

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

Bakalaureusetöö geograafias (12 EAP)

**Kaugseire ja masinõppe abil rohundirikaste kuusikute elupaiga
leviku ja esinduslikkuse prognoosimine Haanja Looduspargi
näitel**

Annabel Metsalu

Juhendajad: PhD Anneli Palo,
PhD Valentina Sagris

Tartu 2026

Annotatsioon

Kaugseire ja masinõppe abil rohundirikaste kuusikute elupaiga leviku ja esinduslikkuse prognoosimine Haanja Looduspargi näitel

Euroopa Liidu Loodusdirektiivi artikli 17 põhjal peavad liikmesriigid esitama iga kuue aasta järel aruande elupaigatüüpide seisundi ja kaitse kohta. Välitööd on kulukad ja aeganõudvad, mistõttu katsetatakse kaugseirel põhinevaid lahendusi. Käesoleva töö eesmärgiks oli testida, kas praegu kasutatava väliinventuuri metoodika kohaselt kogutud andmete, masinõppe meetodi ja satelliitandmete ühendamisel on võimalik prognoosida rohundirikaste kuusikute (9050) levikut ja nende esinduslikkusklassi paranemist ehk looduslikkuse suurenemist. Töös kasutati 2025. aastal Haanja looduspargis läbiviidud inventeerimiseandmeid treenimaks juhumetsa masinõppe mudelit, mille tulemusel saadi klassifitseeritud elupaigatüüpide kaart. Andmete sisuliseks valideerimiseks kasutati 2011. aastal mittemajandatavates sihtkaitsevööndites inventeeritud elupaikade andmeid. Analüüsist selgus, et välitööde, masinõppemudeli ja satelliitandmete kombineerimisel on võimalik prognoosida rohundirikaste kuusikute paiknemist ja esinduslikkust. Samas tuleb arvestada mitmete teguritega, mis võivad prognoosimise täpsust mõjutada.

Märksõnad: kaugseire, metsade kaugseire, juhumets, Natura 2000, rohundirikkad kuusikud

CERCS kood: T181 - Kaugseire

Abstract

Predicting the distribution and structural quality of Fennoscandian herb-rich forests with *Picea abies* using remote sensing and machine learning in the Haanja Nature Park

The European Union Habitats Directive (Article 17) requires Member States to submit reports every six years on the conservation status of habitats within the Natura 2000 network. Field surveys are time-consuming and costly, which has led to increasing interest in remote sensing-based approaches. The aim of this study was to evaluate whether the combination of field inventory data collected according to the currently used methodology, satellite data, and a machine learning approach enables the prediction of the distribution of herb-rich spruce forests (habitat type 9050) and changes in their representativity class, in other words an increase in naturalness. The study used field inventory data collected in 2025 in Haanja Nature Park to train a Random Forest machine learning model, resulting in a classified map of habitat types. For validation of the results, habitat data collected in 2011 from strictly protected zones without forest management were used. The analysis showed that combining field data, machine learning models, and satellite data makes it possible to predict the spatial distribution and condition of herb-rich spruce forests. However, several factors influencing prediction accuracy must be taken into account.

Keywords: remote sensing, forest remote sensing, Random Forest, Natura 2000, herb-rich spruce forests

CERCS code: T181 - Remote sensing

Sisukord

Sissejuhatus	4
1. Teoreetiline taust.....	5
1.1 Kaugseire kasutamine metsade seires	5
1.2 Kaugseire kasutamine metsade looduskaitseväärtuste hindamisel.....	6
1.3 Haanja looduspargi rohundirikkad kuusikud kui looduskaitsete ja majanduslike konfliktide näidis.....	9
2. Andmed ja metoodika	10
2.1 Uurimisala asukoht	10
2.2 Andmed.....	11
2.3 Metoodika	12
3. Tulemused ja arutelu.....	16
3.1 Mudeli täpsus ja tulemuste kaart	16
3.2 Hinnang prognooskaardi täpsusele ristvalideerimismatriksi põhjal	17
3.3 Hinnang metoodikale.....	20
Kokkuvõte	22
Summary.....	24
Tänuavaldused	26
Kasutatud kirjandus.....	27
Lisad	34

Sissejuhatus

Metsad on pidevalt muutuvad ökosüsteemid, mistõttu on ajakohased andmed ja regulaarsed andmekogumised olulised metsade seisundist ülevaate saamiseks. Traditsiooniliselt seisneb metsade kohta informatsiooni kogumine välitöodes, mis tihti on aeganõudvad ja kulukad.

Maa loodusvarade jälgimine ja analüüsimine on pärast Landsati orbiidile saatmist liikunud uuele tasemele. Kaugseire on muutunud oluliseks tööriistaks ka metsade seirel (Hernández-Clemente *et al.* 2019; Lambert *et al.* 2013), sest aitab kokku hoida aega ja kulusid ning parandada välitööde efektiivsust. Metsade kaugseiret kasutatakse mitmesugustel eesmärkidel metsamajandusest looduskaitсени ning üha kasvavaks trendiks on masinõppe meetodite põimimine kaugseire andmetega (Al Saim & Aly 2025).

Kaugseiret rakendatakse aina enam ka Natura 2000 kaitsealade võrgustiku loodusväärtuste seires (Haest *et al.* 2017). Kuna Euroopa Liidu Loodusdirektiivi artikkel 17 põhjal tuleb anda liikmesriikidel ülevaade võrgustikku kuuluvate elupaikade seisundist iga 6 aasta järel, otsitakse üha enam lahendusi, kuidas saada täpset ajakohast infot kiirelt ja madalate kuludega.

Töös käsitletud Haanja looduspargi on rohundirikaste kuusikute (9050) vastutusala Eestis. Intensiivistunud raie kaitseala piiranguvööndis viis kahtlusele, et rohundirikkaid kuusikuid on maha raiutud liiga palju ja nende levikuandmed vajavad uuendamist. Huvipakkuv on ka küsimus, kui kiiresti suureneb struktuuriline mitmekesisus sihtkaitsevööndite kuusikutes, mida enam ei majandata. Selline areng peaks kajastuma metsaelupaiga nn esinduslikkusklassi paranemises.

Käesoleva töö eesmärgiks on hinnata, kas metsaelupaikade inventeerimiseks kasutatava meetodika kohaselt kogutud välitööandmete, valitud topograafiliste andmete, kaugseireandmete ja masinõppe meetodi abil on võimalik prognoosida Loodusdirektiivi metsaelupaiga 9050 ehk rohundirikaste kuusikute levikut ja seisundit Haanja looduspargi aladel ning kui hästi tulemused vastavad varasematele välitöödel kogutud andmetele (2025 *versus* 2011).

1. Teoreetiline taust

1.1 Kaugseire kasutamine metsade seires

Kaugseireks (ingl. k *remote sensing*) nimetatakse tehnoloogiat, mis võimaldab koguda Maa süsteemide kohta andmeid distantilt (Earthdata i.a). Selleks on kasutusel erinevad satelliidid ja muud õhusõidukid (*ibid.*). Maa loodusvarade jälgimine ja analüüsimine kaugseire abil sai alguse, kui *Earth Resources Technology Satellite* (ERTS) programmi osana saadeti 23.juulil 1972 orbiidile satelliit, mille pardal olev skanner oli spetsiaalselt loodud multispektraalsete kaugseireandmete kogumiseks (Boyd & Danson 2005).

1975. aastal saadeti teele teine ERTS-B, mis nimetati pärast ümber Landsat 2-ks (USGS Landsat 2 i.a.). Landsati andmeid kasutatakse laialdaselt maakatte klassifitseerimisel, kuna Landsati satelliidid on töötanud pikaajaliselt ja andmed on tasuta kättesaadavad (Sun & Schulz 2015). Tänapäevaks on Landsati missioonid läbinud mitmeid tehnoloogilisi arengu etappe, mis on viinud andmekvaliteedi järjepideva paranemiseni ajas (USGS i.a). Lisaks on satelliidipõhine seire muutunud üha tavalisemaks ning uuringutes kasutatakse rohkemaid andmeid, mille abil saab maakatte tunnuseid ja dünaamikat täpsemalt kaardistada (Fassnacht *et al.* 2024).

Üheks olulisemaks loodusvaraks, mis on kaugseire arengust kasu saanud pärast ERTS-1 käivitamist, on metsad (Boyd & Danson 2005). Metsaressursside hindamine algas juba 20. sajandi esimesel poolel aerofotograafia abil tehtud kohaliku ulatusega metsakaardistamisega, kuid pärast satelliidi orbiidile saatmist on võimalik hinnata metsaressursse palju suurematel aladel (Boyd & Danson 2005).

Metsade kaugseire peamine eesmärk on anda täpset teavet metsade seisundi ja muutuste kohta, see aitab tõsta välitööde efektiivsust ja parandada metsa omaduste hinnanguid ning alandada kulusid (Hernández-Clemente *et al.* 2019; Lambert *et al.* 2013; Fassnacht *et al.* 2024), näiteks optimeerida metsataksatorite tööd (Lang 2022).

Ajakohased hinnangud ja regulaarsed andmekogumised on olulised, kuna metsad on pidevalt muutuvad ökosüsteemid (Coops *et al.* 2023). Metsade hetkeseisundi ülevaade (ajakohane metsainventuur) on eriti oluline metsade jätkusuutlikuks majandamiseks (Kangas & Maltamo 2006). Selleks jälgitakse metsade seisundi muutusi nii majandamisprotsesside (nt harvendusraie, uuendusraie) kui ka looduslike tegurite (nt kahjurid, põud, tulekahjud) mõjul (Coops *et al.* 2023). Satelliitandmete abil on võimalik uurida nii metsakadu kui juurdekasvu. Näiteks uurisid Peterson & Liira (2012) kaugseire abil Põhja-Euroopa metsasid. Uuringust tuli

välja, et metsasus on vaadeldavas piirkonnas vähenenud just lageraie tõttu. Tänapäeval on metsade kao ja taastuvuse dünaamika hinnang standardmeetod (Maa- ja Ruumiamet 2026a). Metsamuutuste seire sisaldab põhiliselt raiete, mingil määral ka looduslike häiringute tagajärjel hukkunud metsi, mis elupaikadena jäävad siiski väärtuslikuks kuivanud puitu ja lamapuitu vajavatele liikidele (Marchi 2018).

Metsatagavara muutust on kaugseire abil hinnanud ka Tartu Ülikooli geograafia osakonna maastikugeoinformaatika töörühm Eesti metsade alternatiivse hindamise statistilise valikuuringu raames (Maamets jt. 2023).

Vabalt kättesaadavate andmete ja täiustunud arvutusvõimsuste tõttu on kaugseires kasvavaks trendiks masinõppe meetodid (Al Saim & Aly 2025). Üheks enimkasutatavaks on kujunenud juhumetsa algoritm (ingl. k *Random Forest*) tänu oma võimekusele ja täpsusele (*ibid.*). Masinõppe meetoditega saab lihtsustada kaugseire abil kogutud radar-, optilist ja termilist infot ning parandada ekspertide järeldusvõimekust (Caffaratti *et al.* 2021). Näiteks on satelliitandmeid ja masinõpet kombineeritud hindamaks metsade süsinikuvaru ja biomassi (Nguyen & Saha 2024). Eestis on vastavalt kaugseire ja masinõppe meetodite täiustumisele uuritud okas-, sega- ja lehtmetsa eristatavust (Oviir jt. 2008), hiljem on jõutud näiteks Landsat-8 ja Sentinel-2 andmete, mullakaardi ja võrdlusproovide ning masinõppe abil puistute kaardistamiseni dominantpuuliikide alusel (Lang *et al.* 2018).

1.2 Kaugseire kasutamine metsade looduskaitseväärtuste hindamisel

Kaugseire on kujunemas oluliseks tööriistaks looduskaitstes, võimaldades koguda andmeid liikide ja elupaikade leviku kohta. Kaugseire kasutamist looduskaitstes on pidurdanud kaugseire ja looduskaitse valdkondade tehniline erinevus, samuti vajab looduskaitse praktikas palju suurema lahutusvõimega ja kohatäpseid andmeid kui kaugseire esialgu suutis pakkuda (Čahojová *et al.* 2022; Kennedy *et al.* 2009; Turner *et al.* 2003). Kuigi aja jooksul on tehnoloogia arenenud, on kaugseire kasutamine looduskaitstes endiselt keeruline, sest looduslikud kooslused on väga erineva struktuuriga ning nende kohta maapealsete andmete kättesaamine võib osutada keeruliseks ja töömahukaks (Prodromou *et al.* 2024). Samas just seetõttu, et eluslooduse seire välitööd on aeganõudvad ja kulukad, tekkiski surve arendada ja rakendada kaugseire tehnikaid (Čahojová *et al.* 2022; Feilhauer *et al.* 2014).

Kaugseire areng võimaldab koguda aina enam andmeid bioloogilise mitmekesisuse kohta, pakkudes seejuures globaalset kattuvust ja pidevaid mõõtmisi. Tehnoloogia võimaldab vaadelda ökosüsteeme, kooslusi ja suuri organisme, mõtestada ökoloogilist konteksti ning pidevalt jälgida keskkonnategureid (McMahon *et al.* 2014; Reddy, 2021). Näiteks on satelliitpiltide abil loendatud elevanthülgeid (*Mirounga leonina*) (McMahon *et al.* 2014). Eriti olulist teavet annab satelliitseire taimkatte ja üldiste maakattemuutuste kohta, sh elupaikade killustumise kohta. Zerrouk *et al.* (2025) ülevaateartikkel käsitleb viimase kümne aasta märgalade kaardistamise ja sealsete elupaikade seire temaatilisi uuringuid, kus on kasutatud kaugseiret. Uuringu tulemuste põhjal järeldatakse, et kaugseire võimaldab märgalade elupaiku tõhusalt kaardistada ja jälgida.

Kaugseiret on järjest enam rakendatud ka Natura 2000 kaitsealade võrgustiku loodusväärtuste seires, kuna Euroopa Liit vajab looduskaitsepoliitika suunamiseks võrreldavaid ning täpseid andmeid ja seiremeetodeid (Haest *et al.* 2017). Kaitsealade võrgustik rajati 1992. aastal vastuvõetud “Direktiiv looduslike elupaikade ja loodusliku fauna ning floora kaitsest” (*Council directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora*) põhjal, mis on olulisim Euroopa Liidu looduskaitsealast tegevust korraldav seadusandlik akt (lühemalt “loodusdirektiiv” ehk LD). Eriti olulised on LD I lisas loetletud elupaigatüübid, mis on loodusliku leviala piires kadumisohus ning esindavad silmapaistvalt elupaikade tüüpilisi omadusi vähemalt ühes biogeograafilises piirkonnas. Elupaigatüübid on esitatud sarnaste ökoloogiliste tingimustega rühmadena ning igale rühmale on antud originaalne kood (Keskkonnaamet 2025; Kliimaministeerium 2026).

