

Puulajien tunnistus harvasta laserkeilausaineistosta

Tietojenkäsittelytiede
Tietotekniikan laitos, Teknillinen tiedekunta
Kandidaatintutkielma

Laatija:
Andrei Belogurov

Marraskuu 2025

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu
Turnitin OriginalityCheck -järjestelmällä.

Kandidaatintutkielma
Tietotekniikan laitos, Teknillinen tiedekunta
Turun yliopisto

Tutkinto-ohjelma: Tietojenkäsittelytiede

Tekijä: Andrei Belogurov

Otsikko: Puulajien tunnistus harvasta laserkeilausaineistosta

Sivumäärä: 23 sivua

Päivämäärä: Marraskuu 2025

Tieto metsän puulajeista on tärkeä metsänhoidon suunnittelussa ja metsän biodiversiteetin arvioinnissa. Nykyään puulajien selvittämiseen käytetään yhä useammin kaukokartoitusmenetelmiä. Tämä kirjallisuuskatsaus keskittyy erityisesti laserkeilausaineiston hyödyntämiseen harvan pistetiheyden aineistoissa. Tällaisia aineistoja ovat esimerkiksi laajat kansalliset laserkeilausaineistot, joita maanmittauslaitos tarjoaa.

Tutkielmassa esitellään koneoppimismenetelmiä, joita käytetään puulajien luokitteluun. Tutkielmassa kuvataan mallien kouluttamisvaihe ja esitellään yksittäisistä puista laskettavia piirteitä, jotka kuvaavat puun ominaisuuksia. Lisäksi tutustutaan spektrikuvien käyttöön laserkeilausaineiston tukena, aineiston esikäsittelyyn ennen neuroverkoille syöttämistä ja puiden erotteluun laserkeilausaineistosta.

Tulokset osoittavat, että harvallakin laserkeilausaineistolla voidaan luokitella puita hyvin. Puulajien luokitteluun käytetään erilaisia koneoppimismenetelmiä, kuten lineaarista erotteluanalyysiä, satunnaismetsää, tukivektorikonetta ja syväoppivia neuroverkkoja. Luokittelun tarkkuuteen vaikuttaa ainakin luokiteltavat puulajit, tutkimusalue, mittauksen vuodenaika, laserkeilaimen asetukset, koulutusaineiston koko ja mallin testaustapa. Tämän vuoksi mallien yleistettävyyks on rajallinen.

Asiasanat: laserkeilaus, geoinformatiikka, puulajien luokittelu, satunnaismetsä, tukivektorikone, neuroverkot

Sisällysluettelo

1	Johdanto	1
1.1	Tutkimuksen rajaus	1
1.2	Tutkimuskysymykset	2
1.3	Tutkimusmenetelmät ja lähteet	2
2	Laserkeilausaineiston kuvaus	3
3	Luokittelualgoritmeja	5
3.1	Yksittäisten puiden erottelu	6
3.2	Piirteiden laskeminen	8
3.3	Laserkeilausaineiston ja spektrikuvien yhdistäminen	10
3.4	Lineaarinen erotteluanalyysi	11
3.5	Satunnaismetsä	12
3.6	Tukivektorikone	13
3.7	Neuroverkot	14
4	Tutkimusten tuloksia	17
4.1	Luokittelutarkkuudet ja niihin vaikuttavat tekijät	17
4.2	Pistetiheyden vaikutus tarkkuuteen	19
5	Pohdinta	21
	Yhteenveto	23
	Lähteet	24

1 Johdanto

Metsän monipuolinen puulajisto tukee metsän monimuotoisuutta ja tuhonkestävyyttä (Ruotsalainen ym., 2022). Puulajit sietävät ilmastonmuutoksen vaikutuksia eri tavalla. Esimerkiksi kuusi kärsii ilmastonmuutoksesta, sillä se on altis kuivuuden lisääntymiselle ja sen tuholainen juurikäpö hyötyy ilmastonmuutoksesta. Ruotsin metsänhoidon FSC-sertifikaatti vaatii, että vähintään 10 % metsän puista on lehtipuita (FSC, 2020). Lisäksi puulaji on hyödyllinen parametri puun muiden ominaisuuksien selvittämisessä. Esimerkiksi puun paksuuden suhde puun pituuteen vaihtelee lajin mukaan (Sharma ja Parton, 2007).

