.com/blog/how-to-conduct-a-field-survey-using-qgis-and-mergin-maps-a-complete-step-by-step-guide?utm_source=chatgpt.com

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleira, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., & R. Verón, S. (2018). *Remote Sensing and Cropping Practices: A Review*. *Remote Sensing*, 10(1), 99. <https://doi.org/10.3390/rs10010099>

Belgiu, M., & Drăguț, L. (2016). *Random forest in remote sensing: A review of applications and future directions*. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

Behnke, G. D., & Villamil, M. B. (2019). *Cover crop rotations affect greenhouse gas emissions and crop production in Illinois, USA*. *Field Crops Research*, 241, 107580. <https://doi.org/10.1016/j.fcr.2019.107580>

Bousbih, S., Zribi, M., Pelletier, C., Gorrab, A., Chabaane, Z., Baghdadi, N., Aissa, N., & Mougnot, B. (2019). *Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2*. *Remote Sensing*, 11(13), 1520. <https://doi.org/10.3390/rs11131520>

Clevers, J. G. P. W., & Gitelson, A. A. (2013). *Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3*. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344–351. <https://doi.org/10.1016/j.jag.2012.10.008>

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). *Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services*. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

Eesti Entsüklopeedia. (2011). Otepää kõrgustik. https://entsyklopeedia.ee/artikkel/otep%C3%A4%C3%A4_k%C3%B5rgustik1

European Space Agency (ESA). (2015). *Sentinel-2 User Handbook*. https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). *Google Earth Engine: Planetary-scale geospatial analysis for everyone*. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

Hodkiewicz, J., & Paul, M. (2023). *Collecting field data with QGIS and Mergin Maps*. Locate Press.

Immitzer, M., Vuolo, F., & Atzberger, C. (2016). *First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe*. *Remote Sensing*, 8(3), 166. <https://doi.org/10.3390/rs8030166>

Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). *Improved Early Crop Type Identification By Joint Use of High Temporal Resolution SAR And Optical Image Time Series*. *Remote Sensing*, 8(5), 362. <https://doi.org/10.3390/rs8050362>

Kafy, A. A., Khan, M. H. H., Islam, M., Sarker, M. H. S., & Rahman, M. W. (2020). *Prediction of future land surface temperature and its impact on climate change: A remote sensing based approach in Chattogram city. 1st International Student Research Conference 2020, Dhaka University*.

Karmakar, P., Teng, S. W., Murshed, M., Pang, S., Li, Y., & Lin, H. (2024). *Crop monitoring by multimodal remote sensing: A review*. *Remote Sensing Applications: Society and Environment*, 33, 101093. <https://doi.org/10.1016/j.rsase.2023.101093>

Khan, H. R., Gillani, Z., Jamal, M. H., Athar, A., Chaudhry, M. T., Chao, H., He, Y., & Chen, M. (2023). *Early Identification of Crop Type for Smallholder Farming Systems Using Deep Learning on Time-Series Sentinel-2 Imagery*. *Sensors*, 23(4), 1779. <https://doi.org/10.3390/s23041779>

Khanal, S., KC, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). *Remote Sensing in Agriculture—Accomplishments, Limitations, and Opportunities*. *Remote Sensing*, 12(22), 3783. <https://doi.org/10.3390/rs12223783>

Kross, A., McNairn, H., Lapen, D., Sunohara, M., & Champagne, C. (2015). *Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops*. *International Journal of Applied Earth Observation and Geoinformation*, 34, 235–248. <https://doi.org/10.1016/j.jag.2014.08.002>

Liu, Y., Meng, Q., Zhang, L., & Wu, C. (2022). *NDBSI: A normalized difference bare soil index for remote sensing to improve bare soil mapping accuracy in urban and rural areas*. *CATENA*, 214, 106265. <https://doi.org/10.1016/j.catena.2022.106265>

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). *Remote sensing big data computing: Challenges and opportunities*. *Future Generation Computer Systems*, 51, 47–60. <https://doi.org/10.1016/j.future.2014.10.029>

Maa- ja Ruumiamet. (2025). Maainfo kaardirakendus. <https://xgis.maaamet.ee/xgis2/page/app/maainfo>

Mzid, N., Pignatti, S., Huang, W., & Casa, R. (2021). *An analysis of bare soil occurrence in arable croplands for remote sensing topsoil applications*. *Remote Sensing*, 13(3), 474. <https://doi.org/10.3390/rs13030474>

NASA Earthdata. (i.a.). *Remote sensing*. NASA. <https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/remote-sensing>

Newete, S., Abutaleb, K., Chirima, G., Dabrowska-Zielinska, K., & Gurdak, R. (2024). *Phenology-based winter wheat classification for crop growth monitoring using multi-temporal Sentinel-2 satellite data*. *The Egyptian Journal of Remote Sensing and Space Science*, 27, 695-704. <https://doi.org/10.1016/j.ejrs.2024.10.001>

Nguyen, T. C., Chidthaisong, A., Kieu Diem, P., & Huo, L.-Z. (2021). *A modified bare soil index to identify bare land features during agricultural fallow-period in Southeast Asia using Landsat 8*. *Land*, 10(3), 1–18. <https://doi.org/10.3390/land10030231>

Pan, L., Xia, H., Zhao, X., Guo, Y., & Qin, Y. (2021). *Mapping Winter Crops Using a Phenology Algorithm, Time-Series Sentinel-2 and Landsat-7/8 Images, and Google Earth Engine*. *Remote Sensing*, 13(13), 2510. <https://doi.org/10.3390/rs13132510>

Pasquarella, V. J., Brown, C. F., Czerwinski, W., & Rucklidge, W. J. (2023). Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2124–2134).