Euroopa Liidu LD artikkel 17 põhjal tuleb liikmesriikidel iga kuue aasta järel koostada aruanne, kus tuuakse välja direktiivi rakendamine ja selle mõju elupaikadele ja liikidele (European Union 2025).

Praegu kasutatakse reeglina väliinventuure. Eestis sisestatakse inventeeritud elupaikade piirid ja kogutud andmestik Eesti looduse infosüsteemi (EELIS), andmed kasutatakse kaitse- ja hoiualade kaitsekorralduskavades ning elupaikade seires (Keskkonnaportaali 2024; Keskkonnaportaali 2026a). Elupaikade seire eesmärk on hinnata metsade struktuuri, mitmekesisust ja neid mõjutavate tegurite intensiivsust (Liira 2009). Mainitud metsaelupaikade seire välitööde tulemusi kasutatakse Natura 2000 kaitsealade võrgustiku metsaelupaikade riikliku ja Euroopa Liidule esitatava seirearuande koostamisel (Leivits 2019; Palo 2025).

Natura 2000 võrgustiku kaitsealadel olevate elupaigatüüpide leviku ja seisundi uurimiseks on viidud läbi mitmeid teadustöid. Saksamaal viidi läbi uuring, milles tutvustati elupaikade seire modelleerimise raamistikku, kasutades MODIS-i pinnapeegeldusandmeid (Sittaro *et al.* 2022), töö eesmärgiks oli hinnata väljatöötatud mudeli võimalusi ja piiranguid elupaikade seires. Elupaigatüüpide prognoosiks kasutati lisaks satelliitandmetele kolme masinõppe klassifikaatorit ning leiti, et parima arvutustäpsuse annab nende mudelite kombineerimine. Uuringu tulemustest järeldati, et pakutud meetod suudab usaldusväärselt kirjeldada looduslike alade ja elupaikade ruumilist jaotust.

Poolas uuriti kaugseire võimalusi elupaikade klassifitseerimisel (Marcinkowska-Ochtyra *et al.* 2019), keskendudes sobivaimate andmetike leidmisele. Andmetena kasutati satelliitandmeid, treeningandmetena välitööde andmeid, digitaalset kõrgusmudelit ning klassifitseerimine viidi läbi juhumetsa (*Random Forest, RF*) algoritmi abil. Sellest uuringust selgus, et mitmel ajaperioodil kogutud andmete kombineerimine tõstis klassifitseerimise täpsust. Lisaks näitasid tulemused, et topograafiliste indeksite lisamine parandas täpsust peamiselt nende elupaigatüüpide puhul, mille levik on seotud reljeefiga. Sarnaselt kasutati Loodusdirektiivi metsaelupaikade kaardistamiseks kaugseiret koos topograafiliste tunnuste ja juhumetsa algoritmiga Küprosel (Prodromou *et al.* 2024). Uuringu eesmärgiks oli koostada kaugseire andmete ja masinõppe algoritmide ühendamisel võimalikult täpne kaart metsaelupaigatüüpide levikust. Ilmnes, et kõige täpsem tulemus saadi siis, kui lisati satelliitandmetele spektraalindekseid ja topograafilisi tunnuseid. Järeldati, et kaugseire kasutamine on metsade elupaikade klassifitseerimisel juba olulise tähtsusega ning tõhus viis seireks.

Eestis on LD metsaelupaikadega tegelnud Meelis Leivits, kes kasutas masinõppe meetodit nimega *MAXENT* elupaigatüüpide esinemisalade prognoosimiseks (Leivits 2022). Leivitsa töö tulemuseks loodi koondkaardid, mis kirjeldavad metsaelupaikade potentsiaalset esinemist. Samuti selgus, et mudeli täpsuse tõstmiseks on vaja lisada täiendavaid keskkonnatunnuseid. Oma töös jõuab Leivits järeldusele, et parema täpsuse võib anda esialgne prognoos metsa kasvukohatüübi tasemel, mida saab juba hiljem kasutada elupaigatüübi määramisel täiendavaid tunnuseid kasutades.

Eelmainitud uuringud viitavad sellele, et kaugseire ja masinõppe meetodid võivad olla sobivad tööriistad elupaikade kaardistamiseks ja seireks ka Eestis.

1.3 Haanja looduspargi rohundirikkad kuusikud kui looduskaitsete ja majanduslike konfliktide näidis

Haanja looduspargi on LD metsaelupaigatüübi (9050) ehk hariliku kuusega (*Picea abies*) rohunditerikaste Fennoskandia metsade säilitamise vastutusala. Need metsad esinevad looduspargis 666 ha, mis on kogu Haanja looduspargi pindalast 22,9% (Keskkonnaportaali 2026c). See metsaelupaigatüüp esineb peeneteralistel, hea veevarustusega, toiterikastel ning pehme huumusega metsamuldadel (Paal 2000; Palo 2018). Puurindes valitseb kuusk, kuid laialehiste liikide (saar, pärn, jalaks, vaher) osatähtsus võib olla paiguti suur, iseloomulik on hästi välja kujunenud rindeline struktuur (*ibid.*). Rohunditerikkad kuusikud on väga dünaamiline elupaigatüüp, mille väljakujunemine sõltub eelnevatel aastakümnetel toimunud inimtegevusest ja looduslikest häiringutest (Paal 2000, Palo & Gimbutas 2017; Palo 2018). Seetõttu on maastikul palju kohti, kus teistest metsaelupaigatüüpidest kujunevad ajapikku rohundirikkad kuusikud, kuid ka kuusikuteks inventeeritud alad võivad loodusliku arengu tulemusel kujuneda teiseks metsaelupaigatüübiks. See tuleneb Eesti metsaelupaigatüüpide definitsioonist, mis määratlemisel võtab aluseks ka puistu koosseisu valitseva puuliigi järgi, mitte vaid mullastikul põhinevaid kasvukohatüüpe (Paal 2000, Palo & Gimbutas 2017; Palo 2018).

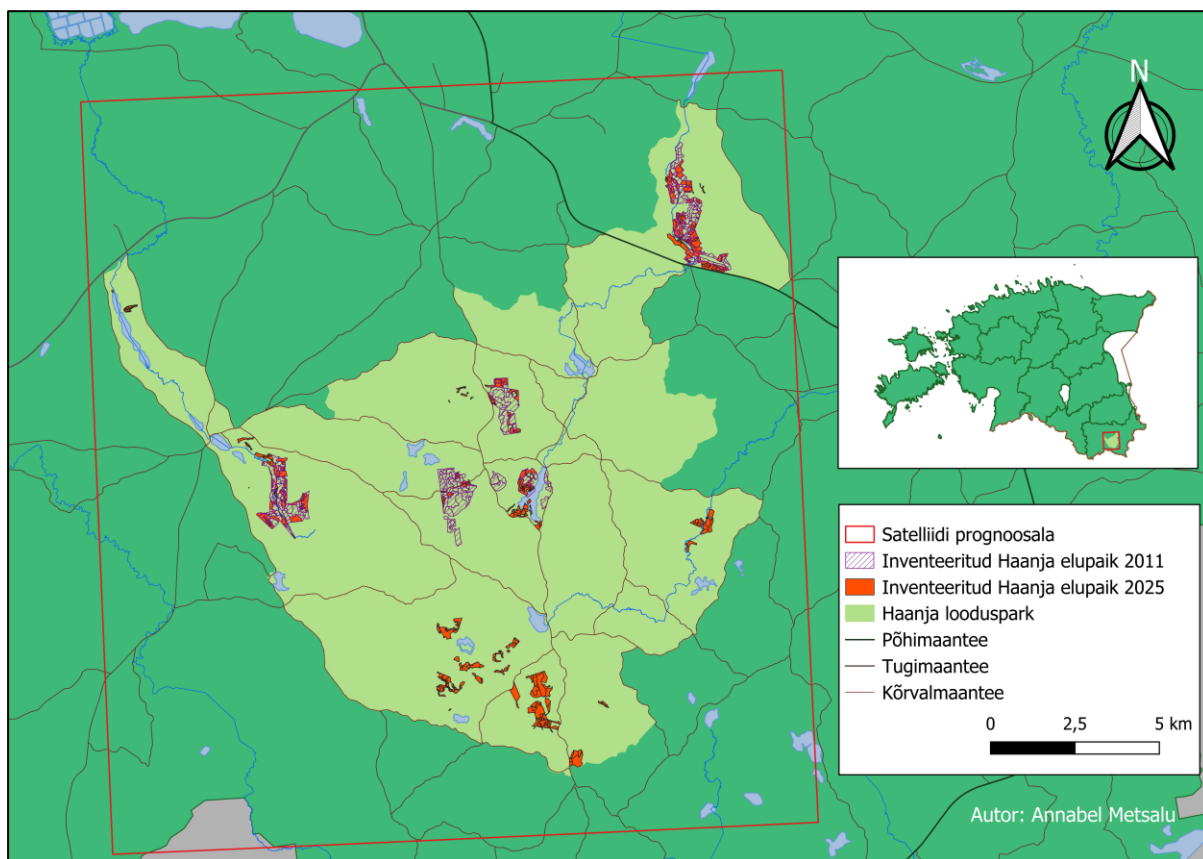
2021. aastaks oli üles tõusnud terav konflikt Haanja looduspargis toimuvate raiete üle. Looduskaitsejad väitsid, et puudub ülevaade metsaelupaikade asukohast ja hulgast piiranguvööndis ning et enamuse ongi juba maha raiutud (EKO 2021; EKO 2022). Kaugseire osutas võimalusele, et Haanjas on elupaigatüüpi säilinud olulisel määral, aga metsamuutuste andmekiht näitas ka ulatuslikke raieid (Leivits 2022; Maa- ja Ruumiamet 2026a).

Euroopa Komisjon algatas 2021. aastal Eesti riigi suhtes rikkumismenetlust, mis heidab ette kergekäelist raielubade andmist metsaelupaikadesse. Menetluse käigus peatas Keskkonnaamet raietegevuse Natura 2000 aladel esialgu 28 kuuks, et ära hoida negatiivne mõju elupaikadele (Keskkonnaamet 2022). Keskkonnaamet käivitas metsaelupaikade inventuurid piiranguvööndites (Harju 2021). Varem katsid metsaelupaikade inventuurid vaid Haanja looduspargi sihtkaitsevööndeid ja üksikuid alasid, mida hiljem siiski ei tsoneeritud sihtkaitsevööndiks. Haanjas inventeeriti metsaelupaiku aastail 2022-2025 (Palo 2022; Palo & Sell 2024; Palo 2025).

2. Andmed ja metoodika

2.1 Uurimisala asukoht

Uurimustöös kasutatavad 2011. ja 2025. aasta metsaelupaikade inventuurid on läbi viidud Haanja looduspargis (edaspidi LP, joonis 1). Satelliidi prognoosala piirati Haanja LP ümbrusega. Tulemuste visualiseerimine ainult huvialal aitas vähendada töödeldavate andmete mahtu ja kiirendada genereerimisprotsessi.



Joonis 1. Uurimisala kaart.

Haanja LP eellasena võeti 1957.a. kaitse alla maastikulised üksikobjektid Suur-Munamägi koos Vaskna ja Tuuljärvega, Vällamägi koos Küla-, Pera- ja Vahejärvega ning Rõuge järvestik (EELIS 2024). Nende baasil moodustati Haanja maastikukaitseala Eesti NSV Ministrite Nõukogu määrusega 1979.a. (Soesoo & Hade 2005). Samades piirides kinnitati 1991.a. keskkonnaministri määrusega Haanja LP kaitse-eeskiri, eesmärgiks kaitsta ja säilitada Haanja kõrgustiku omapärast reljeefi ja väljakujunenud pärandkultuurmaastikku (Haanja LP kaitse-eeskiri; Soesoo & Hade 2005). Alates 2004. aastast kuulub Haanja LP Natura 2000 kaitsealade võrgustikku, olles tervikuna nii linnu- kui loodushoiualade nimekirjas (European Environment Agency i.a; Soesoo & Hade 2005). Haanja LP viimane kaitse-eeskiri kinnitati 2015. aastal. Selle järgi on kaitse-eesmärgiks kaitsta, säilitada ja tutvustada Eesti kõrgeimat kuhjelist

saarkõrgustikku, looduslikku mitmekesisust ja esinduslikku maastikku, samuti kaitsta loodusdirektiivis nimetatud liikide elupaiku ja säilitada ning taastada elupaiku (EELIS 2024). Kokku on alal inventeeritud 30 loodusdirektiivi I lisas nimetatud elupaigatüüpi. Haanja LP pindala on 17 039,6 ha, millest 15 913,7 on piiranguvöönd ja 1125,9 sihtkaitsevöönd (Keskkonnaportaali 2026c).

2.2 Andmed

Lähteandmetena kasutati Anneli Palo koostatud Haanja LP elupaikade inventuuri andmestikku, mis on kogutud 2011. ja 2025. aastal. Välikaardistusel kasutati Keskkonnaameti edastatud andmekihte ja ortofotot. Läbitud elupaikade ja ka mitte-elupaikade piirid digiteeriti Mapinfo programmis, mis on ID kaudu ühenduv MS Exceli tabeliga. Tabel sisaldab inventuurimetoodika kohaselt kogutud elupaikade tunnuseid (Palo 2011; Palo 2025). Käesolevas töös valiti tunnuste seast elupaigatüübi, puistu üldistatud koosseisu (okasmets, segamets, lehtmets) ja esinduslikkuse (A,B,C,p, 0-elupaik) andmed.

Metsade esinduslikkust hinnatakse metsaelupaikade inventeerimise juhendi põhjal kolme tasemega: A – väga esinduslik; B – esinduslik; C – keskmine, arvestatav esinduslikkus. Mida paremini vastab mets tüübikirjeldusele ja mida rohkem elupaigatüübile omaseid metsa struktuurielemente kohapeal leitakse, seda kõrgem on elupaiga esinduslikkuse hinne (Palo 2018). Kuna kõrge esinduslikkusega metsaelupaikade range kaitse on tulemuslikum sidusal metsamaastikul, tuleb inventeerida ka madalama esinduslikkusega metsaelupaigad (C) ja taastumispotentsiaaliga alad ehk potentsiaalsed metsaelupaigad (p) (*ibid.*), sest potentsiaalsed metsaelupaigad aitavad kavandada tõhusa kaitsevööndite tsoneeringu. Maharaiatud metsad või kirjeldusele muul põhjusel mittevastavad metsad ei ole metsaelupaigad, neid tähistatakse kokkuleppeliselt Keskkonnaametiga 0-elupaikadena.

