TARTU ÜLIKOOL

ÖKOLOOGIA JA MAATEADUSTE INSTITUUT

BOTAANIKA OSAKOND

Marta Miia Pärnpuu

LIIKIDE KOOSESINEMISE LEIDMINE TEHISARU ABIL

Bioloogia ja elustiku kaitse

Bakalaureusetöö (12 EAP)

Juhendaja:

Prof. Meelis Pärtel

Tartu 2025

INFOLEHT

Liikide koosesinemise leidmine tehisaru abil

Tehisaru võimaldab automatiseerida liigiandmete kogumist ja analüüsi, tuues uuendusi ka koosluste andmete kogumisse. Käesoleva töö eesmärk on teada saada, kuidas tehisaru abil tuvastada taimeliikide koosesinemisi taimkatte lähifotodelt. Käsitlen liikide koosesinemise põhjuseid ning nende mustrite kujunemist ajalisel ja ruumilisel skaalal. Annan ülevaate nii traditsioonilistest viisidest koosesinemiste uurimiseks kui ka tehisaru kaasavatest lahendustest, keskendudes uurimuslikus osas spetsiifilisemalt tehisaru võimekusele määrata taimeruutude fotodelt mitut liiki. Andmeanalüüsiks lõigatakse fotosid osadeks, eesmärgiga suurendada tõenäosust, et saadud fotoosal jääb esiplaanile üks liik, mida saab määrata PlantNet rakendusega. Lõpuks annan soovitused tehisaru kasutamiseks nii spetsiifiliselt taimeruutude määramisel kui ka üldised soovitused selle edukamaks kaasamiseks ökoloogias.

Märksõnad: tehisintellekt, masinõpe, andmebaasid, elusloodus seire, ökoloogia

CERCS: B270 Taimeökoloogia

Potential Uses of AI in Species Co-occurence Studies

Artificial intelligence enables the automation of species data collection, bringing with it new advances to the study of species assemblages. This thesis investigates the use of AI to identify co-occurrences of plant species in close-up photos of vegetation. The study looks at the reasons behind species co-occurrence as well as how these patterns vary over time and space. It gives a summary of both conventional and AI-based techniques for researching co-occurrence, with the experimental part concentrating on AI's capacity to recognise multiple species of flora from photos of vegetation quadrats. To increase the chances of one species being visually prominent, the photos are split up into smaller segments. The PlantNet application is then used to identify the species in each segment. Lastly, recommendations are made for using AI in the annotation of vegetation quadrats, as well as general suggestions for its successful integration into ecological research.

Keywords: artificial intelligence, machine learning, databases, wildlife monitoring, ecology

CERCS: B270 Plant ecology

Sisukord

SISSEJUHATUS 4
1. LIIKIDE KOOSESINEMISE TEOORIA 5
1.1 Liikide koosesinemise mõiste ja põhjused 5
1.1.1 Liikide koosesinemine erinevatel ruumiskaaladel 6
1.1.2 Liikide koosesinemine erinevatel ajaskaaladel 7
1.1.3 Liikide koosesinemine kui sisend tumeda elurikkuse uurimisse 8
1.2 Liikide koosesinemise matemaatika 8
1.3 Liikide koosesinemise andmete kogumise meetodid 9
1.3.1 Koosluste kirjeldused 9
1.3.2 Liikide koosesinemiste andmete automatiseeritud kogumine 11
1.4 Tehisaru liikide koosesinemiste leidmisel 12
1.4.1 Ülevaade tehisarust 12
1.4.2 Masinõpe ökoloogiliste andmete koondamisel 13
1.4.3 Taimede pildi järgi määramine tehisaruga 14
2. METOODIKA 16
2.1 Andmete allikad ja tehtud uurimuse ülevaade 16
2.2 Andmeanalüüs 19
3. TULEMUSED 21
4. ARUTELU 30
4.1 Probleemid ja edasiarendused 31
KOKKUVÕTE 34
SUMMARY 35
TÄNUAVALDUSED 36
KASUTATUD KIRJANDUS 37
LISAD 49
Lisa 1. R-i skript fotode lõikamiseks ja PlantNetiga määramiseks 49
Lisa 2. R-i skript liigitabelite vormistuse ühtlustamiseks 54
Lisa 3. R-i skript liiginimede ühtlustamiseks 55
Lisa 4. R-i skript tulemuste analüüsimiseks 57
Lisa 5. Ülevaade levinumate taimeliikide määramistulemustest. 63
Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks 67

SISSEJUHATUS

Elurikkuse mõistmine on ökoloogia ja looduskaitse üks keskseid eesmärke (Azeria et al., 2009). Selle saavutamiseks uuritakse muuhulgas liikide koosesinemisi, mis võimaldavad teha järeldusi biootiliste interaktsioonide ja liikide keskkonnaeelistuste kohta (MacKenzie et al., 2004; Tulloch et al., 2018). Juba 1970. aastatest alates on olnud teadlastel selle teema vastu kõrgenenud huvi, mis on viinud mitmesuguste uurimissuundade ja hüpoteeside tekkeni (MacKenzie et al., 2004; Orozco-Arias et al., 2019). Liikide koosesinemiste raamistikus on käsitletud teemasid, nagu koosluste moodustumine (Diamond, 1975), konkurentsi ja keskkonnatingimuste olulisus (Chesson, 1994, 2000) ning erinevate piirkondlike ja globaalsete protsesside roll elustiku mõjutajana (Ricklefs, 2004). Praktiline eesmärk antud valdkonnas on leida keskkonnakaitselisi lahendusi, sh negatiivse inimmõju uurimiseks ja leevendamiseks (M. Pärtel et al., 2025) ning invasiivsete liikide jälgimiseks (Tulloch et al., 2018).

Liike ja nendevahelisi seoseid on palju, seega on liikide koosesinemise mustrite kirjeldamiseks vaja analüüsida suurt hulka andmeid, mistõttu on oluline leida viise, kuidas vähendada analüüsimisel vajalikku inimtööd. Liikide koosesinemise uurimiseks on pakutud erinevaid lahendusi, näiteks nullmudeleid, milles võrreldakse tegelikke koosesinemismustreid juhuslikult koostatuga (Gotelli, 2000). Ühtlasi kasutatakse ka võrgustikuanalüüse (*network approach*) (Bascompte, 2009) ja JSDM mudeleid (*joint species distribution model*) (Pichler & Hartig, 2021). Käesolevas töös käsitlen aga tehisaru kaasamise võimalikkust liikide koosesinemiste uurimiseks. Tehisaru on efektiivne vahend suures kogustes andmete kogumiseks ja analüüsimiseks, kuid sellele vaatamata on tehisaru kasutavad lahendused ökoloogias alles hoogu kogumas ning tehnikavaldkonna välised rakendused piirduvad enamjaolt meditsiiniteadustega (Han et al., 2023; Thessen, 2016).

Uurimistöö eesmärk on teada saada, kuidas tehisaru abil tuvastada taimeliikide koosesinemisi taimkattest tehtud lähifotodelt. Annan kõigepealt ülevaate liikide koosesinemisest ja selle uurimisest traditsioonilistel viisidel. Seejärel tutvustan võimalusi tehisaru kaasamiseks. Toon välja, kuidas tehisaruga on võimalik analüüsida nii andmemassiive kui ka tekste, helisid ja fotosid. Teoreetilisele osale järgneb andmeanalüüs, kus hindan tehisarul põhineva programmi edukust taimede määramisel Eesti puisniitude rohurindest tehtud fotodelt ning fotode lõikamise mõju määramistulemustele.

1. LIIKIDE KOOSESINEMISE TEOORIA

1.1 Liikide koosesinemise mõiste ja põhjused

Mittejuhuslike mustrite esinemist ökoloogilises koosluses nimetame liikide koosesinemiseks (Münzbergová & Herben, 2004). Sellised mustrid tekivad liikide suhestumisel keskkonnaga või teiste liikidega, mistõttu võime ka öelda, et tegemist on ökosüsteemi esilekerkiva omadusega (*emergent property*) – st moodustuv süsteem on komplekssem kui tema komponentide summa (van den Berg et al., 2022; Veech, 2014). Koosesinemine võib olla positiivne, kui liigid esinevad koos sagedamini, kui juhuslikult oodata võiks, või negatiivne, kui liigid „väldivad" üksteist või teatud keskkonnatingimusi (Tulloch et al., 2018). Eelmainitud mustrid on statistiliselt suhteliselt kergesti tuvastatavad, kuid nende tekkepõhjused on komplekssed, olles mõjutatud mitmetest abiootilistest kui ka biootilistest faktoritest ning vajades põhjalikumat uurimist (Ovaskainen et al., 2010; Tulloch et al., 2018).

Koosluste kujunemisel on oluline roll biootilistel vastasmõjudel (Kraan et al., 2020). Vastasmõjudest tingitud koosesinemiste korral on liigi esinemine tõenäolisem kui süsteemis on temaga positiivselt interakteeruv liik, ehk tema mutualist või kommensalist (MacKenzie et al., 2004). Seevastu liigid, kelle vahel on negatiivsed interaktsioonid (nt kisklus, konkurents), esinevad koos vähem tõenäoliselt kui juhuslikult moodustunud süsteemi korral arvata võiks (Mittelbach & McGill, 2019). Kooslustes võib esineda ka vahendatud mõju, mille korral liigid A ja B suhestuvad kolmanda liigi C kaudu (Ovaskainen et al., 2010). Liikudes komplekssemate süsteemide suunas, tuleb meeles pidada, et ühe liigi mõju teisele sõltub ka kontekstist (ressursid, teised liigid). Näiteks Californias invasiivse musta kapsasrohu (*Brassica nigra*) kasvule mõjus soodsalt, kas pärismaise tatralise liigi (*Eriogonum fasciculatum*) või pärismaise pujuliigiga (*Artemisia californica)* koos esinemine. Seevastu kui must kapsasrohi esines koos mõlema pärismaise liigiga, oli neil kapsasrohu kasvule pärssiv mõju (Schlau et al., 2023).

Üheks oluliseks teguriks, mis võib viia liikide esinemismustrite erinemiseni juhuslikest, on nende keskkonnaeelistused (Gotelli & McCabe, 2002). Vaid sarnaste abiootiliste vajadustega liigid on võimelised koos esinema, kombineerudes liikide omavaheliste vastasmõjudega. Keskkonnatingimustest tulenevate koosesinemistega arvestamiseks on võimalik analüüsida liigile sobivaid tingimusi, näiteks temperatuuri, mullaniiskust, toitainesisaldust (Pearson et al., 2006) liikide koosesinemiste modelleerimisel kasutatakse sageli nišimudeleid ja kaasatakse keskkonnatunnuseid, mis aitavad hinnata liigi potentsiaalset levikut, võimaldades uurida,

millist rolli mängivad keskkonna sobivus ja liikide vastastikmõjud koosesinemise kujunemisel (Elith & Leathwick, 2009; MacKenzie et al., 2004).

Varasemad teooriad (Diamond, 1975; Macarthur & Levins, 1967) rõhusid konkurentsi ja niši erinevuste olulisusele koosluste moodustumisel. Näiteks Diamond (1975) uuris Bismarki saarestiku linde ja tuvastas liike, kes kunagi koos ei esinenud – olukorra selgitamisel rõhus ta liikidevahelisele konkurentsile. Järgnes arutelu, mille käigus käsitleti nii koosesinemise mustrite mittejuhuslikkust kui ka selle tekkepõhjuseid (Connor & Simberloff, 1979; Gotelli & McCabe, 2002). Uuemad uuringud väidavad, et liikide mittejuhuslik esinemine tuleneb mitmetest põhjustest, kusjuures roll on ka soodustavatel ja vahendatud vastasmõjudel (Bruno et al., 2003).

1.1.1 Liikide koosesinemine erinevatel ruumiskaaladel

Koosesinemisi määravate protsesside mõju on eri ruumiskaaladel erinev. Saame eristada kahte peamist ruumilist taset – lokaalne ja regionaalne tase (He et al., 2005; Mittelbach & McGill, 2019). Mõned autorid eristavad ka indiviidi ja maastiku taset (Crist et al., 2003; Gleason, 1926). Lokaalsel skaalal domineerivad ökoloogilised vastasmõjud, troofiliste tasemete vahelised suhted ja ressursside saadavus (Chesson, 2000; HilleRisLambers et al., 2012; McIntire & Fajardo, 2014). Seevastu regioonis esinemiseks on oluline liigi levimine kooslusesse ja sealsete keskkonnatinimuste sobivus liigile (Pauchard & Shea, 2006), st võime öelda, et peamisteks koosesinemist mõjutavateks faktoriteks regiooni tasemel on biogeograafilised põhjused, nagu laamade liikumised, kliimamuutused, erinevate elupaiganõudlustega liikide teke ja väljasuremine (Zobel et al., 2011; Weigelt et al., 2015).

Piirkonna topograafia, sidusus ja maakasutus mõjutavad oluliselt liikide levikut, ning peegelduvad seeläbi ka ökoloogilistes kooslustes (Ronk et al., 2015). Isegi kui kaks liiki sobiksid samasse keskkonda, ei pruugi nad tegelikult koos esineda, kui üks liikidest ei ole suutnud kooslusesse levida. Levikupiirangud tekitavad mustreid, mis ei kajasta ei konkurentsi ega keskkonnaeelistusi, vaid pigem biogeograafilisi protsesse. Näiteks paljud Euroopa metsade liigid ei ole viimase jääaja järgselt oma refuugiumitest tagasi klimaatilise piirini levinud, mistõttu need liigid ei esine koos liikidega, kes on paremad levijad (Svenning et al., 2008).

1.1.2 Liikide koosesinemine erinevatel ajaskaaladel

Liikide koosesinemine võib olla määratud minevikus toimunud sündmusest. Sageli võib kooslusi mõjutav sündmus olla lühikese ajalise kestvusega, aga tema mõju ilmnemine võtab aega (Piqueray et al., 2011). Seetõttu tuleb koosluse uurimisel meeles pidada ka piirkonna ajalugu, mida illustreerib ilmekas uuring, kuidas esinesid kividel koos rohe- ja punavetikad (Sousa, 1979). Rohevetikad kui head levijad asustasid kive esimesena, aga aja jooksul hakkas punavetikas kui parem konkurent neid välja tõrjuma. Kuigi nende liikide vaheline konkurents ei luba koosesinemist, võib see juhtuda kui punavetikal ei ole olnud piisavalt aega rohevetika välja tõrjumiseks. Kooslused kividel uuenesid kui lained neid ümber keerasid, kusjuures väikeste kivide kooslused uuenesid tihedamini, sest neid oli lainetel lihtsam ümber keerata. Seetõttu olenes kivil vaadeldavate vetikate koosesinemine sellest, palju aega oli möödunud viimatisest mõjukast lainetusest. Selle näite põhjal võime järeldada, et uurija, kes pole teadlik kividel esinevate koosluste seosest lainetusega võib lihtsasti jõuda valejäreldusele, sest ta ei oska arvestada uuritava piirkonna ajalooga.

Keskkonna muutumisel võib tekkida olukord, kus varasema keskkonna liigid pole veel jõudnud välja surra aga uued liigid on juba kohal – seda nimetatakse väljasuremise võlaks (*extinction debt*) (Cristofoli et al., 2010). Väljasuremise võla ajal esinevad koos liigid, kes stabiilses olukorras koos ei esineks. Cowlishaw (1999) hindas primaatide väljasuremisvõlga Aafrika metsades. Toodi välja, et primaatide liigiline mitmekesisus ei paistnud muutuvat, kuigi neile sobivate tingimustega alasid oli vähemaks jäänud. Cowlishaw leidis liike ja neile sobivaid metsasid mudeldades, et raie tõttu on enamikes riikides üle 30% primaatidest väljasuremisvõlas. Ta rõhus, et ei tohi eeldada liigi jätkusuutlikust ainult tema elupaiga kaitstusest tulenevalt.

Asustamise krediit (*colonization credit*) kirjeldab levimispiirangute tõttu tekkivat ajalist viivet koosluse moodustumises, mistõttu sobivaks muutunud koosluses ei esine kohe kõiki võimalikke koosesinemisi (Cristofoli et al., 2010). Naaf ja Kolk (2015) uurisid nii asustamise krediiti kui ka väljasuremise võlga Saksamaal Prignitzi regiooni metsades, kus osad metsalapid olid endised põllumaad ja osad olid põlismetsad. Metsade hiljutise pindala vähenemise tõttu eeldati, et võib esineda väljasuremise võlg (st põlismetsade kõik liigid pole jõudnud välja surra), samas põllumaade taasmetsastumise tagajärjel pidasid autorid võimalikuks ka asustamise krediidi esinemist (kõik liigid pole veel uude metsa jõudnud). Täheldati, et endistel põllumaadel asuvate metsade liigirikkus oli oodatust madalam ning eriti eraldatud

metsatukkadel ulatus asustuskrediit kuni üheksa liigini. Uuringus näidati, et ka rohkem kui 130-230 aastat hiljem esines metsades asustamiskrediit.