Põllumajanduse Registrate ja Informatsiooni Amet (PRIA). PRIA veebikaart. <https://kls.pria.ee/kaart/>

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). *Monitoring vegetation systems in the Great Plains with ERTS*. *In Third Earth Resources Technology Satellite-1 Symposium* (Vol. 1, pp. 309–317). NASA SP-351.

Sellers, P. J. (1985). *Canopy reflectance, photosynthesis and transpiration*. *International Journal of Remote Sensing*, 6(8), 1335–1372. <https://doi.org/10.1080/01431168508948283>

Space4Water. (2020). *Modified Normalized Difference Water Index (MNDWI)*. <https://www.space4water.org/taxonomy/term/1246>

Tucker, C. J. (1979). *Red and photographic infrared linear combinations for monitoring vegetation*. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)

Tugac, M., Simsek, F., & Torunlar, H. (2024). *Classification of agricultural crops with Random Forest and Support Vector Machine algorithms using Sentinel-2 and Landsat-8 images*. *International Journal of Environment and Geoinformatics*, 11(1), 106–118. <https://doi.org/10.30897/ijegeo.1479116>

U.S. Geological Survey. (i.a.). *Landsat 2*. U.S. Department of the Interior. <https://www.usgs.gov/landsat-missions/landsat-2>

Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., & Brivio, P. A. (2015). *In-Season Mapping of Crop Type with Optical and X-Band SAR Data: A Classification Tree Approach Using*

Synoptic Seasonal Features. *Remote Sensing*, 7(10), 12859-12886.
<https://doi.org/10.3390/rs71012859>

Waldner, F., Lambert, M.-J., Li, W., Weiss, M., Demarez, V., Morin, D., Marais-Sicre, C., Hagolle, O., Baret, F., & Defourny, P. (2015). *Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series*. *Remote Sensing*, 7(8), 10400-10424.
<https://doi.org/10.3390/rs70810400>

Wang, F., Yang, M., Ma, L., Zhang, T., Qin, W., Li, W., Zhang, Y., Sun, Z., Wang, Z., Li, F., & Yu, K. (2022). *Estimation of Above-Ground Biomass of Winter Wheat Based on Consumer-Grade Multi-Spectral UAV*. *Remote Sensing*, 14(5), 1251.
<https://doi.org/10.3390/rs14051251>

Wang, J., Zhang, S., Lizaga, I., Zhang, Y., Ge, X., Zhang, Z., Zhang, W., Huang, Q., & Hu, Z. (2024). *UAS-based remote sensing for agricultural monitoring: Current status and perspectives*. *Computers and Electronics in Agriculture*, 227(Part 1), 109501.
<https://doi.org/10.1016/j.compag.2024.109501>

Xie, G., & Niculescu, S. (2022). *Mapping Crop Types Using Sentinel-2 Data Machine Learning and Monitoring Crop Phenology with Sentinel-1 Backscatter Time Series in Pays de Brest, Brittany, France*. *Remote Sensing*, 14(18), 4437. <https://doi.org/10.3390/rs14184437>

Zarco-Tejada, P. J., Hubbard, N., & Loudjani, P. (2014). *Precision agriculture: An opportunity for EU farmers – potential support with the CAP 2014–2020*. European Commission Joint Research Centre (JRC) Scientific and Policy Reports.
<https://publications.jrc.ec.europa.eu/repository/handle/JRC89859>

Zayani, H., Fouad, Y., Michot, D., Kassouk, Z., Baghdadi, N., Vaudour, E., Lili-Chabaane, Z., & Walter, C. (2023). *Using Machine-Learning Algorithms to Predict Soil Organic Carbon Content from Combined Remote Sensing Imagery and Laboratory Vis-NIR Spectral Datasets*. *Remote Sensing*, 15(17), 4264. <https://doi.org/10.3390/rs15174264>

Zha, Y., Gao, J., & Ni, S. (2003). *Use of normalized difference built-up index in automatically mapping urban areas from TM imagery*. *International Journal of Remote Sensing*, 24(3), 583–594. <https://doi.org/10.1080/01431160210144570>

Zheng, B., Campbell, J. B., Serbin, G., & Galbraith, J. M. (2014). *Remote sensing of crop residue and tillage practices: Present capabilities and future prospects*. *Soil and Tillage Research*, 138, 26–34. <https://doi.org/10.1016/j.still.2013.12.009>

10. Lisad

Lisa 1. Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Emilie Susi,

1. Annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Suve- ja talvekultuuride tuvastamine Sentinel-2 andmete abil”, mille juhendaja on Prof. Evelyn Uemaa, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Emilie Susi

22.05.2025