Keskkonnamuutujatena kasutati metsamaski, metsade kõrguskaardi ja mullastiku andmestikke. (1) Metsamask esitab binaarselt rasterkaardina metsamaa ja mitte-metsamaa ruumilist paiknemist (Keskkonnaportaali 2026b). Metsamaski koostamisel kasutati masinõppe metoodikat ja selle ennustustäpsuseks saavutati statistilise metsainventuuri (SMI) andmetega võrdlemisel 97,4% (*ibid.*). (2) Lisaks on kasutatud Keskkonnaagentuuri (2025) poolt välja antud Eesti metsade kõrguskaarti, mis kirjeldab puistu ülarinde keskmist kõrgust pikslitäpsusega 10 meetrit. Iga aasta mõõdistab MaRu ühe veerandi Eesti territooriumist (Keskkonnaagentuur 2025). Metsade kõrguskaart on kokku pandud Kirde-, Edela-, Loode- ja

Kagu-Eesti eri aastate mõõdistustest (viimase jaoks kasutati 2021. aasta lidariandmeid) (*ibid.*). (3) Mullastiku andmed pärinevad EstSoil-EH andmebaasist (Kmoch *et al.* 2020). Andmestiku jaoks sünteesiti Eesti kohta üle 20 ökohüdroloogilise muutuja numbriliste ja kategooriliste parameetri Eesti mullakaardist (1:10 000), 5m resolutsiooniga kõrgusmudelist, Eesti topograafilisest andmekogust ja EU-HydroSoilGrids kihtidest (Kmoch *et al.* 2020). Igale mullatüübile omistati mullatüüp, tekstuuriklass, mullaprofiilid (kihid, sügavused), tekstuur, kivimisisaldus ja füüsikalised muutujad, mis on seotud vee- ja süsiniku ringega (Kmoch *et al.* 2021). Mullakaardi infost kasutati töö jaoks mullatüüpe.

Satelliitandmetena kasutati Landsat 8 ja 9 (USGS i.a.) kolme komposiitpildikogu (ingl. k *Composite Image Collections*), mis on kättesaadavad Google Earth Engine (GEE) platvormi kaudu (Earth Engine Data Catalog i.a). GEE on veebipõhine platvorm, mis majutab, töötleb ja annab ligipääsu kaugseire andemarhiividele ning pakub kasutajatele tööriistu pildikogudest päringute ja analüüsi tegemiseks (Google Earth Engine i.a). Platvormi kasutamise eeliseks on see, et vajalikud töötused on võimalik läbi viia ilma andmeid kasutaja arvutisse alla laadimata. Komposiitpildikogu on juba eeltöödeldud “toode” (ingl. k *product*), mis võtab kokku kindla ajaperioodi (8, 16 või 32-päeva). Sõltuvalt tootest, säilitatakse piksliväärtustes kas maapinna kõige esinduslikum signaal (nt taimkatteindeksite puhul maksimaalne roheline ajaperioodil) või ajaperioodi keskmine (Earth Engine Data Catalog i.a). Selline lähenemine võimaldab eemaldada pilvedest mõjutatud pikseleid ning vähendada peegeldussignaalis esinevaid äärmusi ja anomaaliaid.

Töös kasutati järgmisi 32-päeva komposiitpildikogusid:

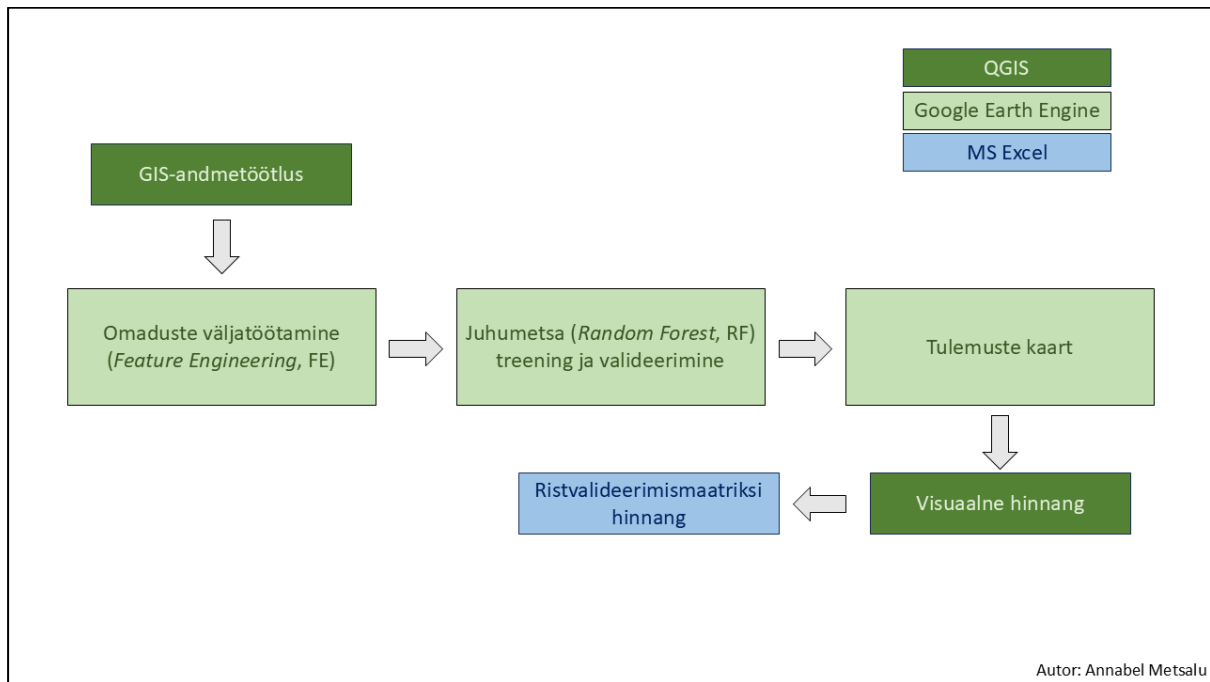
- LANDSAT/COMPOSITES/C02/T1_L2_32DAY_NDVI
- LANDSAT/COMPOSITES/C02/T1_L2_32DAY_EVI
- LANDSAT/COMPOSITES/C02/T1_L2_32DAY

Esimesed komposiidid sisaldavad juba töödeldud taimkatteindekseid, vastavalt Normaliseeritud Vegetatsiooniindeksit (NDVI) ja Täiustatud Vegetatsiooniindeksit (EVI). Viimane komposiit sisaldab pinnapeegelduse kanaleid, milleks on *blue*, *green*, *red*, *nir*, *swirl*, *swirl2* ja *thermal* (Earth Engine Data Catalog i.a).

2.3 Metoodika

Uurimistöös kasutatud metoodika üldistatud skeem on toodud joonisel 2.

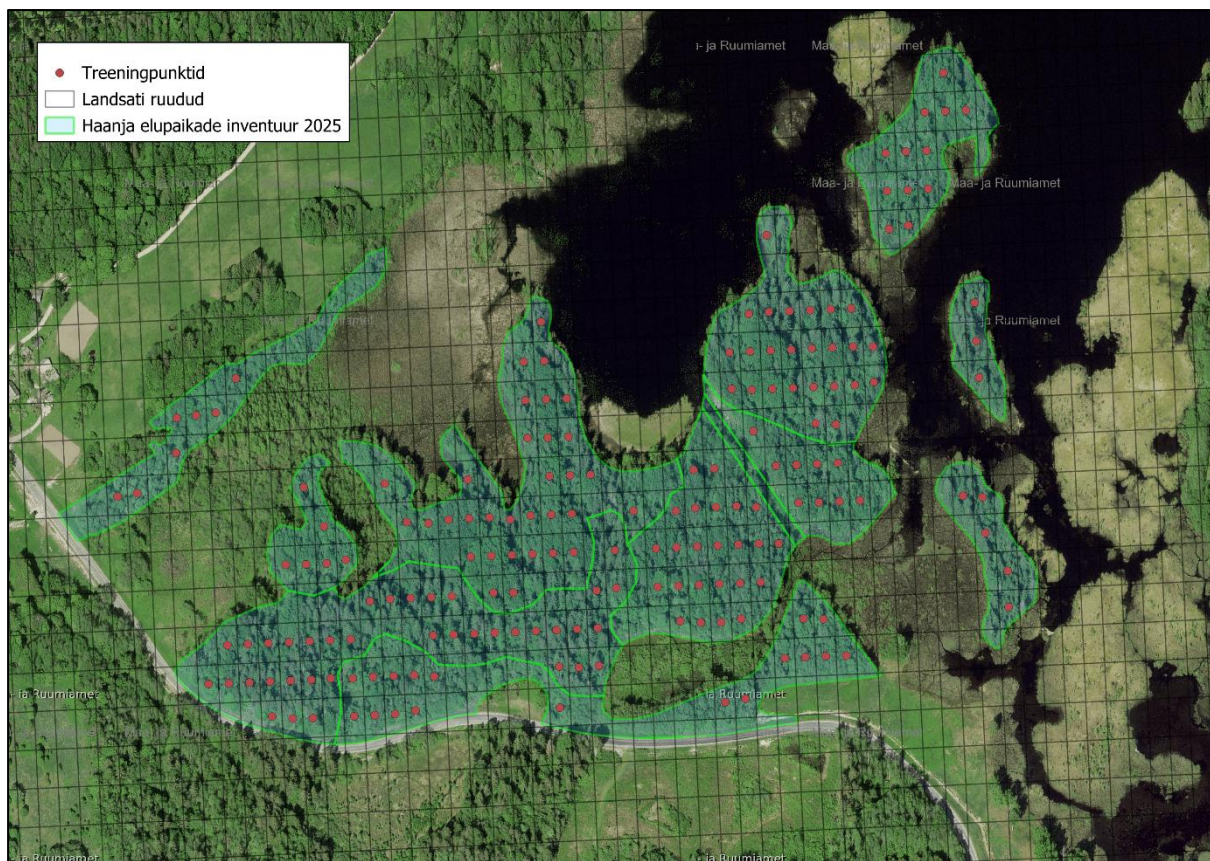
Masinõppe mudeli treenimiseks Google Earth Engine platvormil valmistati QGIS-i (QGIS i.a) tarkvaraga ette treeningpunktid. GIS-andmete töötlust alustati kaardistatud metsaalade atribuutide korrastamisega, kui liideti omavahel metsa elupaigatüüp, esinduslikkus ja puistu andmed ühte veergu. Lisaks üldistati metsade elupaigatüüpe sarnaste tunnuste abil ja eemaldati üleliigsed kirjavahemärgid infomüra vähendamiseks. Selleks, et eristada numbriliselt igat kombinatsiooni, loodi neile unikaalne ID.



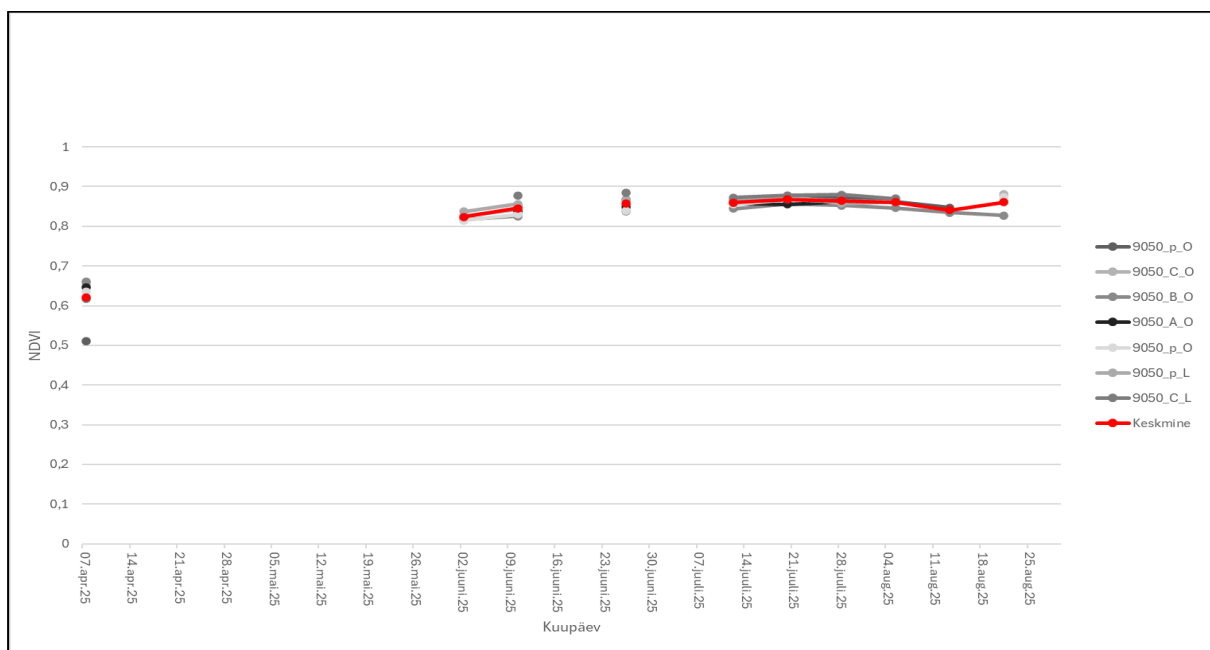
Joonis 2. Metoodika lihtsustatud ülevaateskeem.

Edasi tehti inventeeritud alade polügoonidest punktid. Selleks laaditi esmalt QGIS-i üles Landsati ruudustiku rasterkaart, mis eksporditi polügoonideks *Raster pixels to polygons* tööriistaga. Landsati ruudustikust lõigati välja *Intersectioni* tööriista abil metsade alad. Igast piksli polügoonist moodustati treeningpunkt *Centroids* tööriista abil, mis loob punkti polügooni keskkoha. Välja valiti ainult need polügoonid, milles metsaalade katvus oli üle 80% (joonis 3).

Seejärel viidi läbi taimkatteindeksite tundlikkuse analüüs ning uuriti NDVI ja EVI tundlikkust metsaelupaigatüüpide ja nende tunnuste eristatavuse suhtes. Näiteks joonisel 4 on ilmne, et NDVI puhul on 9050 eritunnuste eristatavus väike. Sarnane tulemus saadi ka EVI analüüsimisel.



Joonis 3. Treeningpunktide valiku näidis 2025. aasta Haanja inventuuri andmete ja Landsat ruudustiku põhjal. Taustaks ortofoto (Maa- ja Ruumiamet 2026b).



Joonis 4. Rohundirikaste kuusikute (9050) metsaelupaigatüübi eristatavus NDVI abil 2025. aasta andmete põhjal (GEE). A,B,C,p – esinduslikkusklassid; O – okasmetsad, L – lehtmetsad.