Kokkuvõtvalt võib öelda, et liikide koosesinemiste uurimisel tuleb arvestada uuritava süsteemi ajalooga. Nii looduslikud kui ka inimeste vahendatud protsessid võivad viia ajas ebastabiilsete koosesinemisteni. Ebastabiilseid kooslusi on võimalik kirjeldada väljasuremise võla ja asustamise krediidi kaudu.

1.1.3 Liikide koosesinemine kui sisend tumeda elurikkuse uurimisse

Liikide koosesinemistel on oluline roll koosluse struktuuri ja funktsioneerimise tagamisel – mõistes, millised liigid kipuvad koos esinema ja miks, on võimalik paremini ennustada koosluste muutusi ka elupaikade kadumise või kliimamuutuste kontekstis. Näiteks teatud liigi kadumisel on ka temaga positiivselt koosesinenud liikide kadumine tõenäolisem (Tulloch et al., 2018). Üks oluline koosesinemiste uurimise väljund, mis võimaldab arvestada ka liikidevaheliste suhete aeg-ruumilise varieeruvusega, on tume elurikkus (M. Pärtel et al., 2011).

Tume elurikkus on kogum liike, mis võiksid kooslusesse ökoloogiliselt sobida, aga mingil põhjusel neid seal pole (M. Pärtel et al., 2011). Tume elurikkus – erinevalt vaadeldud elurikkusest – lubab elurikkust mõõta osakaaluna koosluse potentsiaalsest elurikkusest (liigifondist), ning võimaldab seega omavahel võrrelda erinevaid koosluseid, regioone, taksonoomilisi üksuseid (M. Pärtel et al., 2011). Enimkasutatud tumeda elurikkuse hindamise meetod kasutab liikide koosesinemisi, leides liigid, mis olemasolevate liikidega sageli koos esinevad, aga puuduvad uuritavast kooslusest (M. Pärtel et al., 2025).

1.2 Liikide koosesinemise matemaatika

Koosesinemiste kvantitatiivseks hindamiseks on kasutusel mitmeid matemaatilisi meetodid. Koosesinemisi hinnatakse paariviisiliselt või kõikide liikide-uurimisalade maatriksi põhjal (Veech, 2014). Liikide esinemise kohta kogutud andmeid võrreldakse sobiliku nullmudeliga, et teada saada, kas uuritavates andmetes esinevad koosesinemised erinevad statistiliselt oluliselt koosesinemistest, mida võiksime näha ka juhuslikult (Gotelli, 2000). Näiteks Connor ja Simberloff (1979) võrdlesid koosesinemiste maatriksit Monte Carlo meetodil loodud mudeliga. Nad kasutasid sarnast andmestikku nagu see, mille põhjal Diamond (1975) oli väitnud konkurentsi olulisust koosluste moodustumisel. Vastupidiselt Diamondile jõudsid Connor ja Simberloff järeldusele, et andmestikus täheldatud mustrid ei erine oluliselt juhuslikest.

Hüpergeomeetriline jaotus on üks võimalus koosesinemiste paariviisiliseks analüüsimiseks (Veech, 2013). Meil on võimalik leida tõenäosus (*P*), et uuritavad liigid esinevad koos sagedamini või harvemini kui juhuslikult. Hüpergeomeetriline jaotus on defineeritud valemiga

$$P(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}, \tag{1}$$

mis iseloomustab sündmuse toimumist täpselt *k* kordadel *n*-ist, kus *N* on kõigi sündmuste arv ja *r* on soovitud sündmuse arv (Wagaman & Dobrow, 2021). Koosesinemiste kontekstis on *r* ja *n* liikide esinemiskorrad, *N* on uurimisalade arv ja *k* on valimi suurus (Carmona & Pärtel, 2021; Wagaman & Dobrow, 2021). Eeltoodud valemis tähistavad sulud kombinatsioone, näiteks $\binom{r}{k} = C_k^r$. Seega leiame, kui palju võimalusi on valida uuritavast hulgast *r* objekti *k* kaupa. Saadud tulemuste võrdlemiseks on võimalik neid standardiseerida, kasutades näiteks z-väärtust (*z-score*) (Keil, 2019). Z-väärtus iseloomustab, mitme standardhälbe kaugusel on andmepunkt keskmisest (Díaz et al., 2022).

1.3 Liikide koosesinemise andmete kogumise meetodid

Jõudmaks ökoloogia eesmärgini mõista liikide koosesinemisi looduse eri ruumi- ja ajaskaaladel on vajalik mahukas ning mitmekülgne andmestik (Wüest et al., 2020). Andmete päritolu põhjal on meil võimalik eristada spetsiifilise teadusliku välitöö kampaania käigus kogutud andmeid, harrastusteadusest pärinevaid andmeid ning loodusteaduslike kogude digiteerimise kaudu saadavaid andmeid (Farley et al., 2018). Erinevaid andmeid elusloodusе mõistmiseks on kogutud juba sajandeid ning arenev tehnoloogia avab veel uusi võimalusi, kuidas selliseid andmeid koguda – alates ülemaailmsetest andmebaasidest kuni uusimate sensoriteni (Lürig et al., 2021; Mittelbach & McGill, 2019).

1.3.1 Koosluste kirjeldused

Enamik kaasaegseid ökolooge koguvad andmeid samamoodi nagu sadakond aastat tagasi – välitööde käigus koostatakse vaatlusalasid iseloomustavaid liiginimekirju (Mittelbach & McGill, 2019). Koosesinemiste uurimisel eristatakse metoodikast tulenevalt kahte tüüpi andmestikke. Esiteks, lühiajalise vaatluse käigus kogutud andmed, kus vaatluse määratlemiseks kasutatakse taimeruute, transekte, püünised vms ja saadud andmestik on „proov“ tegelikkusest (Gotelli, 2000). Teiseks, peaaegu täielikud liiginimekirjad saartelt või

kindlalt eristatud uurimisaladelt, mida on eelkõige „populaarsete“ ja lihtsamini uuritavate liikide kohta, nagu selgroogsed (Gotelli, 2000). Klassikalise välitöö tugevuseks on selle detailsus ja usaldusväärsus, samas on sellised andmed sageli ebastandardsed ja nende kättesaadavus piiratud (Parsons et al., 2011).

Tulenevalt ruumilise mõõtme olulisusest ökosüsteemide kirjeldamisel on oluline koondada vaatlusandmeid ka ülemaailmsel skaalal (LaDeau et al., 2017; Michener & Jones, 2012). Sealjuures on osad platvormid andmete hulga suurendamiseks kaasanud ka harrastusteadlasi. Andmebaasid nagu Global Biodiversity Information Facility (GBIF), iNaturalist, eElurikkus/PlutoF koondavad suurel hulgal andmeid, võimaldades seeläbi hinnata näiteks ökosüsteemide vastupanuvõimet erinevates oludes ning teha üldistatavaid järeldusi elurikkuse säilitamise kohta jms. Näiteks GBIF koondab üle kolme miljardi vaatluse ning 2024 jaanuaris ületas andmebaasil põhinevate eelretsenseeritud artiklite arv kümmet tuhandet (GBIF, 2024). Andmebaasidesse koondatud info tugevuseks on selle struktureeritus ja laiem kättesaadavus – GBIF lähtub andmete säilitamisel rahvusvahelisest standardist – ning andmed on juurdepääsupiiranguta (GBIF, 2025).

Loodusteaduslikud kollektsioonid, näiteks herbaariumid on üks andmeallikas koosesinemiste kirjeldamisel – eriti palju on neis potentsiaali koosluste ajaloo uurimisel (Daru, 2025). Herbaariumid, kuhu on kogutud läbi ajaloo lai valik teatud piirkonna liike, võimaldavad ka taasluua liikide varasemaid levilaid (Daru, 2025). Näiteks Tartu Ülikooli loodusmuuseumi botaanilistes kogudes on hoiul üle kolmesaja tuhande herbaareksemplari ning neid lisandub aastas enam kui tuhat. Tartu Ülikooli herbaareksemplaaridest osa on digiteeritud ning nende kohta saab päringuid teha portaalis eElurikkus (Tartu Ülikooli loodusmuuseum, 2022). Vanade andmete digiteerimine võimaldab neid kasutada analüüsides, andes seeläbi teadmisi liikide levilate muutustest ajas ning inimmõjust neile (Daru, 2025).

1.3.2 Liikide koosesinemiste andmete automatiseeritud kogumine

Liikide koosesinemise kohta andmete kogumiseks on arendatud välja erinevaid poolautomatiseeritud ja automatiseeritud lahendusi (Besson et al., 2022). Uuenduslikud tehnoloogiad võimaldavad kiiresti koguda kvaliteetseid ja standardiseeritud mitmemõõtmelisi andmeid (Besson et al., 2022) ning parendavad andmete esindatust eri ruumiskaalade, bioomide ning taksonite lõikes (Wüest et al., 2020).

Suureskaalaliste andmete puudujääki on aidanud täita kõrgresolutsiooniga kaamerate ja sensoritega varustatud mehitamata õhusõidukite kasutuselevõtt (Besson et al., 2022; Wüest et al., 2020). Eriti on neist kasu elupaikade kaardistamisel, keskkonnamuutuste hindamisel, aga ka keskkonnale kahjulike ebaseaduslike tegevuste tuvastamisel (Besson et al., 2022). Võimalikke rakendusi on nii zooloogiliste kui ka botaaniliste andmete kogumisel, näiteks Delplanque *et al.* (2022) kasutasid droone ja masinõpet Aafrika imetajate kaardistamiseks ja tuvastamiseks. Kombineerides masinõppe meetodeid ortofotodega on võimalik ka suurte alade taimekoosluste analüüsimine (Wüest et al., 2020).

Veel üheks võimalikuks andmete kogumise meetodiks on akustiliste sensoritega keskkonnahelide salvestamine, näiteks saab salvestada linnulaulu, konnade või putukate häälitsusi (Sethi et al., 2020). Sellisel viisil kogutud andmed võimaldavad visuaalselt raskesti jälgitavate liikide uurimist, bioloogilise mitmekesisuse analüüsimist ja liikide olemasolu tuvastamist (Besson et al., 2022). Näiteks LeBien *et al.* (2020) töötasid välja konvolutsioonilist närvivõrku kasutava lahenduse, millega tuvastada troopika helisalvestistest linnu- ja konnaliike.

Uuenduslik ja kiiresti arenev viis liikide koosesinemise andmete kogumiseks on keskkonna DNA (eDNA) analüüs. Liigid jätavad oma elutegevuse käigus keskkonda liigispetsiifilist geneetilist materjali, mida on võimalik molekulaarsete meetoditega analüüsida (Besson et al., 2022). Tehnoloogiline areng on võimaldanud eDNA proovide automaatset kogumist ja töötlust. Kuigi DNA järjestamine toimub hetkel veel eraldi etapis, on arendamisel seadmed, mis kasutavad kaasaskantavaid järjestamistehnoloogiaid (nt MinION, SmidgION), mis tulevikus võimaldaksid eDNA täielikult automaatset analüüsi (Besson et al., 2022; Huang et al., 2022).

Ökoloogias on saadavalolevate andmete kogus viimaste aastate jooksul eksponentsiaalselt kasvanud, kusjuures mõningad autorid väidavad isegi, et teadlased „upuvad“ andmetesse (Parsons et al., 2011; Wüest et al., 2020). Uued tehnoloogiad on teatud määral aidanud täita vajakuid ökoloogilistes andmestikes, kuid mitte täielikult (Wüest et al., 2020). Võime öelda, et on tekkinud paradoks, kus andmeid on samal ajal liiga palju ja ka liiga vähe (Smits et al., 2025).

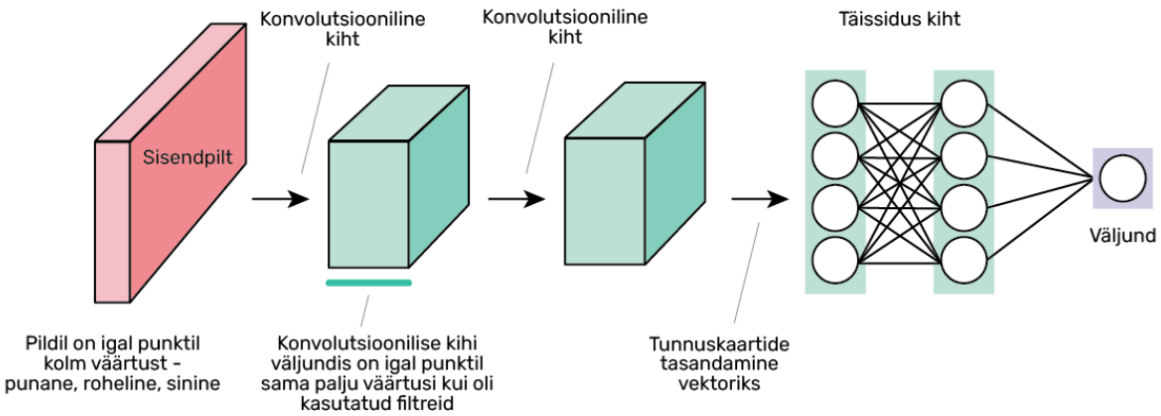
1.4 Tehisaru liikide koosesinemiste leidmisel

1.4.1 Ülevaade tehisarust

Tehisaru termin on üsna lai – kodumasinatest kuni vestlusrobotiteni välja. „[Tehisaru on] tarkvara või isekäituv seade, mis suudab konkreetse ülesande lahendamiseks tunnetada väliseid sündmusi ja neile eesmärgipäraselt reageerida“ (Eesti Keele Instituut, 2025). Üldiselt huvitavad loodusteadlasi eelkõige need tehisaru valdkonnad, mis kaasavad endas ka teatud andmestikust õppimist ja sealt mustrite otsimist – neid valdkondi nimetame masinõppeks (*machine learning*) (Thessen, 2016). Tuntuimad masinõppe lahendused on näiteks Google Assistant, Alexa, YouTube’i videosoovituste algoritm ja isejuhtivad autod, aga masinõppel on palju potentsiaali ka loodusteadustes (Sügis et al., 2024; Thessen, 2016). Lihtsustatult võime öelda – masinõppe valdkonda eristab ülejäänud tehisarust õppimine ehk uutes olukordades vajadusel oma käitumise muutmine (Sügis et al., 2024). Näiteks robottolmuimeja ei liigitu masinõppe alla just seetõttu, et ta käitub ainult vastavalt juhenditele „puhasta põrandat“ ja „väldi mööblit“, aga teda ei ole võimalik õpetada näiteks nõusid pesema.

Teatud gruppi algoritme masinõppe valdkonnas nimetatakse tehisnärvivõrkudeks (*artifical neural networks)*. Tehisnärvivõrkude ülesehitus põhineb loomade ajudel ja nende alamosasid nimetatakse samuti neuroniteks (Tuffery, 2022). Kuigi – segaduse vältimiseks tuleb siin meeles pidada, et tehisnärvivõrgud kui algoritmid on lihtsalt juhised teatud arvutuste järkjärguliseks läbiviimiseks. Väga levinud tehisnärvivõrk – eriti piltidega tegelevates valdkondades – on konvolutsiooniline närvivõrk (lühendatult CNN, inglise keeles *convolutional neural network*) (Ghosh et al., 2020). Sõna „konvolutsiooniline“ tähistab matemaatilist operatsiooni, mis kombineerib kaks signaali, et saada kolmas (Smith, 1997). CNN-i kontekstis on signaalide asemel kihid – sisendpildile rakendatakse konvolutsioonilist kihti ja saadakse väljundkiht. Konvolutsioonilised kihid koosnevad maatriksitest ehk filtritest. Filtrid esindavad tunnuseid, mida närvivõrk pildilt otsib. Filtri rakendamiseks leitakse tema ja sama suure pildiosa skalaarkorrutis. Kui kõiki filtreid on rakendatud kõigile pildi osadele, saadakse väljund, millele omakorda filtreid rakendatakse (Sügis et al., 2024). Lihtsamal juhul rakendabki närvivõrk

järjest erinevaid filtreid („otsib“ järjest keerulisemaid tunnuseid), lõpuks seob saadud tulemused (täissidus kiht) ja annab väljundi – pildil oleva objekti klassi (Joonis 1). Mudelit saab paremaks teha optimeerides selle parameetreid, sh filtrites sisalduvaid väärtuseid (Sügis et al., 2024).