Järgnevalt leiti treeningpunktidele keskmine metsakõrgus, mis arvutati taimkatte kõrguse rasterkaardi abil *Zonal Statistics* tööriistaga. Lisaks omistati punktidele mullainfo *Location by attributes* tööriista abil. Sellist andmete kombineerimist nimetatakse omaduste väljatöötamiseks (ingl. k *Feature Engineering*), mille käigus valitakse, luuakse ja teisendatakse toorandmeid, et parandada õppealgoritmide toimivust (Eldeeb 2024). Selleks, et eristatavus oleks tõenäolisem ja parem, võeti kasutusele lisaks ka LANDSAT/COMPOSITES/C02/T1_L2_32DAY pinnapeegelduse kanalid. Sama eesmärgiga lisati mudelisse ka taimkatte kõrguse ja mullastiku andmed.

Mudeli treenimiseks ja hindamiseks laaditi GIS-töödeldud treeningpunktid vektorkihina GEE keskkonda (vt koodi lisa 3 või [siit](#)). Lisaks laaditi täiendavate sisendandmetena üles metsamask, metsade kõrguskaart ja mullakaart. Treeningpunktid seoti nende ruumiandmete väärtustega GEE funktsiooni *sampleRegions* abil. Seejärel jagati punktid juhusliku valiku alusel treening- ja valideerimisandmeteks vastavalt 70% ja 30%. Andmestiku jagamine kaheks osaks võimaldab mudelit esmalt treenida ning seejärel hinnata täpsust valideerimisandmete abil (Ismayilova *et al.* 2020).

Elupaigatüüpide klassifitseerimiseks kasutati juhumetsa masinõppe meetodit (*Scikit learn* i.a.). Juhumets on tõhus meetod tegemaks täpseid ennustusi ja sobib hästi suurte andmemahitudega töötamiseks (Salman *et al.* 2024). Meetod analüüsib üksikute tunnuste olulisust ning on seda täpsem, mida suurem on üksikute klassifikaatorite mitmekesisus (Janitza & Hornung 2018; Salman *et al.* 2024). Käesolevas töös treeniti juhumetsa treeningandmete põhjal, kasutades 300 otsustuspuud. Iga otsustussõlme juures kaaluti 2 juhuslikku tunnust. Mudeli treenimise alguses oli meil 38 erinevat tunnust. Juhumetsa algoritm annab hinnangu iga tunnuse mõju (ingl. k *Feature Importance*) kohta. Tunnuste täpsused on välja toodud lisa 1 ning analüüsi jaoks kasutati nendest kõige suurema mõjuga tunnuseid.

Treenitud juhumetsa mudelit rakendati kogu uurimisalale, mille tulemuseks saadi elupaigatüüpide klassifitseeritud kaart. Valminud kaardi sisukust hinnati visuaalselt QGIS-is, kus saadud rasterkaardile lisati elupaigatüüpe loogiliselt eristavad värvid ja legend, seejärel kontrolliti tulemuse looduskaitselist sisukust ristvalideerimismatriksi abil. Selleks analüüsiti tunnuste kattuvust 2011.a välitööandmetel põhineva vektorkaardi ja 2025. a. andmete põhjal genereeritud prognooskaardi võrdlemisel.

3. Tulemused ja arutelu

3.1 Mudeli täpsus ja tulemuste kaart

Masinõppe treeningu tulemusena saadi üldiseks täpsuseks 97% (*training overall accuracy*). Valideerimisel saavutati üldiseks täpsuseks 82% (*validation overall accuracy*) ja kappa-kordajaks 0,80 (*kappa*), mis viitab mudeli edukale treeningule, mille käigus tuli välja ka tunnuste olulisus. Lisa 1 graafik näitab, et põhilised taimkatteindeksid juhumetsa arvutuste kohaselt ei osutunud kõige olulisemateks (lisa 1). Üllatuseks osutus kõige suurema mõjuga omaduseks *thermal* kanal. Sellele järgnesid NDVI, kõrgusandmed ja *blue* kanal (lisa 1). Huvitav on see, et kuigi EVI on NDVI-st täiustatum ja täpsem (GeoAI 2023), osutus siin uuringus masina arvutuste põhjal täpsemaks siiski NDVI.

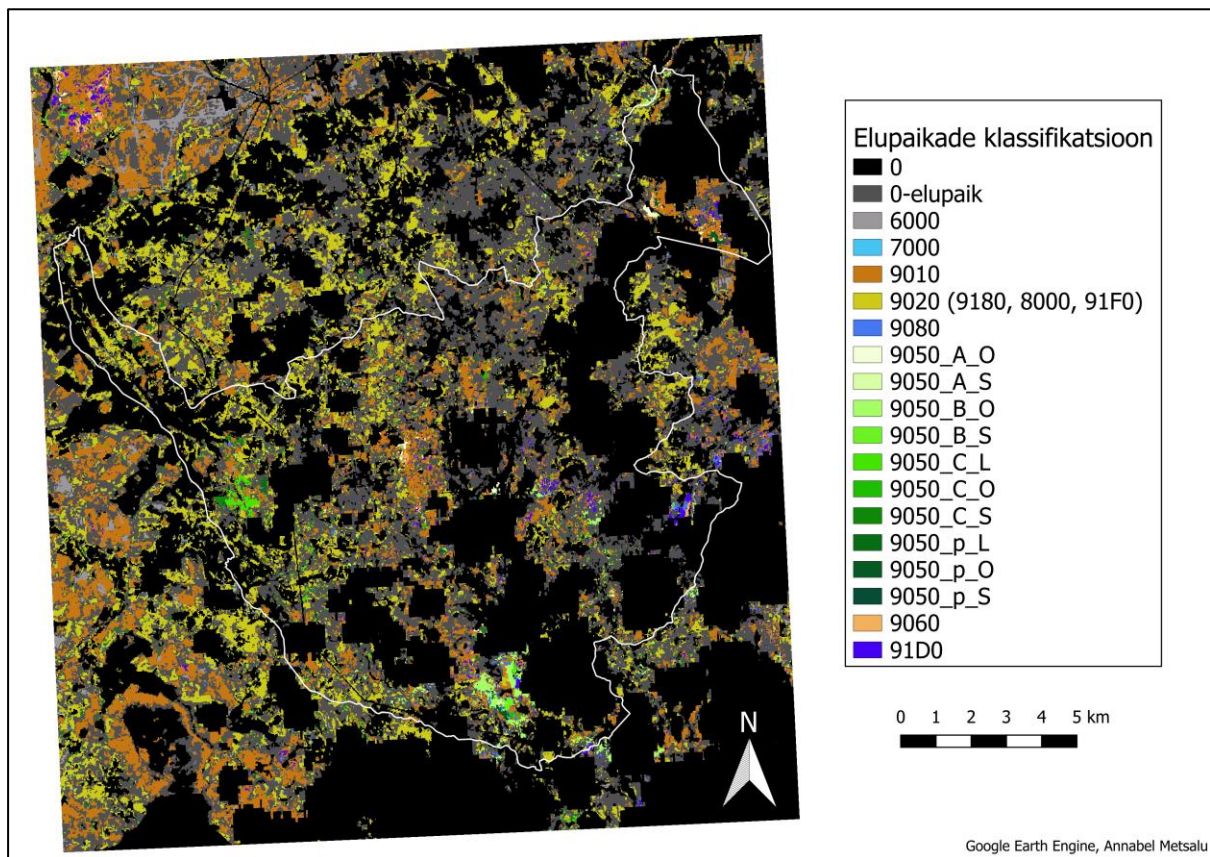
Kõrgusandmete kõrge olulisus on kooskõlas varasemate uuringutega, mille kohaselt parandab taimkatte kõrguse lisamine klassifikatsiooni täpsust (Sankey *et al.* 2017).

Blue kanal (sinine värv) oli oluline keskkonnamuutuja ka Meelis Leivitsa töös (Leivits 2022). See viitab sellele, et *blue* kanalil on kindel roll Eesti elupaikade kaugseires.

Kõige madalamaks prognoosijaks osutusid mullastiku andmed, mistõttu need jäeti lõpliku prognooskaardi genereerimisel kõrvale.

Kaardi prognoosimisel esitati treeningpunktide elupaigatüübi põhjal. Tüübi 9050 puhul kasutati lisaks talle omistatud esinduslikkusklassi ja üldistatud puistu koosseisutunnust kujul: 9050_A_O (9050 – rohundirikkad kuusikud; A – esinduslikkus; O – okasmets). Visuaalne analüüs näitas pärast seda, kui elupaigatüüpidele olid omistatud tüüpide ja klasside sarnasust/erinevust esiletõstvad värvid, et neid on võimalik eristada (joonis 5).

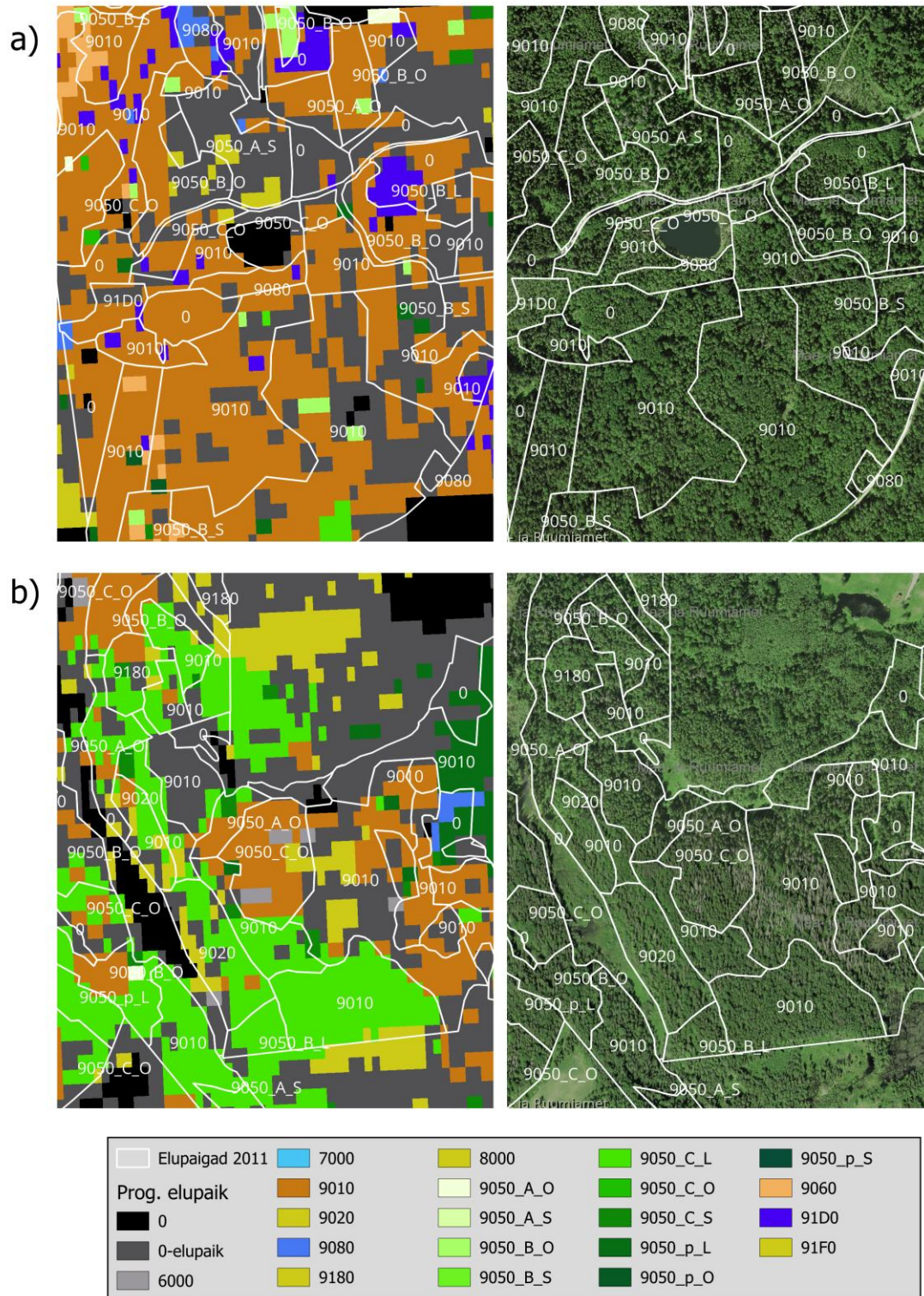
Visuaalsel hinnangul on kaardil (joonis 5) palju musta ala ehk mudel sinna metsa ei prognoosinud, eriti palju on metsata ala Haanja looduspargi idaosas. Visuaalselt jäävad elupaigatüüpidest silma 0-elupaik, 9010 ja 9020. Ka 9050 elupaigad ning nende erinevad esinduslikkusklassid paistavad välja, kuid nende levik on piiratum. Märjemad alad (7000, 9080, 91D0) on kaardil selgelt eristunud ja külgnevad üksteisega. Neid on üldiselt prognoositud sinna, kus esineb ka rohkem mitte-metsa. Väike osa prognoosialast jääb Läti piiridesse, kuhu sisendandmed ei ulatunud. See piirkond on samuti esitatud mitte-elupaigana. Edasi uuriti andmestikku ristvalideerimismatriksi abil.



Joonis 5. Google Earth Engines valminud elupaikade prognooskaart kõikide eristamistunnuste põhjal. 0 – mets puudub, 0-elupaik – mets, mis ei vasta metaelupaiga definitsioonile, 6000 – niidud; 7000 – sood; 9010 – vanad loodusmetsad; 9020 – vanad laialehised metsad; 9180 – rusukallete ja jäärakute metsad; 8000 – paljandid ja koopad, 91F0 – laialehised lammimetsad, 9080 – soostuvad ja soo-lehtmetsad, 9050 – rohunditerikkad kuusikud 9060 – okasmetsad oosidel ja moreenikuhjatistel; 91D0 – siirdesoo- ja rabametsad; A,B,C,p – esinduslikkusklassid; O – okasmetsad, S – segametsad, L – lehtmetsad.

3.2 Hinnang prognooskaardi täpsusele ristvalideerimismatriksi põhjal

Mudeli prognoosimisvõimest täpsema ülevaate saamiseks koostati MS Excelis 2011.a. välitööandmete ja 2025.a. prognooskaardi andmete ristvalideerimismatriks (lisa 2) ja vaadeldi kattuvust kaardil (joonis 6).



Joonis. 6 Mudeli kattuvus 2011.a. metsaelupaikade polügoonidega a) Tuhkrijärve sihtkaitsevööndi 2011.a. elupaigad koos prognooskaardiga, b) Jürihani sihtkaitsevööndi 2011.a. elupaigad koos prognooskaardiga.

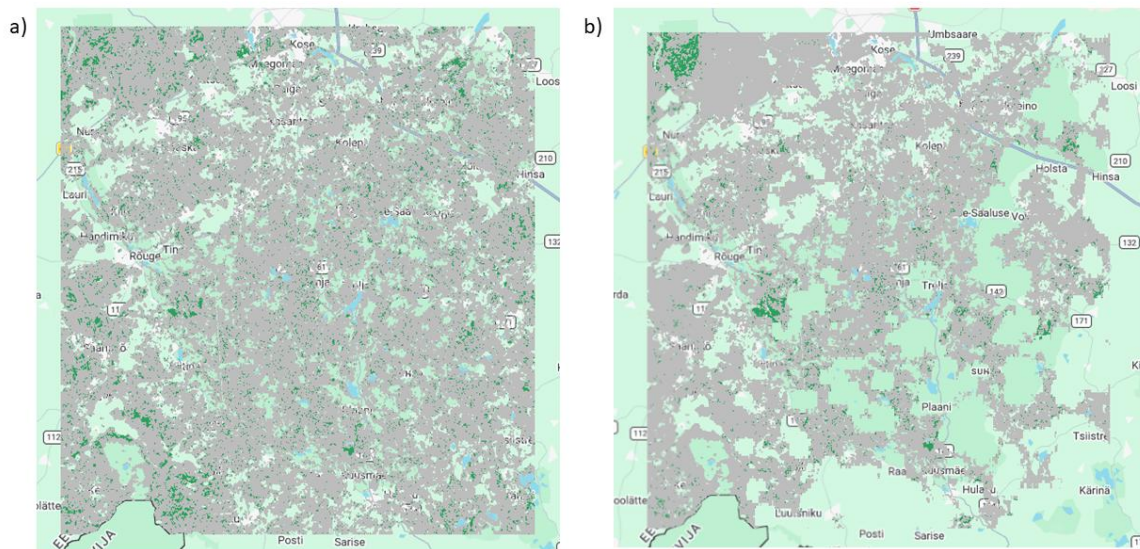
Nagu lisa 2 olevas ristvalideerimismatriksi tabelis näha, suutis mudel sarnaselt inventeerija hinnanguga prognoosida üpris vähesel määral (tabelite diagonaalid näitavad täpseid kattuvaid rastrapunkte ja osakaalusid). Üle 50% prognoositi samaks üksuseks ainult 9050 B esinduslikkusega okasmetsi (56,8%). Okasmetsad olid ilmselt paremini äratuntavad, sest prognoositi õigeks veel 12,2% 9050 A ja 8,2% C esinduslikkusega okasmetsad.

Sarnaselt inventeerijale määrati 0-elupaiku 24,2%, vanu loodusmetsi ehk 9010 36,9%, laialehiseid metsi ehk 9020 34,6%, soo-lehtmetsi ehk 9080 17,0% ja vähesel määral rabametsi ehk 91D0 0,6% (lisa 2). Vähese tuvastamise põhjuseks võib olla erinevate elupaikade liigselt detailne eristamine, mistõttu mudelil on keerulisem vahet teha. Kahjuks mõned täpsemad elupaigatüüpide klassid 2025.a. kaardistuses üldse puudusid, mis selgus alles ristvalideerimismatriksist (nendeks olid 9050_A_S, 9050_B_S). Seetõttu ei ole neid ka prognoositud ja osa 2011.a. vastavast alast on prognoositud lähedastesse klassidesse või 0-elupaigaks.

Tervikuna prognoositi metsaelupaigatüübi 9050 1514st rastrapunktist samaks tüübiks 619 ehk 43%. Kui lisada 259 punkti, mis 2011.a. olid 0-elupaigad, aga teoreetiliselt võisid 14 aastat hiljem vastata p või C esinduslikkusklassile, siis tõuseb prognoositäpsus 58%-le. Teised mineraalmuldade metsaelupaigatüübid, mille eristamine 9050-st on paiguti problemaatiline ja sõltub valitsevast peapuuliigist (Palo 2018 järgi), prognoositakse 9050-ks veel 609 punktis, osa sellest võib puistute sisemise varieeruvuse tõttu olla õige.

Murettekitav on 2025.a. välitöödel kontrollitud 0-elupaikade ehk mitte-elupaikade põhjal saadud prognoosi kattumine 2011.a. esinduslike metsaelupaikadega (joonis 6). Need asuvad kõik Haanja looduspargi sihtkaitsevööndis ja seal ei ole toimunud raieid ega muid protsesse, mis metsaelupaikade kvaliteeti halvendaksid. Ometi on 76% 0-elupaigaks prognoositud aladest tegelikult metsaelupaigad, üldisest punktide arvust moodustab see 16%. Põhjusi tuleb uurida.

Ristvalideerimismatriksist (lisa 2) on näha, et 30,4% punktidest, (22,3% 9050 elupaigatüübist) on jäetud prognoosimata, kuigi 2011.a. andmed metsade kohta olid seal olemas. Saadud prognooskaarti metsamaskiga kattuma pannes tuli välja, et viga tulenes *thermal* kanalist, mis kanalite eraldi uurimisel osutus osaliselt auklikuks ehk satelliitpilt ei olnud täielik (joonis 7). Kui prognoositud kaarti ja *thermal* kanali prognoosi omavahel kontrolliti, kattusid prognoosimata jäetud alad *thermal* kanali aukudega, mis viitab sellele, et see ongi nii suure andmekao põhjus.



Joonis 7. a) prognoos *thermal* kanalit kasutamata; b) prognoos *thermal* kanalit kasutades (väljavõtted Google Earth Engine).

Sellegipoolest jäeti *thermal* kanal prognooskaardi koostamisel sisse, sest see oli kõige täpsema ennustusvõimega (lisa 1). Kui oleks *thermal* kanali eemaldanud, oleks küll saanud suurema metsaalaga prognooskaardi, kuid ennustus täpsus muudel aladel oleks olnud oluliselt madalam.

3.3 Hinnang metoodikale

Uurimuse tulemused viitavad, et elupaigatüüpide inventeerimisandmete põhjal on satelliitandmete ja masinõppe abil võimalik saada klassifitseeritud elupaigatüüpide kaart. Töö käigus saavutatud mudeli täpsus osutus väga kõrgeks, kuid tulemuste võrdlemisel sõltumatute kaardistusandmetega ilmnas mitmeid tegureid, millele peaks tähelepanu pöörama.

Suurimaks puuduseks osutus satelliitandmete mittetäielikkus, mis tähendab seda et parima prognoosivõimega *thermal* kanal osutus auklikuks ehk satelliitpilt ei olnud täielik. Seda saaks ilmselt parandada teist ajaperioodi valides. Selles uurimuses kattus valitud periood 2025.a. välikaardistuse teostamise ajaga ehk augustist oktoobrini, vähempilvised on aga kevadsuvised ajad (Eerme & Aun 2012). Pilvistel aladel on tegelikult meid huvitavad metsaelupaigatüübid olemas, kuid need kaardil ei kajastu. Kui soovitakse saada ülevaadet, kui palju vastavat metsaelupaigatüüpi Haanja looduspargis on, tekiks pindalalistes arvutustes seetõttu sisse viga.

Positiivsena võib esile tuua, et kahte enimlevinud metsaelupaigatüüpi 9010 ja 9050 omavahel eristamata saavutati tervikuna rahuldav valideerimistäpsus seal, kus nii 2011 kui 2025.a. andmed olid olemas. Jättes tähelepanuta ka kõik nende puistutüübid ja esinduslikkusklassid, samuti lugedes õigeks nn juurde tekkinud elupaigad sihtkaitsevööndites seal, kus 14 aastat tagasi elupaika ei olnud, saavutati usutavad tulemused 94%-l prognooskaardi punktidest.

On mitmeid võimalusi prognooskaardi täpsust mõjutavaid tegureid uurida ja selle looduskaitselist ehk praktilist täpsust tõsta. Haanja looduspargis on metsaelupaiku inventeeritud ka aastail 2022-2024 väga suurel pindalal (Palo 2022; Palo & Sell 2024), kuid käesolevas töös kasutati eelkõige 2011. ja 2025.a. andmeid, sest need pärinesid samadelt või külgnevatelt aladelt, mis üldjuhul tõstab kaugseire abil saadud prognoosi täpsust tunduvalt (Karasiak *et al.* 2022). Linder *et al.* 2010 sõnul võib prognoosimistäpsus tunduda sel juhul küll kõrge, kuid kaugemate alade prognoosimisel ei pruugi tulemus olla ületreenimise tõttu nii usaldusväärne. Selles töös seati eesmärgiks saada võimalikult täpne Haanja looduspargi rohundirikaste kuusikute kaart.

Suurema andmehulga kasutamine võimaldaks edaspidi ka uurida, kas inventeerimise kuu ja satelliitandmete aeg peavad kattuma, näiteks võib leht-, segametsade ja okasmetsade eristamisel palju rohkem kasu olla varakevadistest satelliitpiltidest. Ka Lang *et al.* (2018) on kasutanud eri aegadel kogutud satelliitpilte, kuna need annavad täiendavat fenoloogilist infot paremaks puuliigilise koosseisu kaardistamiseks. Uurimist vajaks veel ka aastatevaheline varieeruvus, sest pikem aegrida võib anda paremaid tulemusi (Marcinkowska-Ochtyra *et al.* 2019). Kuna Haanja kohta on nüüd olemas aastaist 2022-2025 suur hulk andmeid, oleks mitme aasta andmete põhjal saadud täpseim prognooskaart edaspidi abiks metsaelupaigatüüpide leviku ja võib-olla ka esinduslikkusklassi seirel, olles selle perioodi kohta teadaolevaks parimaks ülepinnaliseks hinnanguks, millega uusi prognooskaarte võrrelda.

Kokkuvõte

Kaugseire on osutunud metsade uurimisel üheks olulisemaks töövahendiks, seda tänu võimalusele uurida suuri alasid väiksemate kuludega ja kiiremini. Kaugseire on tugevalt põimunud masinõppe meetoditega, mis tõstavad tulemuste täpsust ja aitavad jõuda järeldusteni.

Käesoleva töö eesmärgiks oli uurida, kas metsaelupaikade inventeerimiseks kasutatav meetodika kohaselt kogutud välitööandmete, valitud topograafiliste andmete, kaugseireandmete ja masinõppe meetodi abil on võimalik prognoosida Loodusdirektiivi metsaelupaiga 9050 ehk rohundirikaste kuusikute levikut ja seisundit Haanja looduspargi aladel ning kui hästi tulemused vastavad varasematele välitöödel kogutud andmetele (2025 *versus* 2011).

Selleks loodi treeningpunktid, mis treenisid juhumetsa masinõppe algoritmi, mis omakorda koostas elupaigatüüpide klassifitseeritud kaardi. Mudeli hindamiseks viidi läbi nii visuaalne hinnang kui ka ristvalideerimismatriksi analüüs.

Mudeli täpsuseks saavutati treeningul 97% ja valideerimisel 82%, mis viitas mudeli kõrgele täpsusele. Kõige olulisemaks omaduseks ei osutunud mõni vegetatsiooniindeks, vaid hoopis *thermal* kanal. Kuigi mudeli täpsus oli hea, tulid visuaalsel ja ristvalideerimismatriksil hindamisel välja mõningad tegurid, millele tuleb tähelepanu pöörata. Ligi 30% kogu uurimisalast prognoosis masin mitte-metsaks. See on suuresti mõjutatud *thermal* kanali poolt, kus tuli kontrolli käigus välja satelliitandmete puudulikkus.

Võrreldes inventeerija välitöö tulemustega suutis mudel hinnata täpset samasust väga vähesel määral. Vähese tuvastamise põhjuseks võib olla elupaikade liiga detailne eristamine, mistõttu on masinal keerulisem vahet teha. Uurimist vajab 2025. aasta nn 0-elupaikade prognoosimine 2011.a. inventeeritud elupaikadele, kuna need on ka praegu sihtkaitsevööndites alles. Kahe enimlevinud metsaelupaigatüübi 9010 ja 9050 puhul saavutati neid koos vaadeldes rahuldav valideerimistäpsus (94%), muudeks üksusteks pakuti neid seega üsna vähesel määral. Lisaks viitas prognoos metsade esinduslikkuse võimalikule paranemisele eelmisest inventuurist möödunud 14 aasta jooksul.

Läbi viidud uuring osutab, et metsaelupaikade inventeerimisandmete põhjal on Eestis võimalik satelliitandmete ja masinõppe abil koostada klassifitseeritud elupaigatüüpide kaarti. Uuringu edasiarendamiseks on mitmeid võimalusi. Olemas on suur hulk 2022-2024 aasta Haanja

looduspargi metsaelupaigatüüpide inventeerimise andmeid, mille rakendamisel saaks tõsta prognoosimise täpsust. Tähelepanu saaks rohkem pöörata sesoonsele tegurile, otsides pilvevabu perioode ning uurida aastatevahelist varieeruvust.

Täpsem prognooskaart võimaldaks tulevikus Haanja looduspargis rohundirikaste kuusikute ja ka teiste metsaelupaigatüüpide levikut ning esinduslikkuse muutust seirata kaugseire vahenditega.

Summary

Predicting the distribution and structural quality of Fennoscandian herb-rich forests with *Picea abies* using remote sensing and machine learning in the Haanja Nature Park

Author: Annabel Metsalu

This study investigates the potential of forest remote sensing to map and analyse herb-rich spruce forest habitats (habitat type 9050) within the Natura 2000 network in Haanja Nature Park. The aim of the study was to assess whether the spatial distribution and quality of forest habitats can be analysed using a combination of field data, a machine learning model, and satellite data.

The study utilised Landsat 8 and Landsat 9 image composites, along with inventory data collected during field surveys in 2011 and 2025. In addition, a forest height map, a forest mask, and soil data were included in the analysis. The Random Forest algorithm was used as the machine learning model. The model achieved an overall training accuracy of 97%, an overall validation accuracy of 82%, and a Kappa value of 0.80. The most influential feature was the thermal band, followed by NDVI, elevation data, and the blue band.

As a result of the modelling, a classified habitat map was produced and evaluated both visually and using a confusion-matrix-based validation approach. Both visual and statistical assessments indicated that a relatively large proportion of the area (30.4%) was classified as non-habitat.