Joonis 1. Konvolutsioonilise tehisnärvivõrgu ülesehitus (Sügis et al., 2024).

1.4.2 Masinõpe ökoloogiliste andmete koondamisel

Sajandeid on kogutud erinevaid andmeid elusloodusе mõistmiseks ning uueneva tehnoloogiaga kasvab andmete hulk veelgi (Lürig et al., 2021). Tulenevalt saadaval olevate andmete mahu, mitmelaadilisuse ja koguse eksponentsiaalsest kasvust on (makro-)ökoloogiat hakatud käsitlema ka suurandmete kontekstis (Jarić et al., 2020; LaDeau et al., 2017). Kandes üle teadmisi andmeteadusest, on võimalik välja tuua mitmeid soovitusi, kuidas ökoloogias suurandmeid edukamalt kasutada, – alates metaandmete dokumenteerimisest, standardiseeritud protokollide, taksonoomia ja andmeplatvormideni välja (Farley et al., 2018; Parsons et al., 2011; Smits et al., 2025; Thessen, 2016). Potentsiaalne osa lahendusest on masinõpe, mis võimaldab nii andmete annoteerimist, sorteerimist kui ka analüüsimist (Han et al., 2023; LaDeau et al., 2017).

Masinõppe kaasamine on võimalik mitmes projekti faasis (Sügis et al., 2024). Esiteks on võimalik poolautomatiseerida andmete kogumist ja eeltöötlemist. Näiteks Folk et al. (2024) rakendasid masinõppe meetodeid efektiivsemaks botaanilise kirjanduse analüüsiks. Nad kasutasid keelemudelit, et leida botaanilistest kirjeldustest taimede mõõtmeid, erinevate struktuuride kirjeldusi. See lähenemine võimaldas analüüsida andmeid, mis on juba olemas, aga vajavad ühtlustamist.

Lisaks masinõppe kaasamisele andmete kogumisel, on võimalik selle abil automatiseerida ka projektide järgmisi faase: andmete kirjeldamist ja visualiseerimist, ning analüüsimist (Sügis et al., 2024). Ilmekas uuringus õpetati tehisaru iNaturalist andmebaasi põhjal ennustama liikide levilaid ning tulemusi hinnati ekspertide koostatud levikukaartide võrdlusel (Cole et al., 2023). Loodud mudel osutus paljulubavaks, kuid rõhutati, et eesmärk oli uurida masinõppel põhinevate lahenduste potentsiaali liikide levila hindamisel, mitte luua rakendatavaid levikukaarte.

Kuigi masinõpe lisab ökoloogilistele uuringutele märkimisväärset potentsiaali, seisavad selle laialdasemal rakendamisel ees mitmed praktilised ja metoodilised takistused – masinõppe meetodid on üsnagi ligipääsetavad, aga nende paremaks integreerimiseks loodusteadustesse on vajalik interdistsiplinaarne koostöö ja suhtlus, parem rahastus ning haridus (Thessen, 2016). Lisaks on puudu kvaliteetseid, annoteeritud andmeid, mille põhjal mudeleid treenida (Brito, 2010; Zhai et al., 2024).

1.4.3 Taimede pildi järgi määramine tehisaruga

Taimede määramine tehisaruga võimaldab kaasata andmete kogumisel ja annoteerimisel teadushuvilisi, kellel puudub botaaniline haridus, vähendades seeläbi teadlaste töökoormust (Jones & Jones, 2025). Sarnastel meetoditel on ka olulised rakendusvõimalused põllumajanduses (Picon et al., 2022). Taimede määramise lihtsustamine suurendab ka üldiselt kaasatust botaanikas, isegi kui tegevusel puudub akadeemiline väljund (Jones & Jones, 2025). Android telefonile taimemääramise rakendusi otsides saame vastetena näiteks Flora Incognita, Blossom – Plant Identifier, PlantSnap, Plantum: AI Plant Identifier ja Plantin, mis kõik võimaldavad kasutajal tehisaru abiga taimi määrata, osad võimaldades ka taimehaiguste määramist (Google Play, 2025) .

Üks võimalus taimede määramiseks fotodelt on rakendus PlantNet (Pl@ntNet, 2025a). PlantNeti portaal kasutab masinõppe meetodeid, et määrata kasutajate fotodel olevaid taimi. Rakendust on võimalik kasutada nii veebis kui ka telefonis. Varasemalt kasutas PlantNet automaatsel määramisel CNN meetodeid, nüüd liigutakse keerulisemate ja võimekamate tehisnärvivõrkude – visuaalsete transformerite (*vision transformers*) kasutamise suunas (Pl@ntNet, 2023). Kasutajal on võimalik piirata oma määranguid sobiva piirkonnaga, näiteks Ida-Euroopaga. Lisaks on võimalik täpsustada, milline taimeosa on fotol (leht, õis, vili, koor, muu), võimalik on lisada ühe päringu kohta kuni viis fotot. Vastusena pakub PlantNet võimalikud määrangud koos nende tõenäosusega, misjärel on kasutajal võimalik valida

korrektne määrang vaatlusena esitamiseks (Pl@ntNet, 2025a). Võimalik on teiste kasutajate määranguid kinnitada või ümber lükata. Määranguid ja nendele antud tagasisidet kasutatakse mudeli treenimiseks (Lefort et al., 2025).

Soovides liikuda ühe taimeliigi määramise pealt edasi koosesinemiste andmetele on peamiseks tehniliseks probleemiks vajadus kombineerida nii määramine kui ka segmenteerimine (Goëau et al., 2024). Lahendusi AI-ga taimeruutude fotodelt mitme liigi määramiseks teostati PlantCLEF 2024 võistluse raames. Võistkonnad said ülesandeks luua AI mudel, mis suudaks taimeruutude fotodelt korrektselt määrata võimalikult palju liike. Segmenteerimine vähendab keerulisust, sundides mudelit keskenduma olulistele tunnustele ning aitab vältida näiteks olukorda, kus mudel hakkab seostama tausta/pinnast teatud määranguga (Dyrmann et al., 2016). Taimede puhul aitab segmenteerimine vältida ka olukorda, kus erinevate liikide tunnused on piltidel kõrvuti. PlantCLEF 2024 võistluse parim mudel kasutas valepositiivsete määrangute eemaldamiseks segmenteerimist. Võistluse ülevaates kirjutatakse, et automaatne taimeruutude määramine on keeruline ja tuuakse võimaliku lahendusena välja määranguteta taimeruudu fotodel treenimise. Määranguteta taimeruudu fotosid on odavam toota, aga see eeldab juhendamata (*unsupervised machine learning*) või isejuhendatud õppe (*self-supervised learning*) mudelite loomist (Goëau et al., 2024).

2. METOODIKA

2.1 Andmete allikad ja tehtud uurimuse ülevaade

Käesoleva töö uurimusliku osa üldine eesmärk oli saada esialgne ülevaade, kas ja kuidas on võimalik Eesti andmetel kasutada koosesinevate taimeliikide tuvastamiseks tehisaru abi. Ülevaate saamiseks loodud uurimuse täpne eesmärk oli hinnata tehisarul põhineva taimemääramisprogrammi PlantNet võimekust määrata liike Eesti puisniitude rohurinde – kui väga liigirikka taimekoosluse (Wilson et al., 2012) – taimeruutude fotodelt. Määramise automatiseerimiseks kasutatud rakendus PlantNet on praegusel kujul mõeldud ühelt fotolt määrama ühte liiki, mistõttu oli uurimuses fotode segmenteerimist simuleeriv aspekt, mis seisnes fotode ristkülikuteks lõikamises. Segmenteerimise asendamine robustse lõikamisel põhineva lähenemisega valiti nii lõikamisprotsessi kui ka sellest tuleneva analüüsi lihtsustamiseks. Taimeruutude fotosid väiksemateks osadeks lõigates on tõenäosus, et üks liik on saadud fotoosadelt selgemini eristatav ning tehisaru jaoks lihtsamini määratav – samas tuleks taimefotode lõikamisel vältida taimede oluliste määramistunnuste sattumist mitme fotoosa piirialadele. Püstitati hüpotees, et fotode osadeks lõikamisel kasutatavad parameetrid mõjutavad automaatse liigimäärangu tulemusi. Optimeerimaks nii fotoosade suurust kui ka määramistunnuste paiknevust, kasutati uurimuses kahte parameetrit – kui mitmeks osaks taimeruudu pilt lõigati ja kas lõigatud piltidel lubati ka servapidi kattuda.

Taimeruutude fotode lõikamisparameetrite mõju automaatse liigimäärangu tulemustele uuriti kasutades varasemalt muude projektide raames koostatud andmeid. Taimeruutude fotosid oli kokku 84 tk ja need pärinesid Triin Reitalu ja Elle Roosaluste töörühmade välitöödelt, ning igal fotol olevad liigid olid juba välitööde käigus määratud. Kõik fotod, mida uurimuslikus osas taimemääramisprogrammi võimekuse hindamiseks kasutati, olid tehtud taimeruutudest 1 x 1 m, aga fotodele oli jäänud ka ruutude kõrval olevaid taimi ning varieerus ka kaamera nurk maapinna suhtes. Liigimäärangud olid kantud Microsoft Exceli tabelitesse. Reitalu jagatud fotod olid tehtud avatud puisniitudel ja kinni kasvanud puisniitudel (osad sarnanesid juba salumetsadega) – rohkemal või vähemal määral kinni kasvanud aladel, mida kavatsetakse taastada. Fotod olid tehtud projekti Woodmeadowlife raames 2023 aasta suvel ja olid täistatud tähtnumbriliste koodidega. Reitalu töörühma andmestik sisaldas 46 fotot. Roosaluste jagatud fotod olid tehtud Laelatu puisniidu taastatud aladel. Alad taastati aastal 2020 ja fotod pärinevad 2024. aasta suvest. Fotod olid nummerdatud ning neid oli 38 tk.

Taimeruutude fotodel segmenteerimise simuleerimiseks lõigati neid R-i koodiga (Lisa 1) ristkülikuteks, mis olid defineeritud parameetritega *k* ja *overlap.* Analüüsi korrati erinevate parameetrite väärtustega. Parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse – st võimaldas muuta, mitu väiksemat fotot taimeruudust saadakse (Joonis 2). Uurimuses kasutatud *k* väärtused olid 4, 5 ja 6. *Overlap* iseloomustas foto lõikamisel kattuvuse määra, kus võimalikud väärtused on vahemikus 0 kuni 0,5 ning uurimuses kasutati väärtuseid 0 ja 0,2. *Overlap* parameetri kaasamise eesmärk oli vähendada taimede määramistunnuste sattumist mitme fotoosa piirialadele. *Overlap* väärtuse 0 korral fotod lõigati ilma kattuvuseta ja tulemused erinesid ainult parameeter *k*-st tulenevalt. Nendest kahest parameetrist tulenevalt oli töötluste koguarv 6.

Taimeruutude fotoosad saadeti R-i vahendusel PlantNeti portaalile ühekaupa analüüsida. Uurimuses kasutati PlantNeti rakendusliidese ehk API kaudu (https://my-api.plantnet.org/?tags=my-api). Parameetri *project* väärtuseks määrasime "k-eastern-europe", eesmärgiga piirata võimalikke määranguid Ida-Euroopas esinevate taimeliikidega. Fotodel olid enamuses vegetatiivsed taimed, seega määrasime parameetri *organs* väärtuseks „leaf“, et rakendus otsiks fotodelt peamiselt lehtesid. Tasuta on PlantNeti rakendusliidesega võimalik teha 500 päringut ööpäevas. Teaduslikel eesmärkidel suurendas PlantNeti arendustiim antud töö tarvis lubatud päringute arvu 2000 peale ööpäevas. Kokku määrati 12 936 pildi osa, millele lisaks ka katsetused nii koodi kui ka parameetrite väärtustega, mistõttu tuli ka päevalimiiti arvestada ja jaotada tööd päevade kaupa. Päringu tulemusena tagastab PlantNet määratud liikide ladinakeelsed nimed ja määrangute tõenäosused. Uurimuses valiti määrangutest esimene, mis jäeti analüüsi alles kui selle tõenäosus oli 0,35 või rohkem. Tõenäosuse piir valiti pilootuuringu käigus.

Joonis 2. Taimeruudu foto, mille lõikamisel on kasutatud parameetreid *k*=6 ja *overlap*=0, kus parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse; *overlap* iseloomustas foto lõikamisel kattuvuse määra. Iga pidliosa kohal on kõige suurema täpsusega liigi landinakeelne nimi ja määrangu tõenäosus.

2.2 Andmeanalüüs

Hindamaks automaatse taimemääramisprogrammi võimekust määrata liike Eesti puisniitude rohurinde taimeruutude fotodelt võrdlesime välitööde käigus koostatud liiginimekirjasid samade taimeruutude fotode põhjal tehisaruga koosatud liiginimekirjadega. Analüüsi eeltööna ühtlustati välitööde raames tehtud määrangute andmeid, mis pärinesid kahelt eri töörühmalt ning olid seetõttu erinevates formaatides (Lisa 2. R-i skript liigitabelite vormistuse ühtlustamiseks. Taimeliikide nimetuste ühtlustamiseks kasutati veebirakendust Taxonomic Name Resolution Service (https://tnrs.biendata.org/), millega leiti ühtne sünonüümika nii kahele välitööde käigus koostatud liiginimekirjale kui ka uurimusliku osa käigus loodud andmestikule (Lisa 3. R-i skript liiginimede ühtlustamiseksÜhese sünonüümika leidmise käigus muudeti alamliigid liigini ja apomiktilised mikroliigid liideti perekonnaks: **kortsleht** (*Alchemilla*), **karutubakas** (*Pilosella*), **hunditubakas** (*Hieracium*), **võilill** (*Taraxacum*), **kibuvits** (*Rosa*), **murakas** (*Rubus*) puhul. Seejärel viidi uurimuses kasutavates andmestikes esinevad liigid sünonüümide nimekirja alusel ühtsesse nomenklatuuri ümber.

Analüüsi eeltöö ning analüüs ise viidi mõlemad läbi R-iga, kokku koostati neli R-i faili. Esimene R-i fail oli fotode lõikamiseks ja PlantNetiga määramiseks, teine fail oli korrektsete määrangute vormistuste ühtlustamiseks, kolmas oli liiginimede sünonüümide kõrvaldamiseks ja neljas andmete analüüsimiseks. Lähtekood oli Meelis Pärtli koostatud, seda muutsin vastavalt vajadusele, näiteks päringute piirangu lahendamiseks oli vaja teha täiendusi. Meelis Pärtel koostas ka Taxonomic Name Resolution Service'i abil taimenimede sünonüümika nimekirja ning aitas analüüsi osas ANOVA mudelite koostamisega. Programmeerimisprobleemide tuvastamiseks ja koodide toimetamiseks kasutasin generatiivse IT rakenduse abi. Kõik koodifailid on toodud töö lisades, rahvusvahelise koostöö võimaldamiseks on koodi kommentaarid inglise keeles.

Analüüsi läbi viimiseks tõsteti nii välitöödel saadud määrangud kui ka uurimusliku osa tulemused kujule, kus igale taimeruudule vastas sealt leitud liikide nimekiri. Iga taimeruudu kohta arvutati liiginimekirju võrreldes tõsipositiivsete, valenegatiivsete ja valepositiivsete määrangute arvud, mis tähistati vastavalt *correct*, *missed* ja *wrong*. *Correct* väärtused esinesid nii välitöödel kui ka pildianalüüsi käigus saadud liiginimekirjas. *Missed* väärtused esinesid algses nimekirjas, kuid puudusid uurimusega saadud nimekirjast ning *wrong* väärtused esinesid uurimuse nimekirjas, aga puudusid algsest nimekirjast. Fotode lõikamisel kasutatud parameetrite mõju hindamiseks automaatsetele liigimäärangutele arvutati kolm mõõdikut, mis

kõik saavad varieeruda 0 ja 1 vahel (Sügis et al., 2024). Täpsus (*precision*) iseloomustab mudeli positiivsete ennustuste usaldusväärsust ning on defineeritud valemiga

$$Precision = \frac{Correct}{Correct + Wrong}. \tag{2}$$

Saagis *(recall)* iseloomustab mudeli võimekust tuvastada tõsipositiivseid näiteid ning on defineeritud valemiga

$$Recall = \frac{Correct}{Correct + Missed}. \tag{3}$$

Lõpuks arvutati F1-skoor ehk F1-indeks (*F1-score*), mis iseloomustab mudeli võimekust saavutada tasakaal kõrge täpsuse ja kõrge saagise vahel – perfektse mudeli F1-skoor on 1. F1-skoor on defineeritud valemiga

$$F1_{score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \tag{4}$$

Overlap ja *k* mõju hindamiseks täpsusele, saagisele ja F1-skoorile kasutati lineaarset segamudelit (*lme*) R-i paketist *nlme*. Parameetrit kattuvus ehk *overlap* kasutati kujul *ovelap_yes*, ehk uuriti, kas kattuvuse olemasolu (*overlap*>0) lõikamisel mõjutab tulemusi. Parameetri *k* väärtuseid käsitleti mudelis diskreetsetena. Fotod, mida uurimuses taimemääramisprogrammi võimekuse hindamiseks kasutati, olid pärit kahelt eraldi tegutsenud töörühmalt; seega lisati kofaktorina andmestik (*dataset*), eraldamaks, kumma töörühma andmetega tegu on. Mudelitesse kaasati juhusliku faktorina vaatlusala ehk *code_site*, võimaldamaks arvestada pildi enda sobivust liikide määramisel. Valiti välja parimate parameetritega mudel, mille kohta toodi välja tehisaruga edukamalt ja vähemedukamalt määratud sagedasemad liigid. Liikide määratavuse hindamiseks leiti iga vähemalt kümnel juhul andmestikus esinenud taimeliigi kohta ruutude hulk, kus liik oli kohal ja leiti (*correct_count*); ruutude hulk, kus liik oli kohal aga ei tuvastatud (*missed_count*) ja ruutude hulk, kus liiki ei olnud, aga tehisaru selle liigi tuvastas (*wrong_count*). Seejärel arvutati liikide kohta määramise täpsus, saagis ja F1-indeks.

3. TULEMUSED

Parameetril *k* (ehk mitmeks osaks foto oli lõigatud) oli oluline mõju määrangute täpsusele (*precision*), saagisele (*recall*) ja F1-indeksile (*F1_score*) (Tabel 1, Tabel 2, Tabel 3). Fotoosade kattuvuse mõju saagisele osutus statistiliselt oluliseks (p=0,002), aga mõju määrangute täpsusele ja F1-indeksile ei olnud oluline. Muutuja andmestik ei omanud statistiliselt olulist mõju ühelegi hinnatud mõõdikule.

Tabel 1. Tehisaruga puisniitude rohurinde piltidelt määratud taimeliikide täpsuse seos piltide lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus iseloomustas pildiosade kattuvust kujul esineb/ei esine ning *k* iseloomustas saadud pildiosade arvu. Parameetri *k* väärtuseid käsitleti mudelis diskreetsetena. Andmestik kaasati kofaktorina, sest uurimuses kasutati kahte andmestikku.

	numDF	**denDF**	**F-väärtus**	**p-väärtus**
(vabaliige)	1	415	340,9	<0,0001
kattuvus	1	415	0,0	0,9599
k_faktor	2	415	15,5	<0,0001
andmestik	1	82	1,5	0,2207

Tabel 2. Tehisaruga puisniitude rohurinde piltidelt määratud taimeliikide saagise seos piltide lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus iseloomustas pildiosade kattuvust kujul esineb/ei esine ning *k* iseloomustas saadud pildiosade arvu. Parameetri *k* väärtuseid käsitleti mudelis diskreetsetena. Andmestik kaasati kofaktorina, sest uurimuses kasutati kahte andmestikku.

	numDF	**denDF**	**F-väärtus**	**p-väärtus**
(vabaliige)	1	415	215,6	<0,0001
kattuvus	1	415	9,9	0,0018
k_faktor	2	415	24,8	<0,0001
andmestik	1	82	2,4	0,1284

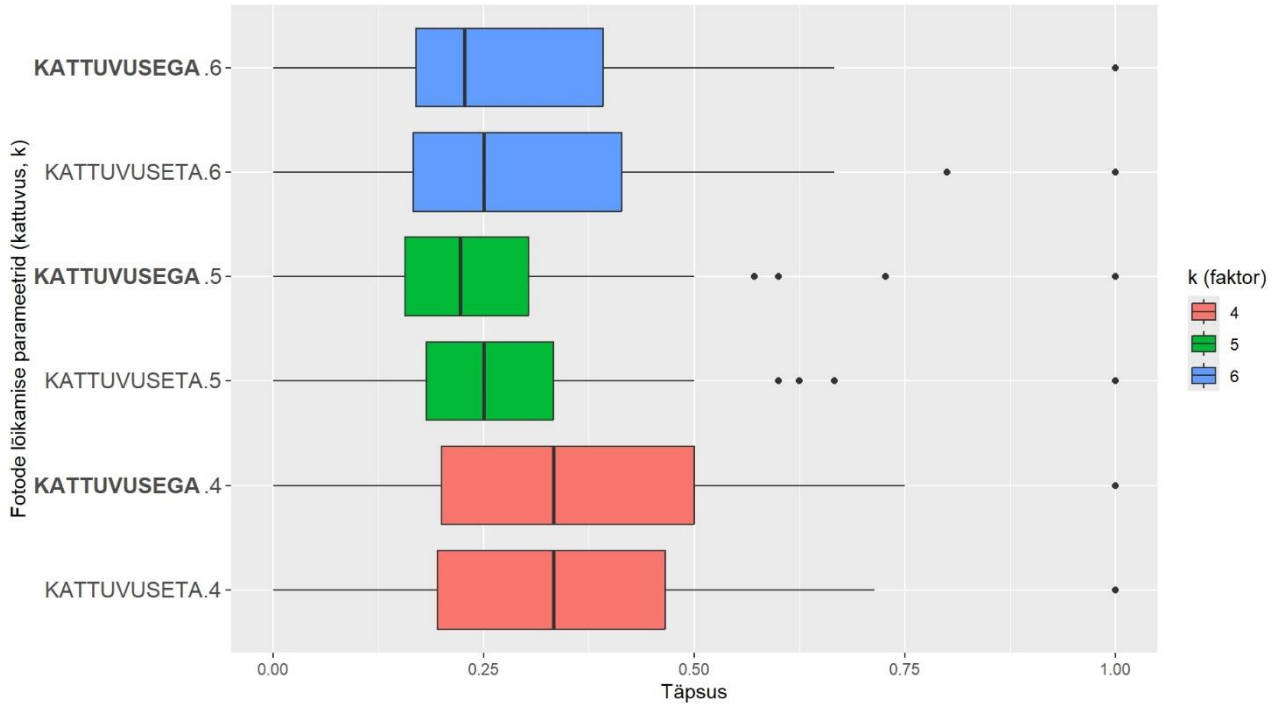
Tabel 3. Tehisaruga puisniitude rohurinde piltidelt määratud taimeliikide F1-indeksi seos piltide lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus iseloomustas pildiosade kattuvust kujul esineb/ei esine ning *k* iseloomustas saadud pildiosade arvu. Andmestik kaasati kofaktorina, sest uurimuses kasutati kahte andmestikku.

	numDF	**denDF**	**F-väärtus**	**p-väärtus**
(vabaliige)	1	415	419,5	<0,0001
kattuvus	1	415	2,2	0,1354
k_faktor	2	415	4,5	0,0112
andmestik	1	82	0,1	0,7023

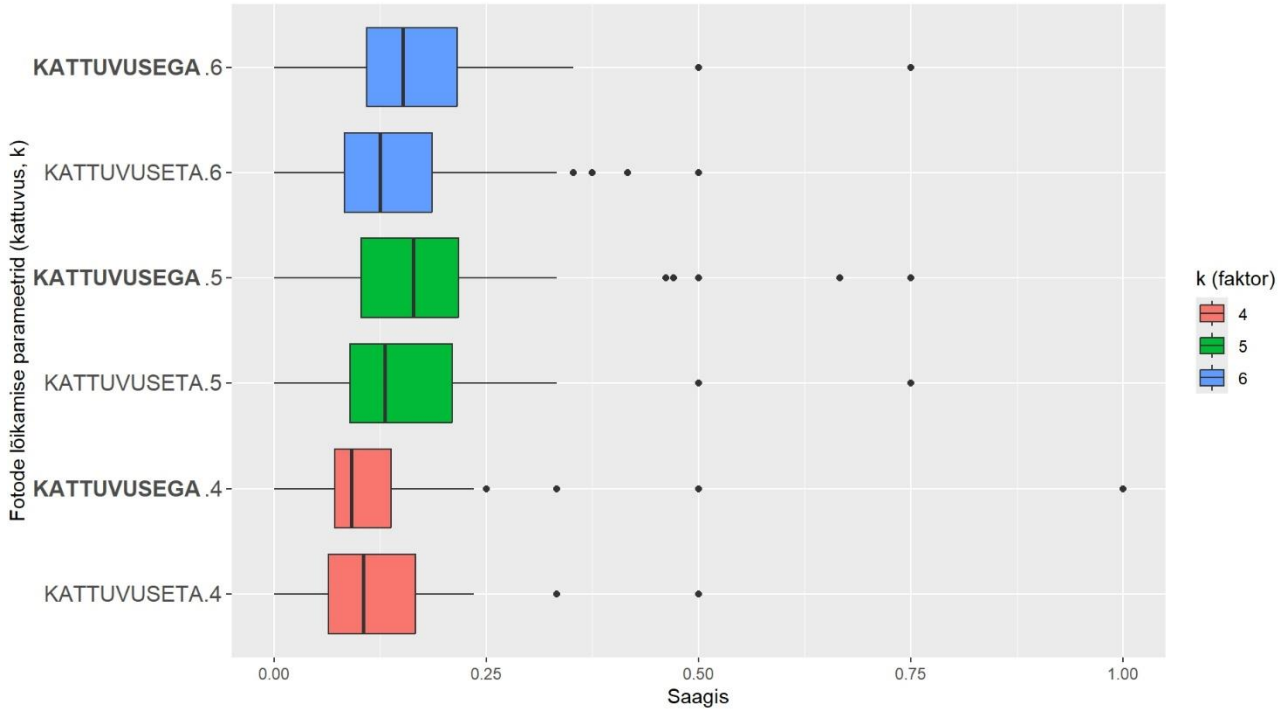
Parimad taimeruutude määramise täpsused saavutati kasutades parameetri väärtust *k*=4 (Joonis 3). Fotoosade kattuvuse (*overlap*) kasutamine või kasutamata jätmine kombineeritud *k*=4 väärtusega, saavutasid samad mediaantäpsused (mediaan=0,33). Suurim ülemine kvartiilväärtus, mis täpsuse hindamisel saavutati, oli 0,5, kasutades parameetrit *k=4* ja kattuvust. Väikseim alumine kvartiil, mis saavutati täpsuse hindamisel, oli 0,16, kasutades parameetrit *k=5* ja fotoosade kattuvust.

Parimad saagised taimeruutude määramisel saavutati kasutades parameetri väärtust *k*=5 ja fotoosade kattuvust, mediaansaagis oli sellel mudelil 0,16 (Joonis 4). Suurim ülemine kvartiilväärtus, mis saagise hindamisel saavutati, oli 0,22, kasutades parameetrit *k=5* ning fotoosade kattuvust. Väikseim alumine kvartiil, mis saavutati saagise hindamisel, oli 0,06, kasutades parameetrit *k=5* ja kasutamata fotoosade kattuvust.

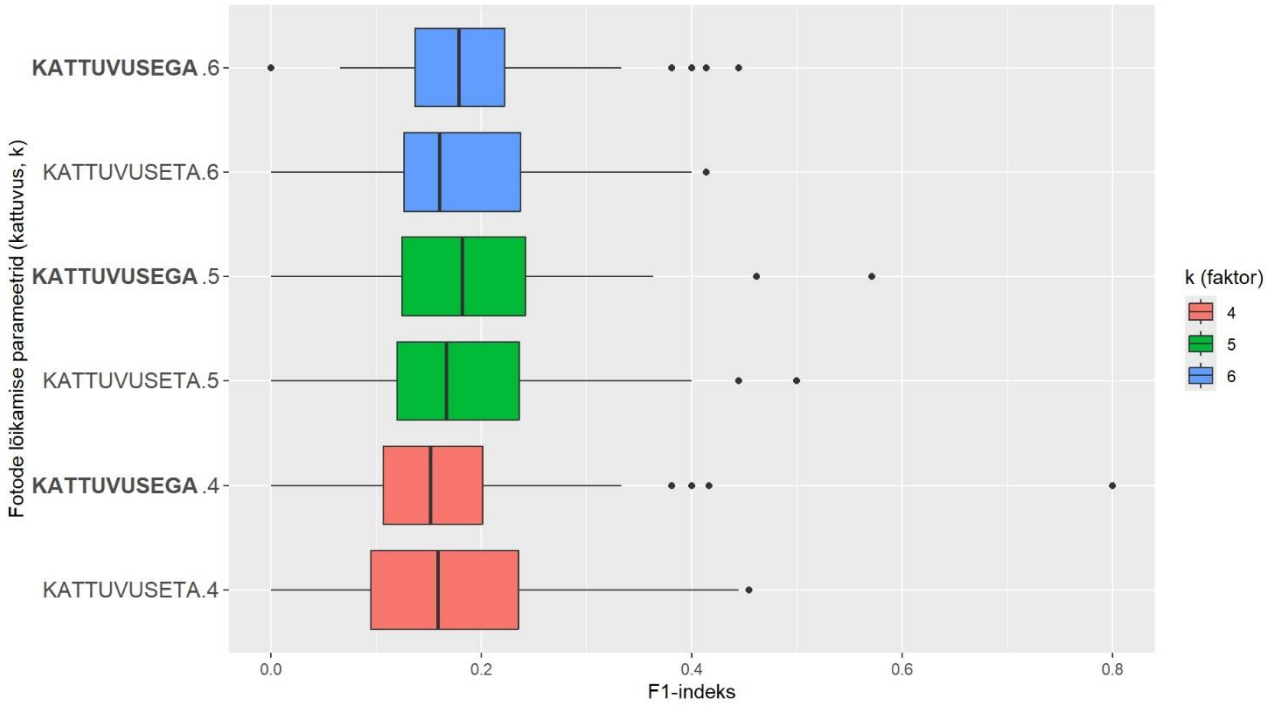
Parimad F1-indeksi väärtused taimeruutude määramisel saavutati kasutades parameetri väärtust *k*=5 ja fotoosade kattuvust (Joonis 5). Tulenevalt F1-indeksi võimekusest iseloomustada mudeli tasakaalu kõrge täpsuse ja kõrge saagise vahel, valiti parim mudel F1-indeksi põhjal. Parim mudel kasutas seega väärtust *k*=5 ja fotoosade kattuvust (*overlap*=0,2); ning selle mudeli mediaan F1-indeks oli 0,18 ja keskmine F1-indeks taimeruudu kohta oli 0,194 ehk 19,4%. Suurim ülemine kvartiilväärtus, mis F1-indeksi hindamisel saavutati, oli 0,24, kasutades parameetrit *k=5* ning fotoosade kattuvust. Väikseim alumine kvartiil, mis F1-indeksi hindamisel saavutati, oli 0,10, kasutades parameetrit *k=4* ja kasutamata fotoosade kattuvust.



Joonis 3. Tehisaruga puisniitude rohurinde fotodelt määratud taimeliikide täpsuse seos fotode lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus on esitatud kujul esineb/ei esine. Parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse; parameeter kattuvus iseloomustas foto lõikamisel fotoosade kattuvuse määra.



Joonis 4. Tehisaruga puisniitude rohurinde fotodelt määratud taimeliikide saagise seos fotode lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus on esitatud kujul esineb/ei esine. Parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse; parameeter kattuvus iseloomustas foto lõikamisel fotoosade kattuvuse määra.



Joonis 5. Tehisaruga puisniitude rohurinde fotodelt määratud taimeliikide F1-indeksi seos fotode lõikamise parameetritega kattuvus (ehk *overlap*) ja *k*, kus kattuvus on esitatud kujul esineb/ei esine. Parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse; parameeter kattuvus iseloomustas foto lõikamisel fotoosade kattuvuse määra.