Based on the confusion matrix, the model showed limited ability to predict habitat types consistently with field-based inventory results, particularly for coniferous forests. One possible reason for this may be the overly detailed classification scheme of habitat types, which made it difficult for the model to distinguish between similar classes. For example, the model predicted habitat type 9050 with representativity class A on 41.9% of the area that was classified as class B in the inventory data. This may be due to minimal structural differences between stands from the model's perspective. However, such discrepancies may also reflect actual ecological changes, as forest condition may have improved over time (with 2011 data used as reference), which is expected in strictly protected areas where no forest management occurs. A similar pattern was observed, with class B habitats predicted in areas mapped as class C.

The analysis demonstrated that combining remote sensing data, field survey data, and machine learning methods enables the modelling of forest habitat distribution and condition. Although the model achieved relatively high accuracy, comparison with independent data revealed several factors that need careful consideration. One of the main limitations was incomplete data, particularly gaps in the thermal band, which likely contributed to the large proportion of non-habitat predictions.

A positive outcome of the study was that, despite these limitations, satisfactory validation accuracy (94%) was achieved for the two most common forest habitat types.

In future studies, the amount of data could be increased by incorporating inventory data from 2022–2024. Additionally, the effect of seasonal variation on model accuracy could be investigated. For example, distinguishing between deciduous, mixed, and coniferous forests may be improved using early spring satellite imagery.

This study confirms that the integration of remote sensing, field data, and machine learning is a promising approach with strong potential as a tool for monitoring the distribution and changes in representativity of herb-rich spruce forests, as well as other forest habitat types, in Haanja Natural Park.

Tänuavaldused

Soovin tänada oma juhendajaid, Anneli Palo ja Valentina Sagrist, kes olid väga abivalmid ja aitasid mind kogu töö koostamise protsessi vältel tagasiside, näpunäidete ja soovitudustega. Lisaks tänan Anneli Palo Haanja metsade inventuuri andmete kasutamise võimaldamise ja looduskaitsealaste soovitude eest. Valentina Sagrist tänan kaugseirealaste juhiste ja nõuannete eest. Lõpetuseks soovin tänada ka Alexander Kmochi, kes jagas viiteid asjakohaste mullastikuandmete jaoks.

Kasutatud kirjandus

Al Saim, A., Aly, M. H. (2025). Machine learning and multisensor data fusion for forest aboveground biomass estimation in Arkansas. *Frontiers in Forests and Global Change*. <https://doi.org/10.1007/s41748-025-00713-z>

Boyd, D. S., Danson, F. M. (2005). Satellite remote sensing of forest resources: Three decades of research development. *Progress in Physical Geography: Earth and Environment*, 29(1), 1–26. <https://doi.org/10.1191/0309133305pp432ra>

Caffaratti, G. D., Marchetta, M. G., Euillades, L. D., Euillades, P. A., Forradellas, R. Q. (2021). Improving forest detection with machine learning in remote sensing data. *Remote Sensing Applications: Society and Environment*, 24, 100654. <https://doi.org/10.1016/j.rsase.2021.100654>

Čahojová, L., Ambroz, M., Jarolínek, I., Kollár, M., Mikula, K., Šibík, J., Šibíková, M. (2022). Exploring Natura 2000 habitats by satellite image segmentation combined with phytosociological data: A case study from the Čierny Balog area (Central Slovakia). *Scientific Reports*, 12(1), 18375. <https://doi.org/10.1038/s41598-022-23066-3>

Coops, N. C., Tompalski, P., Goodbody, T. R. H., Achim, A., Mulverhill, C. (2023). Framework for near real-time forest inventory using multi source remote sensing data. *Forestry: An International Journal of Forest Research*, 96(1), 1–19. <https://doi.org/10.1093/forestry/cpac015>

Earth Engine Data Catalog. (i.a). Datasets tagged landsat-composite in Earth Engine, <https://developers.google.com/earth-engine/datasets/tags/landsat-composite> (viimati vaadatud 07.05.2026).

EELIS (Eesti looduse infosüsteem), Keskkonnaagentuur (2024). Ala andmete kuvamine. Kaitsealune ala või üksikobjekt: Haanja looduspark (KLO1000469), <https://infoleht.keskkonnainfo.ee/ala/1058> (viimati vaadatud 06.05.2026).

Eerme, K., Aun, M. (2012). A review of the variations of optical remote sensing conditions over Estonia in 1958–2011. *International Journal of Remote Sensing Applications*, 2(3), 12–19. <https://archive.org/details/IJRSA10085>

Eesti Keskkonnaühenduste Koda (EKO). (2021). Avalik pöördumine: Soovime heaperemehelikku loodusväärtuste kaitset, <https://eko.org.ee/uudised/avalik-poordumine-soovime-heaperemehelikku-loodusvaartuste-kaitset> (viimati vaadatud 16.05.2026).

Eesti Keskkonnaühenduste Koda (EKO). (2022). EKO: Loodusväärtuste hävitamine kaitsealadel tuleb kohe lõpetada, <https://eko.org.ee/uudised/eko-loodusvaartuste-havitamine-kaitsealadel-tuleb-koheselt-lopetada> (viimati vaadatud 16.05.2026).

Eldeeb, H. A. H. S. (2024). Empowering Machine Learning Pipelines with Automated Feature Engineering. Doctoral thesis. University of Tartu. Computer Science.

European Environment Agency. (i.a). Sites. Haanja, <https://eunis.eea.europa.eu/sites/EE0080613> (viimati vaadatud 06.05.2026).

European Union. (2025). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Article 17, <https://eur-lex.europa.eu/eli/dir/1992/43/oj> (viimati vaadatud 06.05.2026).

Fassnacht, F. E., White, J. C., Wulder, M. A., Næsset, E. (2024). Remote sensing in forestry: Current challenges, considerations and directions. *Forestry: An International Journal of Forest Research*, 97(1), 11–37. <https://doi.org/10.1093/forestry/cpad024>

Feilhauer, H., Dahlke, C., Doktor, D., Lausch, A., Schmidtlein, S., Schulz, G., Stenzel, S. (2014). Mapping the local variability of Natura 2000 habitats with remote sensing. *Applied Vegetation Science*, 17(4), 765–779. <https://doi.org/10.1111/avsc.12115>

GeoAI. (2023). Enhanced Vegetation Index (EVI): A Modern Approach to Vegetation Health Monitoring, <https://www.geoai.au/enhanced-vegetation-index-evi-a-modern-approach-to-vegetation-health-monitoring/> (viimati vaadatud 12.05.26).

Google Earth Engine. (i.a). Platform, <https://earthengine.google.com/platform/> (viimati vaadatud 07.05.2026).

Haanja looduspargi kaitse-eeskiri. Vastu võetud 22.01.2015. Viimane redaktsioon RT I, 23.01.2015, 6.

Haest, B., Vanden Borre, J., Spanhove, T., Thoonen, G., Delalieux, S., Kooistra, L., Mücher, C., Paelinckx, D., Scheunders, P., Kempeneers, P. (2017). Habitat Mapping and Quality Assessment of NATURA 2000 Heathland Using Airborne Imaging Spectroscopy. *Remote Sensing*, 9(3), 266. <https://doi.org/10.3390/rs9030266>

Harju, Ü. (2021). RMK rikkus iseenda korda, et mets enne inventuuri maha raiuda. *Postimees*, 27. September. Kasutatud 07.05.2026, <https://www.postimees.ee/7347100/rmk-rikkus-iseenda-korda-et-mets-enne-inventuuri-maha-raiuda>

Hernández-Clemente, R., Hornero, A., Mottus, M., Penuelas, J., González-Dugo, V., Jiménez, J. C., Suárez, L., Alonso, L., Zarco-Tejada, P. J. (2019). Early Diagnosis of Vegetation Health From High-Resolution Hyperspectral and Thermal Imagery: Lessons Learned From Empirical Relationships and Radiative Transfer Modelling. *Current Forestry Reports*, 5(3), 169–183. <https://doi.org/10.1007/s40725-019-00096-1>

Ismayilova, I. (2020). Suitability Analysis for Alvars in Estonia using Random Forest and GIS based Multi Criteria Decision Making approach. Master's thesis. y. University of Tartu. Geoinformatics for Urbanised Societ.

Janitza, S., Hornung, R. (2018). On the overestimation of random forest's out-of-bag error. *PLOS ONE*, 13(8), Article e0201904. <https://doi.org/10.1371/journal.pone.0201904>

Kangas, A., Maltamo, M. (2006). *Forest inventory: Methodology and applications*. Springer. <https://doi.org/10.1007/1-4020-4381-3>

Karasiak, N., Dejoux, J.-F., Monteil, C., Sheeren, D. (2022). Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, 111(7), 2715–2740. <https://doi.org/10.1007/s10994-021-05972-1>

Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., Adams, P. (2009). Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote Sensing of Environment*, 113(7), 1382–1396. <https://doi.org/10.1016/j.rse.2008.07.018>

Keskkonnaagentuur. (2025). Eesti metsade kõrguskaart. Metsade kõrguskaardi metoodika, <https://keskkonnaportaal.ee/sites/default/files/2025-06/Metsade%20korguskaardi%20metoodika.pdf> (viimati vaadatud 06.05.2026).

Keskkonnaamet. (2022). Keskkonnaamet peatab raied Natura alal asuvates metsaelupaikades, <https://keskkonnaamet.ee/uudised/keskkonnaamet-peatab-raied-natura-alal-asuvates-metsaelupaikades> (viimati vaadatud 06.05.2026).

Keskkonnaamet (2025). Natura 2000 võrgustiku alade metsaelupaigatüüpide ja -liikide administratiivsete kaitsemeetmete rakendamise kava, <https://www.keskkonnaamet.ee/sites/default/files/documents/2025-09/Metsaelupaigat%C3%BC%C3%BCpide%20ja%20metsaliikide%20kaitsep%C3%B5him%C3%B5tete%20rakendamise%20kava%20Natura%202000%20v%C3%B5rgustiku%20alad%2027.08.2025.pdf> (viimati vaadatud 06.05.2026).

Keskkonnaportaal. (2024). Andmeallikad. Andmekogud. EELIS, <https://keskkonnaportaal.ee/et/eelis> (viimati vaadatud 06.05.2026).

Keskkonnaportaal. (2026a). Andmeallikad. Andmekogud. EELIS. Ruumiandmete teenused ja EELISE kasutamine, <https://keskkonnaportaal.ee/et/ruumiandmete-teenused-ja-eelise-kasutamine> (viimati vaadatud 06.05.2026).

Keskkonnaportaal. (2026b). Metsamaskimetoodika loomine, <https://keskkonnaportaal.ee/et/metsamaski-metoodika-loomine> (viimati vaadatud 21.04.2026).

Keskkonnaportaal. (2026c). Publikatsioonid. Väljaanded ja ülevaated. Natura 2000 ja kaitstavad alad. Haanja looduspark, <https://keskkonnaportaal.ee/et/natura-2000-ja-kaitstavad-alad> (viimati vaadatud 06.05.2026).

Kliimaministeerium. (2026). Elurikkus Keskkonnakaitse. Looduskaitse. Kaitsealad ja Natura 2000, <https://kliimaministeerium.ee/elurikkus-keskkonnakaitse/looduskaitse/kaitsealad-ja-natura-2000#maastikukaitseala-l> (viimati vaadatud 06.05.2026).

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Pärtel, M., Ostonen, I., Uuemaa, E. (2020). EstSoil-EH: A high-resolution eco-hydrological modelling parameters dataset for Estonia (dataset) (Version v1.2c) [Dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.3473289>

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Pärtel, M., Ostonen, I., Uuemaa, E. (2021). EstSoil-EH: A high-resolution eco-hydrological modelling parameters dataset for Estonia. *Earth System Science Data*, 13(1), 83–97. <https://doi.org/10.5194/essd-13-83-2021>

Lambert, J., Drenou, C., Denux, J.-P., Balent, G., Cheret, V. (2013). Monitoring forest decline through remote sensing time series analysis. *GIScience & Remote Sensing*, 50(4), 437–457. <https://doi.org/10.1080/15481603.2013.820070>

Lang, M., Kaha, M., Laarmann, D., Sims, A. (2018). Construction of tree species composition map of Estonia using multispectral satellite images, soil map and a random forest algorithm. *Forestry Studies*, 68(1), 5–24. <https://doi.org/10.2478/fsmu-2018-0001>

Lang, M. (2022). Tüüplahendused kaugseire lõimimiseks pideva metsakorralduse süsteemi. *Kaugseire Eestis 2021*. Toim. Urmas Peterson. Tartu Ülikooli Tartu observatoorium, 111–120.

Leivits, M. (2019). Ülevaade loodusdirektiivi metsaelupaikade seisundist (2013-2018) elupaigainventuuride ja seireandmete põhjal, [https://keskkonnaportaal.ee/sites/default/files/2021-12/%C3%BClevaade%20loodusdirektiivi%20metsaelupaikade%20seisundist%20\(2013-2018\)%20elupaigainventuuride%20ja%20seireandmete%20p%C3%B5hjal.pdf](https://keskkonnaportaal.ee/sites/default/files/2021-12/%C3%BClevaade%20loodusdirektiivi%20metsaelupaikade%20seisundist%20(2013-2018)%20elupaigainventuuride%20ja%20seireandmete%20p%C3%B5hjal.pdf) (viimati vaadatud 06.05.2026).

Leivits, M. (2022). Loodusdirektiivi metsaelupaikade kaardistamise võimalused kasutades kaugseire andmeid ja masinõppe meetodeid, <https://keskkonnaportaal.ee/sites/default/files/2022-04/Loodusdirektiivi%20metsaelupaikade%20kaardistamise%20v%C3%B5imalused%20kasutades%20kaugseire%20andmeid%20ja%20masin%C3%B5ppe%20meetodeid.pdf> (viimati vaadatud 06.05.2026)

Liira, J. (2009). Olemasolevate koosluste seiremetoodikate hindamine ning soovitusi Natura 2000 elupaikade seisundi seiremetoodika edendamiseks, <http://www.digar.ee/id/nlib-digar:194631> (viimati vaadatud 06.05.2026).

Linder, M., Jakobson, L., Absalon, E. (2010). The effect of distance correction factor in case-based predictions of vegetation classes in Karula, Estonia. In *ISPRS Archives, Volume XXXVIII, Part 2*. International Society for Photogrammetry and Remote Sensing. https://www.isprs.org/proceedings/XXXVIII/part2/Papers/126_Paper.pdf

Maa- ja Ruumiamet. (2026a). Metsamuutuste kaardirakendus, <https://xgis.maaamet.ee/xgis2/page/app/metsamuutused> (viimati vaadatud 07.05.2026).

Maa- ja Ruumiamet. (2026b). Ortofoto 21.05.2024, <https://kaart.maaamet.ee/wms/alus?>. Litsents: <https://geoportaal.maaruum.ee/avaandmete-litsents>.