Lisaks parameetritele hinnati taimeruutude fotode määratavust. Parimad keskmised F1-skoori tulemused saavutati taimeruutude TAC_TAA_A, 15 ja VOR_KINN määramisel, mille kõrgeimad F1-skoorid olid vastavalt 0,8, 0,6 ja 0,5 (Joonis 6, Joonis 7). Taimeruudu TAC_TAA_A kõrgeima F1-skoori korral olid õigesti määratud liigid: **aasosi** (*Equisetum pratense*), **lillakas** (*Rubus saxatilis*), **harilik jalakas** (*Ulmus glabra*) ja **hall lepp** (*Alnus incana*). Valepositiivsetena määrati: **metsosi** (*Equisetum sylvaticum*) ja **künnapuu** (*Ulmus laevis*). Valenegatiivseid liike polnud. Halvimad keskmised F1-tulemused saavutati taimeruutude SIP_TAA_A3 (Joonis 8), SAA_TAA_B1_LISA1 ja 40 (Joonis 9) määramisel, mille madalaimad F1-skoorid olid vastavalt 0, 0 ja 0,07.

Joonis 6. Reitalu andmestikust pärinev taimeruudu foto koodiga TAC_TAA_A, mis tehisaruga määramisel saavutas ülejäänud andmestikuga võrreldes kõrgeima keskmise F1-skoori; kus F1-skoor iseloomustab mudeli võimekust saavutada tasakaal kõrge täpsuse ja kõrge saagise vahel.

Joonis 7. Roosaluste andmestikust pärinev taimeruudu foto koodiga 15, mis tehisaruga määramisel saavutas ülejäänud Roosaluste andmestiku fotodega võrreldes kõrgeima keskmise F1-skoori; kus F1-skoor iseloomustab mudeli võimekust saavutada tasakaal kõrge täpsuse ja kõrge saagise vahel.

Joonis 8. Reitalu andmestikust pärinev taimeruudu foto koodiga SIP_TAA_A3, mis tehisaruga määramisel saavutas ülejäänud andmestikuga võrreldes madalaima keskmise F1-skoori; kus F1-skoor iseloomustab mudeli võimekust saavutada tasakaal kõrge täpsuse ja kõrge saagise vahel.

Joonis 9. Roosaluste andmestikust pärinev taimeruudu foto koodiga 40, mis tehisaruga määramisel saavutas ülejäänud Roosaluste andmestiku fotodega võrreldes madalaima keskmise F1-skoori; kus F1-skoor iseloomustab mudeli võimekust saavutada tasakaal kõrge täpsuse ja kõrge saagise vahel.

Sobivaimate parameetrite *k*=5 ja *overlap*=0,2 korral täpseimalt leitavad liigid, mille tegelik esinemise sagedus oli kümme või rohkem, olid: **harilik härghein** (*Melampyrum nemorosum*), **põldmurakas** (*Rubus caesius*), **harilik naat** (*Aegopodium podagraria*), **angervaks** (*Filipendula ulmaria*), **harilik haab** (*Populus tremula*), **põldohakas** (*Cirsium arvense*), **lillakas** (*Rubus saxatilis*), **vesihaljas tarn** (*Carex flacca*), **harilik maikelluke** (*Convallaria majalis*), **metsmaasikas** (*Fragaria vesca*), **harilik sinihelmikas** (*Molinia caerulea*), **harilik lodjapuu** (*Viburnum opulus*) (Lisa 5). Sobivaimate parameetrite *k*=5 ja *overlap*=0,2 korral madalaima täpsusega määratavad liigid, mille tegelik esinemise sagedus oli kümme või rohkem, olid: **varvastarn** (*Carex ornithopoda*), **tupptarn** (*Carex vaginata*), **harilik mailane** (*Veronica officinalis*), **mägitarn** (*Carex montana*), **suur teeleht** (*Plantago major*), **võsakannike** (*Viola riviniana*), **tuliohakas** (*Cirsium vulgare*), **humal-lutsern** (*Medicago lupulina*), **kortsleht** (*Alchemilla*), **sookask** (*Betula pubescens*), **mitmeõiene tulikas** (*Ranunculus polyanthemos*), **paju** (*Salix*), **hirsstarn** (*Carex panicea*). Täielik nimekiri sagedasemate liikide määramistulemustest, järjestatud F1-indeksi väärtuste järgi parameetrite *k*=5 ja *overlap*=0,2 korral, on toodud „Lisas 5“. Parima mudeli enim levinud täielikult valed määrangud – liigid, mida taimeruutudel tegelikult üldse ei esinenud – olid **harilik käoraamat** (*Gymnadenia conopsea*, n=11), **tihe tarn** (*Carex otrubae*, n=10), **ahtalehine villpea** (*Eriophorum angustifolium*, n=10) ja **kaunis kuldking** (*Cypripedium calceolus* , n=10).

4. ARUTELU

Liikide koosesinemiste uuringutes on masinõppe kasutamine alles algusjärgus. Oma lõputöös uurisin, kas ja kuidas on võimalik kasutada tehisaru abi puisniidu rohurinde 1 m^2 suurusel fotol koosesinevate taimeliikide määramiseks. Uurimuse eesmärk oli saada esmane ülevaade tehisaru potentsiaalist mitme taimeliigi määramisel. Tehisarul põhinevad taimede määramise rakendused on optimeeritud ühe liigi tuvastamisele, mistõttu jagati taimeruut erineval viisil osadeks; iga fotoosa kohta leidis tehisaru võimalikud liigid. Püstitati hüpotees, et fotode osadeks lõikamisel kasutatavad parameetrid mõjutavad automaatse liigimäärangu tulemusi.

Püstitatud hüpotees fotode osadeks lõikamisel kasutatavate parameetrite mõjust automaatse liigimäärangu tulemustele leidis osaliselt kinnitust. Andmeanalüüsist selgus, et ruutmeetriste puisniidu taimeruutude fotode lõikamisel kasutatud parameetril *k* – mitmeks foto pikkupidi ja laiupidi jagatakse – oli oluline mõju määrangute tulemustele. Optimaalseimaks *k* väärtuseks ostus uuringusse kaasatud väärtustest keskmine, ehk *k*=5. Tänu lõikamisele on konkreetsed liigid saadud fotoosadelt selgemini eristatavad. Siinkohal tuleb ka märkida, et tegemist on esialgsete tulemustega ning soovitatav oleks läbi viia komplekssem uurimus ning analüüs. Fotoosade kattuvuse määr ehk *overlap* mõjutas oluliselt tulemuste saagist, aga mitte täpsust ja F1-indeksit. Võime teha esialgse järelduse, et mõningatel lõikamistel jäid olulised taimeosad overlap=0 korral fotode lõikekoha vahele, mis takistas nende liikide tuvastamist. Parim mudel saavutas taimeruudu kohta keskmise F1-skoori 19%. Eesti puisniitudega tehtud määrangud olid võrreldavad PlantCLEF2024 rahvusvahelise võistluse tulemustega, mille raames võiskonnad saavutasid F1-skoori keskmisi tulemusi vahemikus 5% kuni 29%, kusjuures võistlusrühmade keskmine tulemus oli 17% (Goëau et al., 2024). PlantCLEF2024-i parima tulemuse kohta on välja toodud, et ca 10 protsendipunkti 29-st saavutati lisades mudelile võimekus segmeteerimisega välja jätta fotode piirkonnad, kus polnud määratavaid taimi ning arvestades varasemalt sama asukoha kohta tehtud ennustustega. Võib oletada, et ka käesoleva uurimistöö andmeanalüüsi tulemused pareneksid kui kaasataks mudelisse võimekus välja jätta taimeruudu fotode ebaolulised piirkonnad.

Uurimuses kasutatud andmestikud ja taimeruutude fotod olid koostatud varasemalt teiste projektide raames, ning seetõttu ei olnud tegemist standardiseeritud andmetega, näiteks polnud osad fotod tehtud maapinnaga risti ning mitmetel fotodel oli taimeruudu raamidest väljaspoole jäävaid taimi märkimisväärsel määral. Võrreldav uurimus, PlantCLEF2024, toob samuti välja, et nende andmetes esines variatsioone fotode nurgas ja kvaliteedis, kuid fotod olid valitud

rahvusvahelise võistluse jaoks, ning visuaalsel vaatlusel tundub variatsioon PlantCLEF andmestikus märksa väiksem (Goëau et al., 2024). Arvestades saavutatud tulemuste võrreldavust, pole fotografeerimise protokoll niivõrd määrav, samas visuaalse hinnangu järgi on halvemini määratud taimeruudud tehtud suurema nurga alt.

Käesoleva töö parima mudeli (*k*=5, *overlap*=0,2) kohta toodi välja sagedased tehisaru jaoks paremini ja halvemini määratavad taimeliigid (Lisa 5). Määrataivamate ja sagedaste liikide kasuks räägib ilmselt nii üldiselt paremini eristuvad tunnused, kui ka saadavalolevate treeningandmete suurem hulk (J. Pärtel et al., 2021). Halvemini määratavate sagedaste liikide seas oli mitmeid tarnaliike, nagu näiteks hirsstarn, mägitarn, tupptarn, villtarn ja varvastarn. Samuti esines halvemini määratavate sagedaste liikide seas erinevaid lehtpuu/-põõsaliike, nagu paju, sookask, harilik saar, harilik paakspuu, harilik türnpuu ja harilik tamm. Tarnade puhul võime välja tuua, et nende määramisel on sageli roll määramistunustel, mida pildiosalt ei erista (näiteks lehetuped, eri taimeosade värvused), mistõttu võib olla fotolt liigi määramine välistatud. Puuliikide puhul võis madal määratavus tuleneda mudeli tuginemisest peamiselt lehtede kujule, arvestamata võimalikke värvi- ja tekstuurimuutusi, mis võivad tekkida (Kaur & Kaur, 2019).

Käesoleva uurimuse tulemuste põhjal võib soovitada niitude 1 m^2 suuruste taimeruutude automaatsel määramisel kasutada parameetrit *k*=5 ning kaasata ka lõigatud pildiosade kattuvust. Võime oletada, et koosesinevate taimeliikide tuvastamisel Eesti puisniitude taimeruutude fotodelt on potentsiaali kui arendada fotode lõikamise meetodeid, ning komplekssemaid segmenteerimismeetodeid.

4.1 Probleemid ja edasiarendused

Masinõppe mudelite jaoks on fundamentaalselt olulised kaks määramatuse tüüpi, mis mängivad rolli ka taimefotode määramisel (Pl@ntNet, 2021). Esiteks, juhuslikkusest tingitud määramatus (*aleatoric uncertainty*), mille puhul on näiteks taime lõplikuks määramiseks vajalik info fotolt puudu. Võib juhtuda, et määramiseks vajalikku tunnust ei ole fotol või on terve taim teise katnud. Teiseks, puudulikest teadmistest tulenev määramatus (*epistemic uncertainty*), mis siinkohal tuleneb näiteks sellest, et määraja treenimisel on teatud liigi kohta liiga vähe fotosid (Hüllermeier & Waegeman, 2021; Pl@ntNet, 2021; Weinstein, 2018). Ka parim mudel ei saa ennustamisel vältida juhuslikkusest tingitud määramatust, seega tasub

liikide koosesinemiste tuvastamise parendamisel keskenduda puudulikest teadmistest tuleneva määramatuse vähendamisele, ehk koguda paremaid andmeid ja luua paremaid mudeleid (Hüllermeier & Waegeman, 2021).

Tehnilisest vaatepunktist on saadavalolevate andmete madal kvaliteet üheks oluliseimaks probleemiks liikide koosesinemiste tuvastamisel tehisaruga (Lefort et al., 2025). Kvaliteetsed andmed on täielikud, õiged ja neis ei esine puuduvaid väärtusi (Sügis et al., 2024). Ökoloogia andmetes esineb erineval määral vajakajäämisi kõigis kolmes eelmainitud omaduses (König et al., 2019; Smits et al., 2025).

Ökoloogia andmete mittetäielikkus on teatud määral mõistetav – praeguste võimekuste juures on naiivne loota dokumenteerida kõigi maailma liike ja nende esinemisi (Sugai & Llusia, 2019). Peamine probleem andmete mittetäielikkusega seisneb kallutatuses – on kindlaid piirkondi, bioome ja taksoneid, mille esindatus andmebaasides on väga madal (Beck et al., 2012). Näiteks Euroopa ja Põhja-Ameerika on ökoloogilistes andmestikes suhteliselt hästi esindatud, aga mitmekesised regioonid, eriti Aasias ja Aafrikas on halvasti esindatud (Beck et al., 2012). Andmete esindatuse parendamiseks on pakutud muuhulgas välja erinevaid poolautomatiseeritud ja automatiseeritud lahendusi, nagu näiteks sensoritega varustatud mehitamata õhusõidukite kasutamine (Besson et al., 2022; Wüest et al., 2020). Sotsiaal-majanduslikust vaatapunktist on ökoloogiliste andmestike puudujääkide vähendamiseks vajalik parem rahvusvaheline koostööd, vähemarenenud piirkondade kohaliku teaduse toetamine ning nende piirkondade teadlaste kaasamine andmete kogumisel (Armitage et al., 2020; Meyer et al., 2015).

Andmete õigsuse kui ka puuduvate väärtuste vähendamiseks on võimalik teha muudatusi juba andmete kogumise faasis, järgides protokolle taksonoomiliste nimede ühtlustamiseks, standardiseerides metoodikaid, andmete vormingut jne (König et al., 2019; Wieczorek et al., 2012). Üks kasutuses olevatest standarditest on Darwin Core, mida kasutavad näiteks GBIF (GBIF, 2025), PlantNet (Pl@ntNet, 2025b) ja Ocean Biodiversity Information System (EurOBIS, 2025). Juba kogutud andmete annoteeringute õigsuse tagamiseks on võimalik kasutada näiteks harrastusteadlaste abi, võimaldades kasutajatel teiste määranguid kontrollida (Lefort et al., 2025).

Tehisaru mudelite parendamiseks taimeruutude määramise kontekstis on võimalik arendada erinevaid viise, kuidas mudelit rohkem suunata määrama taimi, mitte tausta, kõrvalist objekti vms. Üheks võimalikuks lahenduseks on visuaalsete transformerite kasutuselevõtt, mis

võimaldavad mudelil keskenduda olulistele fotoosadele, parendades seeläbi visuaalselt komplekssete fotode annoteerimist (Goëau et al., 2024). Samuti on võimalik arendada segmenteerimist, eesmärgiga mitme taimeliigiga fotolt liike eraldada, ning neid seejärel ühe kaupa määrata (Kirillov et al., 2023). Treeningandmete vähesusest üle saamiseks on võimalik kasutada isejuhendatud õpet, kus lastakse tehisarul kõigepealt annoteerimata andmetel seoseid luua ning seejärel peenhäälestatakse mudelit vähesel hulgal annoteeritud andmetel (Goëau et al., 2024; Sügis et al., 2024). Sotsio-kultuurilisest vaatepunktist on peamiseks takistuseks paremate masinõppe lahendusteni jõudmisel endiselt nõrk suhtlus masinõppe ja loodusteadlaste kogukondade vahel. Vajalik oleks ökoloogide teadlikkuse suurendamine olemasolevatest masinõppe tööriistadest ja nende kasutamise oskuse arendamine – nii ülikoolides kui ka erialastel koolitustel (Michener & Jones, 2012; Thessen, 2016).

KOKKUVÕTE

Liikide koosesinemiste uurimine on pika ajalooga uurimissuund. Varasemad käsitlused rõhutasid konkurentsi tähtsust koosluste kujunemisel, kuid hilisemad uuringud on toonud esile ka soodustavate ja vahendatud interaktsioonide ning levikupiirangute rolli. Tänapäevased lähenemised kombineerivad nii statistilisi mudeleid kui ka mitmekülgseid andmekogumeid, et mõista, millised tegurid määravad, miks teatud liigid esinevad koos ja teised mitte. Koosesinemiste uurimine aitab paremini mõista ökosüsteemide struktuuri ja toimimist ning võimaldab hinnata ka kooslustest puudu olevat ehk tumedat elurikkust.

Töö eesmärk oli uurida, kas ja kuidas on võimalik kasutada tehisaru taimeliikide koosesinemise tuvastamiseks Eesti puisniitude rohurinde taimeruutude fotodelt. Uurimisküsimusele vastamiseks viidi läbi uurimus, kus hinnati PlantNeti rakenduse võimet määrata liike taimeruutude fotodelt, millel esineb mitmeid liike. Uurimus keskendus fotode lõikamisel kasutavate parameetrite muutmise mõjule määramistulemustele. Selgus, et mitmeks ristkülikuks taimeruudu foto oli lõigatud mõjutas oluliselt määramist – kõige paremad tulemused saavutati lõigates fotot viieks nii risti- kui pikkupidi. Pildiosade osaline kattuvus mõjutas statistiliselt oluliselt saagist, võimaldades tuvastada rohkem andmestikus tegelikult esinevaid liike. Tulemused näitasid, et mõned liigid olid tehisaru jaoks kergemini ära tuntavad, samas kui osad määrati sagedamini valesti.

Tehisarul on potentsiaal toetada liikide koosesinemise uurimist, eriti kui on vaja analüüsida järjest suurenevaid andmehulkasid. Samas on mitmeid kitsaskohti, nagu andmete puudulik kvaliteet, määramatuse eri allikad ning vajadus laiapõhjalisema interdistsiplinaarse koostöö järele. Edasine areng tehisaru kasutamisel liikide koosesinemiste leidmiseks piltidelt võiks liikuda segmenteerimise täiustamise suunas. Kuigi töö raames läbi viidud uurimus oli esmane ja lihtsustatud, viitab see siiski võimalusele kaasata tehisaru liikide koosesinemiste tuvastamisse, pakkudes uut võimalust elurikkuse analüüsimiseks.

SUMMARY

The study of species co-occurrences has a long history. Previous approaches highlighted the importance of competition in the formation of communities, but later studies have emphasised the role of facilitation, indirect interactions, and dispersal limitations. Modern approaches combine statistical models and multifaceted datasets to understand which factors determine why particular species occur together, while others do not. The study of species co-occurrences helps to better understand the structure and functioning of ecosystems, whilst making it possible to evaluate the missing, or dark, diversity.

This thesis aimed to investigate whether and how artificial intelligence can be used to detect the co-occurrences of plant species from photographs of vegetation quadrats of the herbaceous layer of Estonian wooded meadows. A study was conducted to evaluate the ability of the PlantNet application to identify species from photographs of vegetation quadrats in which multiple species occur. The analysis part of this thesis focused on the effect that altering the parameters used for splitting photos had on species identification. It turned out that the amount of cuttings made significantly influenced the species identification results – the best results were achieved when splitting the quadrat photos into five parts, both lengthwise and widthwise. The partial overlap of the split image segments had a statistically significant effect on recall, increasing the number of correctly classified positive samples. The results showed that some species were more easily identifiable for AI, while others were often incorrectly identified.

Artificial intelligence has the potential to further the research of species co-occurrences, especially with the increasing amounts of data that needs to be analyzed. At the same time, there are multiple bottlenecks, such as the lack of quality data, the different sources of uncertainty in machine learning, and the need for broader interdisciplinary collaboration. Further development could move towards improvement in segmentation. Although the research carried out within the framework of this work was primary and simplified, it still supports the possibility of involving AI in the detection of species co-occurrence, thus offering new opportunities for the analysis of biodiversity and communities.

TÄNUAVALDUSED

Soovin tänada oma juhendajat Meelis Pärtlit, kes oli suureks abiks töö kirjutamisel, andes alati põhjalikku ja kasulikku tagasisidet. Suur aitäh Triin Reitalule ja Elle Roosalustele, kes lahkesti jagasid andmestikke, millel töö praktiline osa põhineb, ning vastasid tekkinud küsimustele.

KASUTATUD KIRJANDUS

Armitage, D., Mbatha, P., Muhl, E.-K., Rice, W., & Sowman, M. (2020). Governance principles for community-centered conservation in the post-2020 global biodiversity framework. *Conservation Science and Practice*, *2*. https://doi.org/10.1111/csp2.160

Azeria, E. T., Fortin, D., Hébert, C., Peres-Neto, P., Pothier, D., & Ruel, J.-C. (2009). Using null model analysis of species co-occurrences to deconstruct biodiversity patterns and select indicator species. *Diversity and Distributions*, *15*(6), 958–971. https://doi.org/10.1111/j.1472-4642.2009.00613.x

Bascompte, J. (2009). Disentangling the web of life. *Science*, *325*(5939), 416–419. https://doi.org/10.1126/science.1170749

Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., & Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, *35*(8), 673–683. https://doi.org/10.1111/j.1600-0587.2012.07364.x

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, *25*(12), 2753–2775. https://doi.org/10.1111/ele.14123

Brito, D. (2010). Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology*, *11*(8), 709–713. https://doi.org/10.1016/j.baae.2010.09.007

Bruno, J. F., Stachowicz, J. J., & Bertness, M. D. (2003). Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution*, *18*(3), 119–125. https://doi.org/10.1016/S0169-5347(02)00045-9

Carmona, C. P., & Pärtel, M. (2021). Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. *Global Ecology and Biogeography*, *30*(1), 316–326. https://doi.org/10.1111/geb.13203

Chesson, P. (1994). Multispecies competition in variable environments. *Theoretical Population Biology*, *45*(3), 227–276. https://doi.org/10.1006/tpbi.1994.1013

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology, Evolution, and Systematics*, *31*(Volume 31, 2000), 343–366. https://doi.org/10.1146/annurev.ecolsys.31.1.343

Cole, E., Horn, G. V., Lange, C., Shepard, A., Leary, P., Perona, P., Loarie, S., & Aodha, O. M. (2023). *Spatial Implicit Neural Representations for Global-Scale Species Mapping* (No. arXiv:2306.02564). arXiv. https://doi.org/10.48550/arXiv.2306.02564

Connor, E. F., & Simberloff, D. (1979). The assembly of species communities: Chance or competition? *Ecology*, *60*(6), 1132–1140. https://doi.org/10.2307/1936961

Cowlishaw, G. (1999). Predicting the pattern of decline of African primate diversity: An extinction debt from historical deforestation. *Conservation Biology*, *13*(5), 1183–1193. https://doi.org/10.1046/j.1523-1739.1999.98433.x

Crist, T. O., Veech, J. A., Gering, J. C., Summerville, K. S., & Boecklen, A. E. W. J. (2003). Partitioning species diversity across landscapes and regions: A hierarchical analysis of α, β, and γ diversity. *The American Naturalist*, *162*(6), 734–743. https://doi.org/10.1086/378901

Cristofoli, S., Piqueray, J., Dufrêne, M., Bizoux, J., & Mahy, G. (2010). Colonization credit in restored wet heathlands. *Restoration Ecology*, *18*(5), 645–655. https://doi.org/10.1111/j.1526-100X.2008.00495.x

Daru, B. H. (2025). Tracking hidden dimensions of plant biogeography from herbaria. *New Phytologist*, *246*(1), 61–77. https://doi.org/10.1111/nph.70002

Diamond, J. M. (1975). Assembly of species communities. *Ecology and Evolution of Communities* (lk 342–444). The Belknap Press of Harvard University Press.

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I. C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P. B., Moles, A. T., Dickie, J., Zanne, A. E., Chave, J., … Zotz, G. (2022). The global spectrum of plant form and function: Enhanced species-level trait dataset. *Scientific Data*, *9*(1), 755. https://doi.org/10.1038/s41597-022-01774-9

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, *151*, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Eesti Keele Instituut. (2025, märts 10). *Sõnaveeb*. Termini „tehisintellekt“ tähendused. https://sonaveeb.ee/search/unif/dlall/dsall/tehisintellekt/1/est

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(Volume 40, 2009), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

EurOBIS. (2025). *EurOBIS data formats*. Data and file formats. https://www.eurobis.org/data_formats

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances, challenges, and solutions. *BioScience*, *68*(8), 563–576. https://doi.org/10.1093/biosci/biy068

Folk, R. A., Guralnick, R. P., & LaFrance, R. T. (2024). FloraTraiter: Automated parsing of traits from descriptive biodiversity literature. *Applications in Plant Sciences*, *12*(1), e11563. https://doi.org/10.1002/aps3.11563

GBIF. (2024). *More than 10,000 scientific papers enabled by GBIF-mediated data*. https://www.gbif.org/news/7wQdwQiUN5qF33Fu0CWgHV/more-than-10000-scientific-papers-enabled-by-gbif-mediated-data

GBIF. (2025). *Data standards*. Global Biodiversity Information Facility Data Standards. https://www.gbif.org/standards

Ghosh, S., Das, N., Das, I., & Maulik, U. (2020). Understanding deep learning techniques for image segmentation. *ACM Computing Surveys*, *52*(4), 1–35. https://doi.org/10.1145/3329784

Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, *53*(1), 7–26. https://doi.org/10.2307/2479933

Goëau, H., Espitalier, V., Bonnet, P., & Joly, A. (2024). *Overview of PlantCLEF 2024: Multi-species plant identification in vegetation plot images*. Conference and Labs of the Evaluation Forum. https://www.semanticscholar.org/paper/Overview-of-PlantCLEF-2024%3A-Multi-species-Plant-in-Go%C3%ABau-Espitalier/8094a07863a506a2a615eb611653733ec015a7fd

Google Play. (2025, märts 30). *Plant Identifier Apps on Google Play*. https://play.google.com/store/apps/details?id=com.scaleup.plantid&hl=en_US

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, *81*(9), 2606–2621. https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2

Gotelli, N. J., & McCabe, D. J. (2002). Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, *83*(8), 2091–2096. https://doi.org/10.2307/3072040

Han, B. A., Varshney, K. R., LaDeau, S., Subramaniam, A., Weathers, K. C., & Zwart, J. (2023). A synergistic future for AI and ecology. *Proceedings of the National Academy of Sciences*, *120*(38), e2220283120. https://doi.org/10.1073/pnas.2220283120

He, F., Gaston, K. J., Connor, E. F., & Srivastava, D. S. (2005). The local–regional relationship: Immigration, extinction, and scale. *Ecology*, *86*(2), 360–365. https://doi.org/10.1890/04-1449

HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., & Mayfield, M. M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics*, *43*(Volume 43, 2012), 227–248. https://doi.org/10.1146/annurev-ecolsys-110411-160411

Huang, S., Yoshitake, K., Watabe, S., & Asakawa, S. (2022). Environmental DNA study on aquatic ecosystem monitoring and management: Recent advances and prospects. *Journal of Environmental Management*, *323*, 116310. https://doi.org/10.1016/j.jenvman.2022.116310

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., Firth, J. A., Gaston, K. J., Jepson, P., Kalinkat, G., Ladle, R., Soriano-Redondo, A., Souza, A. T., & Roll, U. (2020). iEcology: Harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*(7), 630–639. https://doi.org/10.1016/j.tree.2020.03.003

Jones, H. G., & Jones, A. J. (2025). Application and pitfalls of the use of plant ID apps for urban flora and citizen science studies. *Plant Ecology & Diversity*, 1–9. https://doi.org/10.1080/17550874.2025.2476938

Kaur, S., & Kaur, P. (2019). Plant species identification based on plant leaf using computer vision and machine learning techniques. *Journal of Multimedia Information System*, *6*(2), 49–60. https://doi.org/10.33851/JMIS.2019.6.2.49

Keil, P. (2019). Z-scores unite pairwise indices of ecological similarity and association for binary data. *Ecosphere*, *10*(11), e02933. https://doi.org/10.1002/ecs2.2933

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). *Segment Anything* (No. arXiv:2304.02643). arXiv. https://doi.org/10.48550/arXiv.2304.02643

Kraan, C., Thrush, S. F., & Dormann, C. F. (2020). Co-occurrence patterns and the large-scale spatial structure of benthic communities in seagrass meadows and bare sand. *BMC Ecology*, *20*(1), 37. https://doi.org/10.1186/s12898-020-00308-4

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., & Kreft, H. (2019). Biodiversity data integration—The significance of data resolution and domain. *PLOS Biology*, *17*(3), e3000183. https://doi.org/10.1371/journal.pbio.3000183

LaDeau, S. L., Han, B. A., Rosi-Marshall, E. J., & Weathers, K. C. (2017). The next decade of big data in ecosystem science. *Ecosystems*, *20*(2), 274–283. https://doi.org/10.1007/s10021-016-0075-y

Lefort, T., Affouard, A., Charlier, B., Lombardo, J.-C., Chouet, M., Goëau, H., Salmon, J., Bonnet, P., & Joly, A. (2025). Cooperative learning of Pl@ntNet's Artificial Intelligence algorithm: How does it work and how can we improve it? *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.14486

Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A., & Tsuboi, M. (2021). Computer Vision, Machine Learning, and the Promise of Phenomics in Ecology and Evolutionary Biology. *Frontiers in Ecology and Evolution*, *9*, 642774. https://doi.org/10.3389/fevo.2021.642774

Macarthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*. https://doi.org/10.1086/282505

MacKenzie, D. I., Bailey, L. L., & Nichols, J. D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, *73*(3), 546–555.

McIntire, E. J. B., & Fajardo, A. (2014). Facilitation as a ubiquitous driver of biodiversity. *New Phytologist*, *201*(2), 403–416. https://doi.org/10.1111/nph.12478

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*(1), 8221. https://doi.org/10.1038/ncomms9221

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, *27*(2), 85–93. https://doi.org/10.1016/j.tree.2011.11.016

Mittelbach, G. G., & McGill, B. J. (2019). *Community ecology* (2. tr). Oxford University Press. https://doi.org/10.1093/oso/9780198835851.001.0001

Münzbergová, Z., & Herben, T. (2004). Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos*, *105*(2), 408–414.

Naaf, T., & Kolk, J. (2015). Colonization credit of post-agricultural forest patches in NE Germany remains 130–230 years after reforestation. *Biological Conservation*, *182*, 155–163. https://doi.org/10.1016/j.biocon.2014.12.002

Orozco-Arias, S., Núñez-Rincón, A. M., Tabares-Soto, R., & López-Álvarez, D. (2019). Worldwide co-occurrence analysis of 17 species of the genus Brachypodium using data mining. *PeerJ*, *6*, e6193. https://doi.org/10.7717/peerj.6193

Ovaskainen, O., Hottola, J., & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, *91*(9), 2514–2521. https://doi.org/10.1890/10-0173.1

Parsons, M. A., Godøy, Ø., LeDrew, E., de Bruin, T. F., Danis, B., Tomlinson, S., & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, *37*(6), 555–569. https://doi.org/10.1177/0165551511412705

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, *33*(10), 1704–1711. https://doi.org/10.1111/j.1365-2699.2006.01460.x

Pichler, M., & Hartig, F. (2021). A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods in Ecology and Evolution*, *12*(11), 2159–2173. https://doi.org/10.1111/2041-210X.13687

Picon, A., San-Emeterio, M. G., Bereciartua-Perez, A., Klukas, C., Eggers, T., & Navarra-Mestre, R. (2022). Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Computers and Electronics in Agriculture*, *194*, 106719. https://doi.org/10.1016/j.compag.2022.106719

Piqueray, J., Cristofoli, S., Bisteau, E., Palm, R., & Mahy, G. (2011). Testing coexistence of extinction debt and colonization credit in fragmented calcareous grasslands with complex historical dynamics. *Landscape Ecology*, *26*(6), 823–836. https://doi.org/10.1007/s10980-011-9611-5

Pl@ntNet. (2021, märts 30). A Pl@ntNet dataset for machine learning researchers. *Pl@ntNet*. https://plantnet.org/en/2021/03/30/a-plntnet-dataset-for-machine-learning-researchers/

Pl@ntNet. (2023, juuli 5). Covering all countries floras & new identification AI. *Pl@ntNet*. https://plantnet.org/en/2023/07/05/covering-all-countries-floras-new-identification-ai/

Pl@ntNet. (2025a, märts 30). *Pl@ntNet home page*. Pl@ntNet. https://plantnet.org/en/

Pl@ntNet. (2025b, märts 30). *Pl@ntNet observations information*. https://ipt.plantnet.org/resource?r=observations

Pärtel, J., Pärtel, M., & Wäldchen, J. (2021). Plant image identification application demonstrates high accuracy in Northern Europe. *AoB PLANTS*, *13*(4), plab050. https://doi.org/10.1093/aobpla/plab050

Pärtel, M., Szava-Kovats, R., & Zobel, M. (2011). Dark diversity: Shedding light on absent species. *Trends in Ecology & Evolution*, *26*(3), 124–128. https://doi.org/10.1016/j.tree.2010.12.004

Pärtel, M., Tamme, R., Carmona, C. P., Riibak, K., Moora, M., Bennett, J. A., Chiarucci, A., Chytrý, M., De Bello, F., Eriksson, O., Harrison, S., Lewis, R. J., Moles, A. T., Öpik, M., Price, J. N., Amputu, V., Askarizadeh, D., Atashgahi, Z., Aubin, I., … Zobel, M. (2025). Global impoverishment of natural vegetation revealed by dark diversity. *Nature*. https://doi.org/10.1038/s41586-025-08814-5

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, *7*(1), 1–15. https://doi.org/10.1046/j.1461-0248.2003.00554.x

Ronk, A., Robert Szava-Kovats, & Meelis Pärtel. (2015). Applying the dark diversity concept to plants at the European scale. *Ecography*, *38*, 1015–1025. https://doi.org/10.1111/ecog.01236

Schlau, B. M., Huxman, T. E., Mooney, K. A., & Pratt, J. D. (2023). Three-way species interactions reverse the positive pairwise effects of two natives on an exotic invader. *Plant Ecology*, *224*(4), 349–359. https://doi.org/10.1007/s11258-023-01304-6

Sethi, S. S., Ewers, R. M., Jones, N. S., Signorelli, A., Picinali, L., & Orme, C. D. L. (2020). SAFE Acoustics: An open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. *Methods in Ecology and Evolution*, *11*(10), 1182–1185. https://doi.org/10.1111/2041-210X.13438

Smith, S. W. (1997). *The scientist and engineer's guide to digital signal processing* (1st ed). California Technical Pub.