Maamets, L., Kuuba, R., Lõhmus, A., Uuemaa, E., Rosenvald, R., Kont, R. (2023). Eesti metsade alternatiivne hindamine. Postimees Kirjastus, <http://www.digar.ee/id/nlib-digar:980341> (viimati vaadatud 16.05.2026).

Marchi, N., Pirotti, F., Lingua, E. (2018). Airborne and terrestrial laser scanning data for the assessment of standing and lying deadwood: Current situation and new perspectives. *Remote Sensing*, 10(9), 1356. <https://doi.org/10.3390/rs10091356>

Marcinkowska-Ochtyra, A., Gryguc, K., Ochtyra, A., Kopec, D., Jarocińska, A., Slawik, L. (2019). Multitemporal Hyperspectral Data Fusion with Topographic Indices—Improving Classification of Natura 2000 Grassland Habitats. *Remote Sensing*, 11(19), 2264. <https://doi.org/10.3390/rs11192264>

McMahon, C. R., Howe, H., van den Hoff, J., Alderman, R., Brolsma, H., Hindell, M. A. (2014). Satellites, the all-seeing eyes in the sky: Counting elephant seals from space. *PLOS ONE*, 9(3), e92613. <https://doi.org/10.1371/journal.pone.0092613>

NASA Earthdata. (i.a). Learn. Earth Observation Data Basics, <https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/remote-sensing> (viimati vaadatud 06.05.2026).

Nguyen, A., Saha, S. (2024). Machine learning and multi-source remote sensing in forest carbon stock estimation: A review. *arXiv preprint arXiv:2411.17624*. <https://doi.org/10.48550/arXiv.2411.17624>

Oviir, M., Remm, K., Linder, M. (2008). Eesti põhikaardi okas-, sega- ja lehtmetsa eristatus kaugseireandmete ja mullakaardi järgi, kasutades näidistele tuginevat järeldamist ja tehisoõpet. *Kaugseire Eestis 2008. Toim. Katrin Väljataga, Karmen Kaukver. Tartu Ülikooli Tartu observatoorium*, 69-77.

Paal, J. (2000). Loodusdirektiivi elupaigatüüpide käsiraamat, <http://www.digar.ee/id/nlib-digar:10577> (viimati vaadatud 06.05.2026).

Palo, A. (2011). Haanja looduspargi Kütioru, Vällamäe, Tuhkrijärve, Vaskna, Suur-Munamäe ja Jürihani sihtkaitsevööndite loodusdirektiivi metsaelupaigatüüpide inventeerimine. *Tartu*.

Palo, A., Gimbutas, M. (2017). Successional dynamics of tree composition in mixed boreo-nemoral stands from Habitat Directive forest types *9020, 9050, 9060, *9010. *Baltic Forestry*, 23(3), 546–555.

Palo, A. (2018). Loodusdirektiivi metsaelupaikade inventeerimise juhend. *Toim.-d Viilma, K., Paal, J., Türrpu, T., Maamets, L., Otsus, M., Animägi, A., Sell, I. Käsikiri. Tellija Eesti Keskkonnaministeerium*.

Palo, A. (2022). Haanja elupaikade inventuur (metsad). Käsikiri. Tellija: Keskkonnaamet. Täitja: OÜ Metsamutt.

Palo, A., Sell, I. (2024). Haanja elupaikade inventuur (metsad). Käsikiri. Tellija: Keskkonnaamet. Täitja: OÜ Metsamutt, MTÜ Puuseen.

Palo, A. (2025). Haanja elupaikade inventuur. Käsikiri. Tellija: Keskkonnaamet. Täitja: OÜ Metsamutt.

Peterson, U., Liira, J. (2012). MONITORING OF FOREST AREA CHANGES USING SNOW COVERED SATELLITE IMAGERY IN NORTHERN EUROPE. Proceedings of International Scientific Conference Forests in the Future – Sustainable Use, Risks and Challenges: International Scientific Conference Forests in the Future – Sustainable Use, Risks and Challenges, 4-5 October 2012, Belgrade, Republic of Serbia. Ed. Ljubinko Rakonjac. Belgrade, Republic of Serbia: Institute of Forestry, Belgrade, Serbia, 123–129, <https://www.forest.org.rs/pdf/proceedings-conference2012.pdf>

Prodromou, M., Theocharidis, C., Gitas, I. Z., Eliades, F., Themistocleous, K., Papasavvas, K., Dimitrakopoulos, C., Danezis, C., Hadjimitsis, D. (2024). Forest Habitat Mapping in Natura2000 Regions in Cyprus Using Sentinel-1, Sentinel-2 and Topographical Features. *Remote Sensing*, 16(8), 1373. <https://doi.org/10.3390/rs16081373>

QGIS. (i.a). <https://qgis.org/> (viimati vaadatud 12.05.2026)

Reddy, C. S. (2021). Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodiversity and Conservation*, 30, 2617–2631. <https://doi.org/10.1007/s10531-021-02216-5>

Sankey, T., Donager, J., McVay, J., Sankey, J. B. (2017). UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sensing of Environment*, 195, 30–43. <https://doi.org/10.1016/j.rse.2017.04.007>

Salman, H. A., Kalakech, A., Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 7. <https://doi.org/10.58496/BJML/2024/007>

Scikit-learn. (i.a). User Guide. Random Forests, <https://scikit-learn.org/stable/modules/ensemble.html#forest> (viimati vaadatud 05.05.2026).

Sittaro, F., Hutengs, C., Semella, S., Vohland, M. (2022). A Machine Learning Framework for the Classification of Natura 2000 Habitat Types at Large Spatial Scales Using MODIS Surface Reflectance Data. *Remote Sensing*, 14(4), 823. <https://doi.org/10.3390/rs14040823>

Soesoo, A., Hade, S. (2005). Haanja looduspark. Küljendus: Eesti Loodusfoto

Sun, G., Schulz, K. (2015). The Landsat satellite record for land cover classification. *Remote Sensing*, 7(7), 8368–8386. <https://doi.org/10.3390/rs70708368>

Zerrouk, M., Ait El Kadi, K., Sebari, I., Fellahi, S. (2025). Machine and deep learning for wetland mapping and bird-habitat monitoring: A systematic review of remote-sensing applications (2015–April 2025). *Remote Sensing*, 17(21), 3605. <https://doi.org/10.3390/rs17213605>

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18(6), 306–314. [https://doi.org/10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)

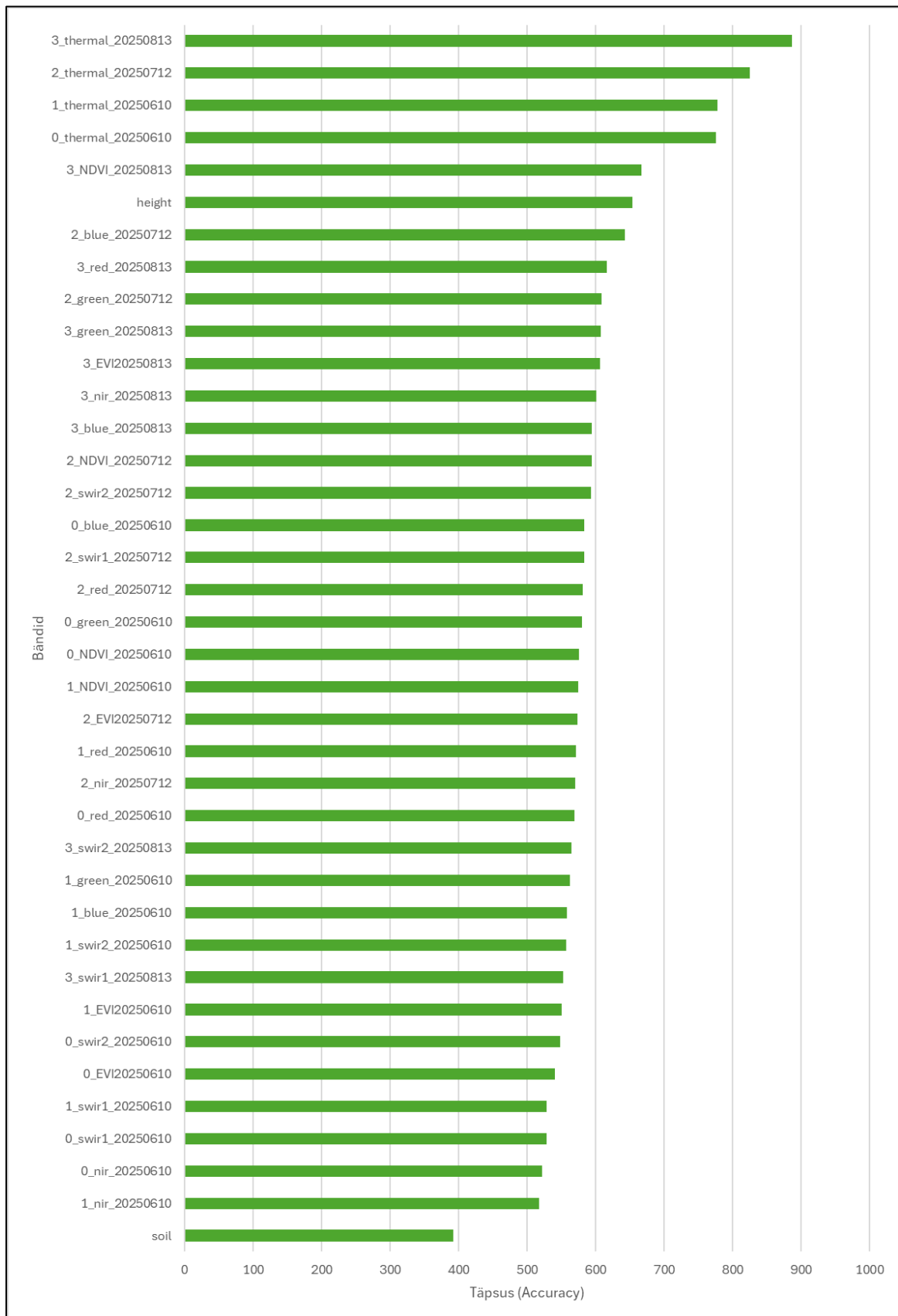
U.S. Geological Survey, Landsat 2 (i.a). Landsat 2. U.S. Department of the Interior.

<https://www.usgs.gov/landsat-missions/landsat-2> (viimati vaadatud 06.05.2025).

U.S. Geological Survey. (i.a). Landsat Satellite Missions, <https://www.usgs.gov/landsat-missions/landsat-satellite-missions> (viimati vaadatud 06.05.2026).

Lisad

Lisa 1. Tunnuste olulisuse (ingl. k *Feature Importance*) graafik.



Lisa 2. Ristvalideerimismatriks, mis on koostatud prognooskaardi ja 2011. aasta inventeeritud andmete kattumise põhjal.

a) Rasterpunktide kattuvus.

	2025. aasta põhjal prognoositud elupaigatüübid																						
2011	0	6000	7000	8000	9010	9020	9060	9080	9180	9050_A_O	9050_A_S	9050_B_O	9050_B_S	9050_C_L	9050_C_O	9050_C_S	9050_p_L	9050_p_O	9050_p_S	91D0	91F0	(blank)	Grand Total
0	658	19			458	40	31	33	6	1		6		34	36	69	61	37	22	47		1433	2991
9010	782	3			1365	84	123	5		30		43		251	10	7	123	15		56		1088	3985
9020	69	1			15	132								83		39						1	340
9060	25				36	2								5		3				4		45	120
9080	13				25			9												4			51
9180	43				22	32								15		5						186	303
9050_A_O	344				217	9	2	1		9		3		24	11	4	5			3		398	1030
9050_A_S	109				45	13								11			1					1	180
9050_B_L	17				3									16									36
9050_B_O	370	3			731	49	19	2		31		42		87	8	28	5	4		27		161	1567
9050_B_S	52				127	5	14	1		2		3		9		4	1	2		1			221
9050_C_L	9														2								11
9050_C_O	180	20			560	14	17			1		10		71	6	40	13			4		140	1076
9050_C_S	16				2	2								9								9	38
9050_p_L	1				1									50									52
9050_p_O	15				42									3		102	2			9		146	319
91D0	11				53			2												1		116	183
91F0																						116	116
(blank)																							
Grand Total	2714	46			3702	382	206	53	6	74		107		668	73	301	211	58	22	156		3840	12619

b) Rasterpunktide kattuvuse osakaal.

	2025. aasta põhjal prognoositud elupaigatüübid																						
2011	0	6000	7000	8000	9010	9020	9060	9080	9180	9050_A_O	9050_A_S	9050_B_O	9050_B_S	9050_C_L	9050_C_O	9050_C_S	9050_p_L	9050_p_O	9050_p_S	91D0	91F0	(blank)	Grand Total
0	24,2	41,3	0,0	0,0	12,4	10,5	15,0	62,3	100,0	1,4	0,0	8,1	0,0	5,1	49,3	22,9	28,9	63,8	100,0	30,1	0,0	37,3	2991
9010	28,8	6,5	0,0	0,0	36,9	22,0	59,7	9,4	0,0	40,5	0,0	58,1	0,0	37,6	13,7	2,3	58,3	25,9	0,0	35,9	0,0	28,3	3985
9020	2,5	2,2	0,0	0,0	0,4	34,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	12,4	0,0	13,0	0,0	0,0	0,0	0,0	0,0	0,0	340
9060	0,9	0,0	0,0	0,0	1,0	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,7	0,0	1,0	0,0	0,0	0,0	2,6	0,0	1,2	120
9080	0,5	0,0	0,0	0,0	0,7	0,0	0,0	17,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,6	0,0	0,0	51
9180	1,6	0,0	0,0	0,0	0,6	8,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,2	0,0	1,7	0,0	0,0	0,0	0,0	0,0	4,8	303
9050_A_O	12,7	0,0	0,0	0,0	5,9	2,4	1,0	1,9	0,0	12,2	0,0	4,1	0,0	3,6	15,1	1,3	2,4	0,0	0,0	1,9	0,0	10,4	1030
9050_A_S	4,0	0,0	0,0	0,0	1,2	3,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,6	0,0	0,0	0,5	0,0	0,0	0,0	0,0	0,0	180
9050_B_L	0,6	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	36
9050_B_O	13,6	6,5	0,0	0,0	19,7	12,8	9,2	3,8	0,0	41,9	0,0	56,8	0,0	13,0	11,0	9,3	2,4	6,9	0,0	17,3	0,0	4,2	1567
9050_B_S	1,9	0,0	0,0	0,0	3,4	1,3	6,8	1,9	0,0	2,7	0,0	4,1	0,0	1,3	0,0	1,3	0,5	3,4	0,0	0,6	0,0	0,0	221
9050_C_L	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	11
9050_C_O	6,6	43,5	0,0	0,0	15,1	3,7	8,3	0,0	0,0	1,4	0,0	13,5	0,0	10,6	8,2	13,3	6,2	0,0	0,0	2,6	0,0	3,6	1076
9050_C_S	0,6	0,0	0,0	0,0	0,1	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	38
9050_p_L	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	7,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	52
9050_p_O	0,6	0,0	0,0	0,0	1,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	33,9	0,9	0,0	0,0	5,8	0,0	3,8	319
91D0	0,4	0,0	0,0	0,0	1,4	0,0	0,0	3,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	3,0	183
91F0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,0	116
(blank)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	
Grand Total	2714	46			3702	382	206	53	6	74		107		668	73	301	211	58	22	156		3840	12619
% 2025 valimist (veeru summast)	21,5	0,4	0,0	0,0	29,3	3,0	1,6	0,4	0,0	0,6	0,0	0,8	0,0	5,3	0,6	2,4	1,7	0,5	0,2	1,2	0,0	30,4	