Smits, A. P., Hall, E. K., Deemer, B. R., Scordo, F., Barbosa, C. C., Carlson, S. M., Cawley, K., Grossart, H.-P., Kelly, P., Mammola, S., Pintar, M. R., Robbins, C. J., Ruhi, A., & Saccò, M. (2025). Too much and not enough data: Challenges and solutions for generating information in freshwater research and monitoring. *Ecosphere*, *16*(3), e70205. https://doi.org/10.1002/ecs2.70205

Sousa, W. P. (1979). Disturbance in marine intertidal boulder fields: The nonequilibrium maintenance of species diversity. *Ecology*, *60*(6), 1225–1239. https://doi.org/10.2307/1936969

Sugai, L. S. M., & Llusia, D. (2019). Bioacoustic time capsules: Using acoustic monitoring to document biodiversity. *Ecological Indicators*, *99*, 149–152. https://doi.org/10.1016/j.ecolind.2018.12.021

Svenning, J.-C., Normand, S., & Skov, F. (2008). Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. *Ecography*, *31*(3), 316–326. https://doi.org/10.1111/j.0906-7590.2008.05206.x

Sügis, E., Tampuu, A., Aljanaki, A., Fišel, M., & Kull, M. (2024). *Praktiline andmeteadus: Kõrgkooliõpik* (1. tr). Tartu Ülikooli arvutiteaduse instituut.

Zhai, Z.-M., Glaz, B., Haile, M., & Lai, Y.-C. (2024). *Learning to learn ecosystems from limited data—A meta-learning approach* (No. arXiv:2410.07368). arXiv. https://doi.org/10.48550/arXiv.2410.07368

Zobel, M., Otto, R., Laanisto, L., Naranjo-Cigala, A., Pärtel, M., & Fernández-Palacios, J. M. (2011). The formation of species pools: Historical habitat abundance affects current local diversity. *Global Ecology and Biogeography*, *20*(2), 251–259. https://doi.org/10.1111/j.1466-8238.2010.00593.x

Tartu Ülikooli loodusmuuseum. (2022, detsember 5). *Botaanilised kogud Tartu Ülikool*. https://natmuseum.ut.ee/et/botaanilised-kogud

Thessen, A. E. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, *1*. https://doi.org/10.3897/oneeco.1.e8621

Tuffery, S. (2022). *Deep learning: From big data to artificial intelligence with R*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/tartu-ebooks/detail.action?docID=7133414

Tulloch, A. I. T., Chadès, I., & Lindenmayer, D. B. (2018). Species co-occurrence analysis predicts management outcomes for multiple threats. *Nature Ecology & Evolution*, *2*(3), 465–474. https://doi.org/10.1038/s41559-017-0457-3

van den Berg, N. I., Machado, D., Santos, S., Rocha, I., Chacón, J., Harcombe, W., Mitri, S., & Patil, K. R. (2022). Ecological modelling approaches for predicting emergent properties in microbial communities. *Nature Ecology & Evolution*, *6*(7), 855–865. https://doi.org/10.1038/s41559-022-01746-7

Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, *22*(2), 252–260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

Veech, J. A. (2014). The pairwise approach to analysing species co-occurrence. *Journal of Biogeography*, *41*(6), 1029–1035. https://doi.org/10.1111/jbi.12318

Wagaman, A. S., & Dobrow, R. P. (2021). *Probability: With applications and R*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/tartu-ebooks/detail.action?docID=6640833

Weigelt, P., Daniel Kissling, W., Kisel, Y., Fritz, S. A., Karger, D. N., Kessler, M., Lehtonen, S., Svenning, J.-C., & Kreft, H. (2015). Global patterns and drivers of phylogenetic

structure in island floras. *Scientific Reports*, *5*(1), 12213. https://doi.org/10.1038/srep12213

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*(3), 533–545. https://doi.org/10.1111/1365-2656.12780

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLOS ONE*, *7*(1), e29715. https://doi.org/10.1371/journal.pone.0029715

Wilson, J. B., Peet, R. K., Dengler, J., & Pärtel, M. (2012). Plant species richness: The world records. *Journal of Vegetation Science*, *23*(4), 796–802. https://doi.org/10.1111/j.1654-1103.2012.01400.x

Wüest, R., Zimmermann, N., Zurell, D., Alexander, J., Fritz, S., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Székely, E., Thuiller, W., Wikelski, M., & Karger, D. N. (2020). Macroecology in the age of Big Data – Where to go from here? *Journal of Biogeography*, *47*(1), 1–12. https://doi.org/10.1111/jbi.13633

LISAD

Lisa 1. R-i skript fotode lõikamiseks ja PlantNetiga määramiseks

Uurimusliku osa läbi viimiseks koostatud R-i skript, mis võimaldab lõigata taimeruutude fotosid ja saata seejärel fotoosad PlantNetti automaatseks määramiseks. Koodist on välja jäetud jooksutamiseks vajalik privaatne API võti (*key*), mis on tasuta saadaval PlantNet kasutajatele. Analüüsitavad pildifailid peavad asuma alamkataloogis „data“. Tulenevalt uurimuslikus osas kasutatud kahe andmestiku formaatide erinevustest on loodud pildifalide protsessimises üks erand – „data“ juurkaustas asuvad fotodest protsessitakse kõiki, aga „data“ alamkaustades asuvatest failidest töödeldakse ainult neid, mille nimetuses on märksõna „ruut“. Tööjärje salvestamiseks loob skript txt-formaadis faili „analyzed_files“. Päringute tulemused salvestatakse Excel formaadis kataloogi „results“.

```
library(httr)
library(jpeg)
library(future)
library(furrr)
library(openxlsx)

#--------------1.DEFINE---------------------------------------------

API_URL <- "https://my-api.plantnet.org/v2/identify"
key <- # Your API key here
project <- "k-eastern-europe"
lang <- "en"
includeRelatedImages <- FALSE

##---------------UPLOAD and RESPONSE--------------------------------

upload_image <- function(image_array) {
  raw_conn <- rawConnection(raw(0), "wb")
  writeJPEG(image_array, raw_conn)
  image_raw <- rawConnectionValue(raw_conn)
  close(raw_conn)

  temp_file <- tempfile(fileext = ".jpg")
  writeBin(image_raw, temp_file)

  URL <- paste0(API_URL, "/", project, "?lang=", lang, "&include-
related-images=", includeRelatedImages, "&api-key=", key, "&no-
reject=true")

  data <- list("images" = httr::upload_file(temp_file), "organs" =
"leaf")
```

```
  response <- httr::POST(URL, body = data, encode = "multipart")

  unlink(temp_file)

  if (response$status_code == 429) {
    warning("Error: 429 - Too Many Requests. Saving progress and
pausing.")
    pause_and_save()
  } else if (response$status_code == 404) {
    warning("Error: 404 - Not Found. Setting response to an empty
list.")
    return(list())
  } else if (response$status_code != 200) {
    warning("Error: ", response$status_code, " - ", content(response,
"text"))
    return(NULL)
  }

  return(content(response))
}

##------------------SPLIT WITH OVERLAP---------------
splitWithOverlap <- function(vec, k, overlap) {
  seg.length <- length(vec) / k
  starts <- seq(1, length(vec), by = seg.length - overlap *
seg.length)
  ends <- starts + seg.length - 1
  ends[ends > length(vec)] <- length(vec)
  splits <- lapply(1:length(starts), function(i)
vec[starts[i]:ends[i]])
  if (overlap > 0) splits <- splits[1:(length(splits) - 1)]
  splits
}

#----------------------3.IMAGE PROCESSING---------------------------
process_image_file <- function(file.name, k, overlap, progress) {
  image <- readJPEG(file.name)

  dims <- dim(image)
  image_list <- list()

  dx <- splitWithOverlap(1:dims[2], k, overlap)
  dy <- splitWithOverlap(1:dims[1], k, overlap)

  c <- 0
  for (ix in 1:length(dx)) {
    for (iy in 1:length(dy)) {
      c <- c + 1
      if (c > progress) {
        image_part <- image[dy[[iy]], dx[[ix]], ]
        image_list[[c]] <- image_part
      }
    }
```

```
  }

  results <- future_map(image_list, upload_image)

  display_name <- if (dirname(file.name) == "." || dirname(file.name)
== "./data") {
    sub("\\.JPG$", "", basename(file.name))
  } else {
    basename(dirname(file.name))
  }

  return(list(results = results, display_name = display_name, k = k,
overlap = overlap, progress = length(image_list) + progress))
}

#--------------------4.SAVE TO EXCEL--------------------------------
save_results_to_excel <- function(results_list, output_file) {
  # Existing or new dataframe
  if (file.exists(output_file)) {
    existing_data <- read.xlsx(output_file)
  } else {
    existing_data <- data.frame(Code_site = character(), Species =
character(), Score = numeric(), K = numeric(), Overlap = numeric(),
stringsAsFactors = FALSE)
  }

  new_data <- data.frame(Code_site = character(), Species =
character(), Score = numeric(), K = numeric(), Overlap = numeric(),
stringsAsFactors = FALSE)

  for (result_info in results_list) {
    results <- result_info$results
    display_name <- result_info$display_name
    k <- result_info$k
    overlap <- result_info$overlap

    for (result in results) {
      if (is.list(result) && !is.null(result$results)) {
        species_name <-
result$results[[1]]$species$scientificNameWithoutAuthor
        score <- result$results[[1]]$score
        new_data <- rbind(new_data, data.frame(Code_site =
display_name, Species = species_name, Score = score, K = k, Overlap =
overlap, stringsAsFactors = FALSE))
      } else {
        new_data <- rbind(new_data, data.frame(Code_site =
display_name, Species = "N/A", Score = "N/A", K = k, Overlap =
overlap, stringsAsFactors = FALSE))
      }
    }
  }

  combined_data <- rbind(existing_data, new_data)
```

```
  write.xlsx(combined_data, output_file)
}

#------------------5.KEEP TRACK OF PROGRESS TXT---------------------
save_analyzed_files <- function(analyzed_files, file_path) {
  write.table(analyzed_files, file_path, row.names = FALSE, col.names
= FALSE, sep = ",")
}

##--------------------LOAD PROGRESS--------------------------------
load_analyzed_files <- function(file_path) {
  if (file.exists(file_path)) {
    analyzed_files <- read.table(file_path, sep = ",",
stringsAsFactors = FALSE, col.names = c("file_name", "progress"))
    analyzed_files$progress <- as.numeric(analyzed_files$progress)
    return(analyzed_files)
  } else {
    return(data.frame(file_name = character(), progress = integer(),
stringsAsFactors = FALSE))
  }
}

##---------------------PAUSE FOR ERROR-----------------------------
pause_and_save <- function() {
  save_analyzed_files(analyzed_files, analyzed_files_path)
  save_results_to_excel(all_results, "./results/results.xlsx")
  cat("Results saved to ./results/results.xlsx\n")
  stop("Paused by user request")
}

#--------------------------6.DEFINE k and o-------------------------
k <- 5
overlap <- 0
analyzed_files_path <- "./analyzed_files.txt"

#-------------------------7.MAIN!----------------------------------

analyzed_files <- load_analyzed_files(analyzed_files_path)

subfolders <- list.dirs("./data", recursive = FALSE)
image_files_root <- list.files("./data", pattern = "\\.JPG$",
full.names = TRUE)

all_results <- list()

# Process images in the root "data" folder
for (file in image_files_root) {
  if (file %in% analyzed_files$file_name) {
    progress <- analyzed_files$progress[analyzed_files$file_name ==
file]
    progress <- as.numeric(progress)
    if (length(progress) > 1) {
      progress <- progress[1]
```

```
    }
  } else {
    progress <- 0
  }

  plan(multisession)
  result_info <- process_image_file(file, k, overlap, progress)
  all_results <- c(all_results, list(result_info))
  analyzed_files <- rbind(analyzed_files, data.frame(file_name = file,
progress = result_info$progress, stringsAsFactors = FALSE))
  save_analyzed_files(analyzed_files, analyzed_files_path)
}

# Process images in subfolders
for (subfolder in subfolders) {
  image_files <- list.files(subfolder, pattern = "ruut.*\\.jpg$",
full.names = TRUE)

  for (file in image_files) {
    if (file %in% analyzed_files$file_name) {
      progress <- analyzed_files$progress[analyzed_files$file_name ==
file]
      progress <- as.numeric(progress)
      if (length(progress) > 1) {
        progress <- progress[1]
      }
    } else {
      progress <- 0
    }

    plan(multisession)
    result_info <- process_image_file(file, k, overlap, progress)
    all_results <- c(all_results, list(result_info))
    analyzed_files <- rbind(analyzed_files, data.frame(file_name =
file, progress = result_info$progress, stringsAsFactors = FALSE))
    save_analyzed_files(analyzed_files, analyzed_files_path)
  }
}

save_results_to_excel(all_results, "./results/results.xlsx")
cat("Analysis complete\n")
## Analysis complete
```

Lisa 2. R-i skript liigitabelite vormistuse ühtlustamiseks

Analüüsi eeltööks koostatud R-i skript, mida kasutasin, et viia kahe erineva töörühma välitööde raames koostatud määrangute andmestikud samale vormingule.

```
library(tidyverse)
library(readxl)
library(writexl)

# Convert Roosaluste data
data <- read_excel("./results/muuda.xlsx")

# Extract species names (first column) and cover values (remove header
row)
species <- data[-1, 1, drop = TRUE]
cover_data <- data[-1, -1]

# Convert cover values to numeric
cover_data <- cover_data %>%
  mutate(across(everything(), as.numeric))

# Add species names as a new column
cover_data <- cover_data %>%
  mutate(Species = species)

# Pivot to long format
muudetud_df <- cover_data %>%
  pivot_longer(
    cols = -Species,
    names_to = "Code_site",
    values_to = "Cover"
  ) %>%
  filter(!is.na(Cover) & Cover > 0)      # Keep only pos cover values

# Summarize duplicates
muudetud_df <- muudetud_df %>%
  group_by(Code_site, Species) %>%
  summarize(Cover = sum(Cover, na.rm = TRUE), .groups = 'drop')

# Save to Excel
write_xlsx(muudetud_df, "./results/oiged_2.xlsx")
```

Lisa 3. R-i skript liiginimede ühtlustamiseks

Analüüsi eeltööks koostatud R-i skript, mida kasutasin välitööde käigus saadud andmete ja PlantNetiga saadud liiginimede ühtlustamiseks.

```
library(openxlsx)
library(dplyr)
library(tidyr)

#------1.LOAD RESULTS----------------------
# Load the results from the Excel file
results <- read.xlsx("./results/results_2_uuesti.xlsx")

# Filter rows based on Score
results <- results %>% filter(Score >= 0.35)

# Load correct identifications
correct_first <- read.xlsx("./results/oiged_2.xlsx")
correct_second <- read.xlsx("./results/oiged_1.xlsx")

# Function to summarize species
summarize_species <- function(df, species_col = "Species",
summary_col) {
  df %>%
    group_by(across(all_of(species_col))) %>%
    summarise("{summary_col}" := paste(unique(.data[[species_col]]),
collapse = ", "), .groups = 'drop')
}

# Create summarized species lists
results_sum <- summarize_species(results, summary_col =
"Species_results")
master_first <- summarize_species(correct_first, summary_col =
"Species_first")
master_second <- summarize_species(correct_second, summary_col =
"Species_second")

# Combine all species lists
master_list <- results_sum %>%
  full_join(master_first, by = "Species") %>%
  full_join(master_second, by = "Species")

# Filter rows where at least one column is NA
df_filtered <- master_list[!apply(!is.na(master_list), 1, all), ]

#-----LOAD NEW NAMES TO REPLACE SPECIES NAMES-------

new_names <- read.xlsx("./results/ymber_nimetatud.xlsx")
correct_1 <- read.xlsx("./results/oiged_2.xlsx")
correct_2 <- read.xlsx("./results/oiged_1.xlsx")
results_rn <- read.xlsx("./results/results_2_uuesti.xlsx")
```

```r
# Function to replace species names
replace_species_names <- function(df, species_col, lookup_df,
lookup_col) {
  df[[species_col]] <- ifelse(
    df[[species_col]] %in% lookup_df[[lookup_col]],
    lookup_df$Accepted_name[match(df[[species_col]],
lookup_df[[lookup_col]])],
    df[[species_col]]
  )
  return(df)
}

# Apply species name replacement
correct_1 <- replace_species_names(correct_1, "Species", new_names,
"Species_first")
correct_2 <- replace_species_names(correct_2, "Species", new_names,
"Species_second")
results_rn <- replace_species_names(results_rn, "Species", new_names,
"Species")

save(results_rn, file = "results_rn.RData")
save(correct_1, file = "correct_1.RData")
save(correct_2, file = "correct_2.RData")
```

Lisa 4. R-i skript tulemuste analüüsimiseks

R-i skript tulemuste analüüsimiseks ja graafikute joonestamiseks.

```
library(openxlsx)
library(dplyr)
library(ggplot2)
library(nlme)
library(tidyr)
library(purrr)
#------1.LOAD RESULTS----------------------
load("results_rn.RData")
load("correct_1.RData")
load("correct_2.RData")

results_rn <- results_rn %>%
  mutate(
    Code_site = gsub("-", "_", Code_site),
    Overlap = ifelse(SplittingMethod == "Without Overlap", 0,
Overlap)
  )
results_rn <- results_rn[results_rn$Score >= 0.35, ]

#-------------2.KEEP SPECIES by CODE SITE---------
combined_results <- results_rn %>%
  group_by(Code_site, SplittingMethod, K, Overlap) %>%
  summarise(Species = paste(unique(Species), collapse = ", "))
## `summarise()` has grouped output by 'Code_site',
'SplittingMethod', 'K'. You can override using the `.groups`
## argument.
#--------------3.CHECK CORRECT RESULTS-----------
correct_df <- bind_rows(correct_1, correct_2) %>%
  mutate(Code_site = gsub(" ", "", gsub("Ruut", "", gsub("-",
"_", Code_site))))

#--------------------4.COMBINE CORRECT RESULTS-------
combined_correct <- correct_df %>%
  group_by(Code_site) %>%
  summarise(Species = paste(unique(Species), collapse = ", "))

#---------5.COMPARE FUNCTION------
compare_species <- function(combined_results, combined_correct)
{
  combined_results <- combined_results %>%
    mutate(Species = strsplit(as.character(Species), ", "))

  combined_correct <- combined_correct %>%
    mutate(Species = strsplit(as.character(Species), ", "))
```

```
  results_list <- list()

  for (code_site in unique(combined_results$Code_site)) {
    correct_species <- combined_correct %>%
      filter(Code_site == code_site) %>%
      pull(Species) %>%
      unlist()

    code_site_methods <- combined_results %>%
      filter(Code_site == code_site)

    for (i in 1:nrow(code_site_methods)) {
      species_results <- code_site_methods$Species[[i]]

      common_species <- intersect(species_results,
correct_species)
      missed_species <- setdiff(correct_species,
species_results)
      wrong_species <- setdiff(species_results, correct_species)

      results_list <- append(results_list, list(
        data.frame(
          Code_site = code_site,
          SplittingMethod =
code_site_methods$SplittingMethod[i],
          K = code_site_methods$K[i],
          Overlap = code_site_methods$Overlap[i],
          Correct = length(common_species),
          Missed = length(missed_species),
          Wrong = length(wrong_species),
          Correct_Species = paste(common_species, collapse = ",
"),
          Missed_Species = paste(missed_species, collapse = ",
"),
          Wrong_Species = paste(wrong_species, collapse = ", ")
        )
      ))
    }
  }

  do.call(rbind, results_list)
}

#-------------------------6. CALL----------------------------
summary_df_return <- compare_species(combined_results,
combined_correct)

###### Meelis 12.03.2025
summary_df_return <- summary_df_return %>%
```

```
  mutate(
    Precision = Correct / (Correct+ Wrong) ,
    Recall = Correct / (Correct + Missed),
    F1_score = (2 * Precision * Recall) / (Precision + Recall))

table(summary_df_return[,c("Overlap","K")])
##        K
## Overlap  4  5  6
##     0   84 84 83
##     0.2 84 84 83
summary_df_return <- summary_df_return %>%
  mutate(F1_score = ifelse(is.nan(F1_score), 0, F1_score))

summary_df_return$Overlap_yes=summary_df_return$Overlap>0
summary_df_return$K_factor=as.factor(summary_df_return$K)
summary_df_return$dataset=ifelse(is.na(as.numeric(summary_df_ret
urn$Code_site)),"Reitalu","Roosaluste")
## Warning in
ifelse(is.na(as.numeric(summary_df_return$Code_site)),
"Reitalu", : NAs introduced by coercion
tab3D=table(summary_df_return$Overlap_yes,summary_df_return$K_fa
ctor,summary_df_return$Code_site)
full.index=apply(tab3D,3,sum)==9

#---7. ANALYSIS-------

##----F1 ANOVA------
anova_result_f <- lme(F1_score ~ Overlap_yes + K_factor +
dataset, data = summary_df_return, random = ~1|Code_site)
anova(anova_result_f)
summary(anova_result_f)
#hist(residuals(anova_result_f))
#summary(glht(anova_result_f, linfct=mcp(K_factor="Tukey")))
par(mar=c(5,12,2,1))

# Reorder the interaction variable by dataset, Overlap_yes, and
K_factor
summary_df_return <- summary_df_return %>%
  mutate(interaction_var = interaction(Overlap_yes, K_factor))
%>%
  mutate(interaction_var = factor(interaction_var,
                                  levels =
unique(interaction_var[order( K_factor,Overlap_yes)])))
##----F1 plot-----
ggplot(summary_df_return, aes(x = interaction_var, y =
F1_score)) +
  geom_boxplot(aes(fill = K_factor)) +
  coord_flip() +
  labs(x = "Töötlus (overlap_yes, K)", y = "F1-skoor") +
```

```
  theme(axis.text.y = element_text(size = 12),
        plot.margin = margin(5, 12, 2, 1, "pt"))
##-----Precision ANOVA-----
anova_result_p <- lme(Precision ~ Overlap_yes + K_factor +
dataset, data = summary_df_return, random = ~1|Code_site)
anova(anova_result_p)
#hist(residuals(anova_result_p))

##----Precision plot-----
ggplot(summary_df_return, aes(x = interaction_var, y =
Precision)) +
  geom_boxplot(aes(fill = K_factor)) +
  coord_flip() +
  labs(x = "Töötlus (overlap_yes, K)", y = "Täpsus") +
  theme(axis.text.y = element_text(size = 12),
        plot.margin = margin(5, 12, 2, 1, "pt"))
##-----Recall ANOVA------
anova_result_r <- lme(Recall ~ dataset+Overlap_yes + K_factor,
data = summary_df_return, random = ~1|Code_site)
anova(anova_result_r)
#hist(residuals(anova_result_r))
summary(anova_result_r)
##------Recall plot-----
ggplot(summary_df_return, aes(x = interaction_var, y = Recall))
+
  geom_boxplot(aes(fill = K_factor)) +
  coord_flip() +
  labs(x = "Töötlus (overlap_yes, K)", y = "Saagis") +
  theme(axis.text.y = element_text(size = 12),
        plot.margin = margin(5, 12, 2, 1, "pt"))
#

# F1 Score Summary (Q1, Median, Q3)
summary_df_return %>%
  group_by(interaction_var) %>%
  summarise(
    Q1_F1 = quantile(F1_score, 0.25, na.rm = TRUE),
    Median_F1 = quantile(F1_score, 0.50, na.rm = TRUE),
    Q3_F1 = quantile(F1_score, 0.75, na.rm = TRUE)
  ) %>%

# Recall Summary (Q1, Median, Q3)
summary_df_return %>%
  group_by(interaction_var) %>%
  summarise(
    Q1_R = quantile(Recall, 0.25, na.rm = TRUE),
    Median_R = quantile(Recall, 0.50, na.rm = TRUE),
    Q3_R = quantile(Recall, 0.75, na.rm = TRUE)
```

```
  ) %>%
  arrange(desc(Median_R))
# Precision Summary (Q1, Median, Q3)
summary_df_return %>%
  group_by(interaction_var) %>%
  summarise(
    Q1_P = quantile(Precision, 0.25, na.rm = TRUE),
    Median_P = quantile(Precision, 0.50, na.rm = TRUE),
    Q3_P = quantile(Precision, 0.75, na.rm = TRUE)
  ) %>%
  arrange(desc(Median_P))
#-----8. EXCEL OUTPUT----

anova_df <- rbind(as.data.frame(anova(anova_result_r)),
                  as.data.frame(anova(anova_result_p)),
                  as.data.frame(anova(anova_result_f)))
write.xlsx(anova_df, file = "anova_results_all.xlsx")

#------9. Calculate------
# Choose best model parameters
summary_df_return <- summary_df_return %>%
  filter(Overlap_yes == TRUE, K_factor == 5)
# Count occurrences with whitespace removal for each category
# Correct species
correct_counts <- summary_df_return %>%
  mutate(Correct_Species =
strsplit(trimws(as.character(Correct_Species)), ",")) %>%
  unnest_longer(Correct_Species) %>%
  mutate(Correct_Species = trimws(Correct_Species)) %>%
  distinct() %>%
  count(Correct_Species, sort = TRUE) %>%
  rename(Species = Correct_Species, Correct_Count = n)

# Missed species
missed_counts <- summary_df_return %>%
  mutate(Missed_Species =
strsplit(trimws(as.character(Missed_Species)), ",")) %>%
  unnest_longer(Missed_Species) %>%
  mutate(Missed_Species = trimws(Missed_Species)) %>%
  distinct() %>%
  count(Missed_Species, sort = TRUE) %>%
  rename(Species = Missed_Species, Missed_Count = n)

# Wrong species
wrong_counts <- summary_df_return %>%
  mutate(Wrong_Species =
strsplit(trimws(as.character(Wrong_Species)), ",")) %>%
  unnest_longer(Wrong_Species) %>%
  mutate(Wrong_Species = trimws(Wrong_Species)) %>%
```

```
  distinct() %>%
  count(Wrong_Species, sort = TRUE) %>%
  rename(Species = Wrong_Species, Wrong_Count = n)

# Merge all counts by Species
merged_counts <- reduce(list(correct_counts, missed_counts,
wrong_counts), full_join, by = "Species")

# Replace NA values with 0
merged_counts[is.na(merged_counts)] <- 0

# Calculate data for classification classes
merged_counts <- merged_counts %>%
  mutate(
    Class_precision = Correct_Count /
(Correct_Count+Wrong_Count),
    Class_recall = Correct_Count/(Correct_Count+Missed_Count),
    Class_F = (2 * Class_precision* Class_recall) /
(Class_precision + Class_recall),
    Support = Correct_Count + Missed_Count

  )


# Final result as a data frame
merged_counts <- as.data.frame(merged_counts)


#----Common species------

species_10plus <- merged_counts %>%
  mutate(Support = Correct_Count + Missed_Count) %>%
  filter(Support >= 10) %>%
  select(Species, Correct_Count,
Missed_Count,Wrong_Count,Class_F)

write.xlsx(species_10plus, file = "species_10.xlsx", rowNames =
FALSE)
```

Lisa 5. Ülevaade levinumate taimeliikide määramistulemustest.

Tabelis on välja toodud taimeliikide, mille tegelik esinemise sagedus oli kümme või rohkem, määramistulemused sobivaimate parameetrite *k*=5 ja *overlap*=0,2 korral. Parameeter *k* iseloomustas, mitmeks foto pikkupidi ja laiupidi jagatakse; *overlap* iseloomustas foto lõikamisel kattuvuse määra.Iga liigi kohta on toodud, mitmel korral see liik õigesti määrati ja mitmel korral liik määramata jäi. Tabel on järjestatud F1-indeksi väärtuste alusel.

Ladinakeelne liiginimi	**Eestikeelne liiginimi**	**Määratud õigesti**	**Määramata**	**F1-indeks**
Melampyrum nemorosum	harilik härghein	32	9	0,9
Rubus caesius	põldmurakas	18	13	0,7
Aegopodium podagraria	harilik naat	16	20	0,6
Filipendula ulmaria	angervaks	9	12	0,5
Populus tremula	harilik haab	14	25	0,5
Cirsium arvense	põldohakas	7	14	0,5
Rubus saxatilis	lillakas	16	29	0,5
Carex flacca	vesihaljas tarn	12	28	0,4
Convallaria majalis	harilik maikelluke	16	48	0,4
Fragaria vesca	metsmaasikas	10	32	0,4
Molinia caerulea	harilik sinihelmikas	5	12	0,4
Viburnum opulus	harilik lodjapuu	4	14	0,4
Corylus avellana	harilik sarapuu	4	13	0,3

Brachypodium pinnatum	sulg-aruluste	9	27	0,3
Galium boreale	värvmadar	3	11	0,3
Scorzonera humilis	madal mustjuur	2	10	0,3
Prunella vulgaris	harilik käbihein	4	16	0,3
Paris quadrifolia	harilik ussilakk	5	24	0,3
Geum rivale	ojamõõl	6	27	0,3
Primula veris	nurmenukk	5	30	0,2
Serratula tinctoria	värvi-paskhein	2	11	0,2
Sonchus arvensis	põld-piimohakas	2	13	0,2
Anemonoides nemorosa	võsaülane	2	15	0,2
Angelica sylvestris	harilik heinputk	3	17	0,2
Acer platanoides	harilik vaher	2	16	0,2
Hepatica nobilis	harilik sinilill	6	50	0,2
Brachypodium sylvaticum	mets-aruluste	3	17	0,2
Maianthemum bifolium	leseleht	1	10	0,2
Melica nutans	longus helmikas	3	33	0,2
Cornus sanguinea	verev kontpuu	2	23	0,1
Taraxacum	võililll	2	30	0,1

Frangula alnus	harilik paakspuu	1	14	0,1
Rhamnus cathartica	harilik türnpuu	1	16	0,1
Dactylis glomerata	harilik kerahein	1	12	0,1
Quercus robur	harilik tamm	1	17	0,1
Viola mirabilis	imekannike	2	40	0,1
Lathyrus vernus	kevadine seahernes	1	19	0,1
Crepis paludosa	soo-koeratubkas	1	15	0,1
Potentilla erecta	tedremaran	1	22	0,1
Carex tomentosa	villtarn	1	39	0,0
Fraxinus excelsior	harilik saar	1	42	0,0
Veronica chamaedrys	külmamailane	0	24	0
Carex ornithopoda	varvastarn	0	20	0
Carex vaginata	tupptarn	0	20	0
Veronica officinalis	harilik mailane	0	20	0
Carex montana	mägitarn	0	18	0
Plantago major	suur teeleht	0	18	0
Viola riviniana	võsakannike	0	18	0
Cirsium vulgare	tuliohakas	0	16	0
Medicago lupulina	humal-lutsern	0	15	0

Alchemilla	kortsleht	0	13	0
Betula pubescens	sookask	0	13	0
Ranunculus polyanthemos	mitmeõiene tulikas	0	11	0
Salix	paju	0	11	0
Carex panicea	hirsstarn	0	10	0

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Marta Miia Pärnpuu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Liikide koosesinemise leidmine tehisaru abil“, mille juhendaja on prof Meelis Pärtel, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi 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.

Marta Miia Pärnpuu

18.05.2025