Lisa 3. Google Earth Engine kood

```

/*****
* 1) KASUTAJA SISEND
*****/
var roi = ee.Geometry.Rectangle([26.86, 57.62, 27.21, 57.82]);

var startDate = '2025-06-01'; // vali ulatus, mida soovid
var endDate = '2025-09-01';

// Treeningandmed: punktshapefile (Assets)
var trainingPoints = ee.FeatureCollection('projects/ee-annabelmetsalu/assets/GEE_4_LABELS_training_points');
var classProperty = 'Label4_id'; // väli shapefile'is, mis sisaldab klassi koodi (int)
print (trainingPoints.size())
print('First N features', trainingPoints.limit(10));

//var forestMask = ee.Image("projects/ee-annabelmetsalu/assets/metsamask");
var forestMask = metsamask.clip(roi);
var m_mask = forestMask.eq(1); // forest is 1
var forestcolor = { min: 0, max: 1, palette: ['#cccccc', '#228b22'] };

var taim_korgus = taimkatekorgus.clip(roi).select(['b1'], ['height']); //
take firsth band and rename to 'height'
// var soilBandNamesToUse = null; // e.g., ['Band1'] or null to use all
bands
// var NoData = -9999; // <-- EDIT if your GeoTIFF uses a different NoData
value

var muld = mullad.clip(roi).select(['b1'], ['soil']); // take firsth band
and rename to 'height'
// print(muld.bandNames())

//32 days composite - selected days list
var dateList = ee.List(['2025-03-06', '2025-06-10', '2025-07-12', '2025-
08-13' ]);
// Convert to ee.Date objects
dateList = dateList.map(function(d) { return ee.Date(d); });
print(dateList)

/*****
* 2) LANDSAT N-PÄEVASED NDVI KOMPOSIITID * (kasutame kõiki pilte
ajavahehikud)
*****/
// Andmestik: LANDSAT/COMPOSITES/C02/T1_L2_8DAY_NDVI (NDVI -1..1, 8-päevane
komposiit) [docs]
// var ndviCol =
ee.ImageCollection('LANDSAT/COMPOSITES/C02/T1_L2_8DAY_NDVI')
var ndviCol = ee.ImageCollection('LANDSAT/COMPOSITES/C02/T1_L2_32DAY_NDVI')
.filterDate(startDate, endDate)

```

```

        .filterBounds(roi)
        .select('NDVI'); // võta ainult NDVI bänd
// (Andmestik kirjeldatud Earth Engine'i kataloogis; komposiit kasutab
// perioodi kõige hilisemat piksli väärtust.)
[1] (https://developers.google.com/earth-engine/datasets/catalog/LANDSAT\_COMPOSITES\_C02\_T1\_L2\_8DAY\_NDVI)

var ndviCol_2 = ee.ImageCollection(
  dateList.map(function(d) {
    d = ee.Date(d); // convert to ee.Date inside map()
    var img = ndviCol
      .filter(ee.Filter.gte('system:time_start', d.millis()))
      .first();
    // Fallback for missing dates
    img = ee.Image(img);
    return img .clip(roi)
      //.set('requested_date', d);
  })
);

// Anna igale pildile unikaalne kanali nimi kuupäeva järgi (1 pilt = 1
// kanal) var ndviNamed =
ndviCol_2.sort('system:time_start').map(function(img) {
  var dateStr =
ee.Date(img.get('system:time_start')).format('YYYYMMdd');
  return img.rename(ee.String('NDVI_').cat(dateStr));
});

// Muuda kollektsioon mitmebändiliseks pildiks (ajasari)
var ndviSeries =
ee.ImageCollection(ndviNamed).toBands().clip(roi).updateMask(m_mask);

// Kontroll: mitu bändi (peaks võrduma piltide arvuga)
print('NDVI ajasari bändid:', ndviSeries.bandNames());
print('Piltide arv:', ndviCol.size());

// EVI, NDVI ausel

var eviCol = ee.ImageCollection('LANDSAT/COMPOSITES/C02/T1_L2_32DAY_EVI')
  .filterDate(startDate, endDate)
  .filterBounds(roi)
  .select('EVI');

var eviCol_2 = ee.ImageCollection(
  dateList.map(function(d) {
    d = ee.Date(d); // convert to ee.Date inside map()
    var img = eviCol
      .filter(ee.Filter.gte('system:time_start', d.millis()))
      .first();
    // Fallback for missing dates
    img = ee.Image(img);
    return img
      .clip(roi)

```

```

        //set('requested_date', d);
    })
};

var eviNamed = eviCol_2.sort('system:time_start').map(function(img) {
    var dateStr =
        ee.Date(img.get('system:time_start')).format('YYYYMMdd'); return
        img.rename(ee.String('EVI').cat(dateStr));
});

var eviSeries =
ee.ImageCollection(eviNamed).toBands().clip(roi).updateMask(m_mask);

print('EVI ajsari bändid:', eviSeries.bandNames());
print('EVI piltide arv:', eviCol.size());

var composites = ee.ImageCollection('LANDSAT/COMPOSITES/C02/T1_L2_32DAY')
    .filterDate(startDate, endDate)
    .filterBounds(roi)
    .select(['blue', 'thermal']);

var composites_2 = ee.ImageCollection(
    dateList.map(function(d) {
        d = ee.Date(d);
        var img = composites
            .filter(ee.Filter.gte('system:time_start', d.millis()))
            .first();
        return ee.Image(img).clip(roi);
    })
);

var compositesNamed = composites_2.map(function(img) {
    var dateStr =
ee.Date(img.get('system:time_start')).format('YYYYMMdd');
    var bandNames = img.bandNames().map(function(b) {
        return ee.String(b).cat('_').cat(dateStr);
    });
    return img.rename(bandNames);
});

var compositesSeries = ee.ImageCollection(compositesNamed)
    .toBands()
    .clip(roi)
    .updateMask(m_mask);

//predictor stack
var predictors =
ndviSeries.addBands(eviSeries).addBands(compositesSeries).addBands(taim_kor
gus); //addBands(muld);
var predictorNames = predictors.bandNames();

```

```

print('Predictor bands (NDVI + EVI + COMPOSITES + Soil + height):',
predictorNames.size(), predictorNames);

/*****
* 3) TRENN + RANDOM FOREST
*****/

var inputBands = predictors.bandNames();

// Proovid treeningpunktides kõigilt ajasari-bändidelt
var training = predictors.sampleRegions({
  collection: trainingPoints,
  properties: [classProperty],
  scale: 30,
  geometries: false
});

/***** TRAIN/VALIDATION SPLIT *****/

var withRandom = training.randomColumn({columnName: 'rand', seed: 42});
var split = 0.7;
var trainSet = withRandom.filter(ee.Filter.lte('rand', split));
var validSet = withRandom.filter(ee.Filter.gt('rand', split));

print('Train size:', trainSet.size(), 'Valid size:', validSet.size());

// Random Forest (vajadusel muuda puude arvu jms)
var classifier = ee.Classifier.smileRandomForest({
  numberOfTrees: 200,
  minLeafPopulation: 1,
  seed: 13
}).train({
  features: training,
  classProperty: classProperty,
  inputProperties: inputBands
});

//train random forest
var nTrees = 300;
var variablesPerSplit = 1; // 0 -> default (sqrt of features)
var minLeafPopulation = 2;
var bagFraction = 0.632;

var rf = ee.Classifier.smileRandomForest({
  numberOfTrees: 300,
  variablesPerSplit: 2, // default: sqrt of features
  minLeafPopulation: 2,
  bagFraction: 0.632,
  seed: 42
}).train({
  features: trainSet,
  classProperty: classProperty,

```

```

        inputProperties: inputBands
    });

    // Feature importance (by Gini importance)
    var explanation = rf.explain();
    var importances = ee.Dictionary(explanation.get('importance'));
    print('Feature importance:', importances);

    /***** EVALUATION *****/

    var trainClassified = trainSet.classify(rf);
    var trainConfMat = trainClassified.errorMatrix(classProperty,
        'classification');
    print('Training overall accuracy:', trainConfMat.accuracy());

    var validClassified = validSet.classify(rf);
    var validConfMat = validClassified.errorMatrix(classProperty,
        'classification');
    print('Validation confusion matrix: ', validConfMat);
    print('Validation overall accuracy: ', validConfMat.accuracy(), 'Kappa: ',
        validConfMat.kappa(), 'Producers Accuracy: ',
        validConfMat.producersAccuracy(), 'Consumers Accuracy:',
        validConfMat.consumersAccuracy());

    var classValues = ee.List.sequence(1, 16);

    var errorMatrix = validClassified.errorMatrix(
        classProperty, // tegelik klass
        'classification', // ennustatud klass
        classValues // fikseeritud klassid
    );

    print('Error matrix:', errorMatrix);

    // Klassifitseeri kogu ajasari-pilt
    var classified = predictors.classify(rf);

    // Rakenda ristkülikmask (mask = ROI)
    var classifiedMasked = classified.updateMask(ee.Image(1).clip(roi));

    /*****
    * 4) PALETID JA VISUALISEERIMINE
    *****/

    var classVis = {
        min: 1,
        max: 46,
        palette: [
            // 1-9 → MUU (niit, raba, allikas jne) - hall
            '#bdbdbd', '#bdbdbd', '#bdbdbd', '#bdbdbd', '#bdbdbd',
            '#bdbdbd', '#bdbdbd', '#bdbdbd', '#bdbdbd',

```

```

// 10-18 → 9010 + 9020 - roheline
'#2ca25f', '#2ca25f', '#2ca25f', '#2ca25f', '#2ca25f',
'#2ca25f', '#2ca25f', '#2ca25f', '#2ca25f',

// 19-29 → 9050 - erkroheline (kõige olulisem)
'#7FFF00', '#7FFF00', '#7FFF00', '#7FFF00', '#7FFF00',
'#7FFF00', '#7FFF00', '#7FFF00', '#7FFF00', '#7FFF00',
'#7FFF00',

// 30-32 → muud metsad - pruun
'#8c510a', '#8c510a', '#8c510a',

// 33-37 → soo-lehtmetsad - sinine
'#1f78b4', '#1f78b4', '#1f78b4', '#1f78b4', '#1f78b4',

// 38-39 → muud metsad - pruun
'#8c510a', '#8c510a',

// 40-46 → siirdesoo-, rabametsad, lammimetsad, uhtvalli metsad
- sinine
'#1f78b4', '#1f78b4', '#1f78b4', '#1f78b4', '#1f78b4',
'#1f78b4', '#1f78b4'
]
};

// Kuvame kihid
Map.centerObject(roi, 10);
Map.addLayer(roi, {}, 'region of interest', false);
Map.addLayer(m_mask, forestcolor, 'metsa mask', false);
Map.addLayer(taim_korgus, {min: 0.0, max: 50.0, palette: ['00FFFF',
'0000FF'] }, 'taimkate kõrgus', false);

Map.addLayer(classifiedMasked, classVis, 'classified', false);
// Map.addLayer(classifiedMasked, { min:1, max: 5, palette: paletteColors}
, 'classified');
Map.addLayer(trainingPoints, {color: 'black'}, 'points', false)

Map.addLayer(roi, {}, 'region of interest');

Map.addLayer(classifiedMasked, {min: 0, max: 50, palette: ['red', 'green',
'blue']});
// (Lisavõimalus: näita viimatist NDVI komposiiti kontekstiks)
var lastNDVI = ee.Image(ndviCol.sort('system:time_start', false).first());
Map.addLayer(lastNDVI.clip(roi), {min: -0.1, max: 0.8, palette:
  ['white', '#d9f0a3', '#78c679', '#238443', '#004529']}, 'NDVI (viimane 8-
p komposiit)', false);
Map.addLayer(trainingPoints, {color: 'black'}, 'points')
Map.addLayer(m_mask, forestcolor, 'Mets ja muu ala');

var lastEVI = ee.Image(eviCol.sort('system:time_start', false).first());

Map.addLayer(

```

```

        lastEVI.clip(roi), {min: -0.1, max: 0.8, palette:
['white', '#fee5d9', '#fcae91', '#fb6a4a', '#cb181d', '#67000d']},
        'EVI',
        false
    );

    /*****
    * 5) EKSPORDID
    *****/

    var exportDescriptionPrefix = 'RF_results_'
    print(exportDescriptionPrefix)

    // 1) Ekspordi toores klassifitseeritud pilt (algsete klassikoodidega,
    maskitud metsa maskiga)
    Export.image.toDrive({
        image: classifiedMasked,
        description: exportDescriptionPrefix + '_map',
        folder: 'GEE_Exports',
        fileNamePrefix: exportDescriptionPrefix + '_map',
        region: roi,
        scale: 30,
        maxPixels: 1e13
    });

    // 2) Export validation Accuracy Matrix
    var val_matrix = ee.Feature(null, {
        matrix: validConfMat.array()
    });
    Export.table.toDrive({
        collection: ee.FeatureCollection([val_matrix]),
        description: exportDescriptionPrefix + '_val_matrix',
        folder: 'GEE_Exports',
        fileFormat: 'CSV'
    });

    // 3) Export evaluation metrics (accuracy) as a small table
    var metrics = ee.Feature(null, {
        train_accuracy: trainConfMat.accuracy(),
        valid_accuracy: validConfMat.accuracy(),
        matrix: validConfMat.array()
    });
    Export.table.toDrive({
        collection: ee.FeatureCollection([metrics]),
        description: exportDescriptionPrefix + '_metrics',
        folder: 'GEE_Exports',
        fileFormat: 'CSV'
    });

```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Annabel Metsalu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Kaugseire ja masinõppe abil rohundirikaste kuusikute elupaiga leviku ja esinduslikkuse prognoosimine Haanja Looduspargi näitel“, mille juhendajad on Anneli Palo ja Valentina Sagris, 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.

Annabel Metsalu

20.05.2026