Perinteisesti metsän puulajit selvitetään maastomittauksilla. Tietyn alueen puulajit mitataan ja tulos yleistetään koko metsää koskeväksi (Hyypä ym., 2008). Menetelmä vie aikaa ja rahaa, sillä mittaajien tarvitsee matkustaa alueelle ja mitata kukin puu yksi kerrallaan. Nykyisin manuaalista mitaamista voidaan automatisoida paikkatiedon avulla. Automatisointiin voidaan käyttää erilaisia kaukokartoitusmenetelmiä, kuten ilmakuvia, spektrikuvia, laserkeilausta tai niiden yhdistelmää. Tämä tutkielma keskittyy laserkeilaukseen, joka luo metsästä kolmiulotteisen mallin.

Laserkeilauksen luoma kolmiulotteinen malli koostuu joukosta pisteitä, joka luodaan laserin avulla. Pistetiheydellä tarkoitetaan pisteiden määrää neliometriä kohden. Ympäristöä voidaan mitata joko maassa olevalla laserkeilaimella (engl. Terrestrial laser scanning, TLS) tai ilmasta käsin (engl. Airborne laser scanning, ALS). (Petrie ja Toth, 2008)

1.1 Tutkimuksen rajaus

Tämän tutkielman tavoite on selvittää, miten laserkeilaamalla saatua aineistoa voi käyttää puulajien tunnistukseen. Tutkielmassa keskitytään erityisesti lentokoneesta mitattuun diskreettiin (ks. Luku 2) aineistoon, jossa pistetiheys on pieni. Pienellä pistetiheydellä tarkoitetaan pistetiheyttä, joka on alle 20 pistettä neliometriä kohden. Pistetiheys vaikuttaa laserkeilauksen hintaan merkittävästi (Jakubowski ym., 2013). Pienen pistetiheyden aineistoja on tarjolla enemmän. Esimerkiksi maanmittauslaitos tarjoaa maanlaajuista aineistoa, jossa on noin viisi pistettä neliometriä kohden (Maanmittauslaitos, 2023). Tutkielmassa on tarkoitus selvittää, miten näitä aineistoja voi käyttää havumetsävyöhykkeen metsiin. Havumetsissä yleisimmät puulajit ovat kuusi, mänty ja koivu.

1.2 Tutkimuskysymykset

Tutkielmassa keskitytään käytännönläheisiin menetelmiin ja pistetiheyden merkitykseen.

Tutkimuskysymykset ovat seuraavat:

TK1: Mitä algoritmeja voi käyttää puulajin luokitteluun laserkeilausaineistoa käytettäessä?

TK2: Miten pistetiheys vaikuttaa luokittelutarkkuuteen?

1.3 Tutkimusmenetelmät ja lähteet

Tutkielma on kirjallisuuskatsaus. Aineisto on haettu Google Scholar- ja Web of Science -tietokannoista, koska niistä löytyi eniten hyödyllisiä artikkeleita. Tutkielman tekeminen aloitettiin käyttämällä hakulauseita:

1. tree "species classification" lidar
2. tree "species classification" lidar ("low-density" OR "low density") -tropical

Tutkielmaa varten etsittiin siis yleisiä artikkeleita puulajien tunnistuksesta ja artikkeleita, jotka käsittelevät erityisesti pienen tiheyden aineistoja. Osoittautui kuitenkin, että termin "low-density" merkitys on muuttunut tekniikan kehittyessä. Vanhoissa tutkimuksissa käytetään yleensä laserkeilausaineistoja, joiden pistetiheys on pienempi kuin uusissa tutkimuksissa.

Suurin osa tutkielman lähteistä on löydetty hakusanoilla löydettyjen artikkelien

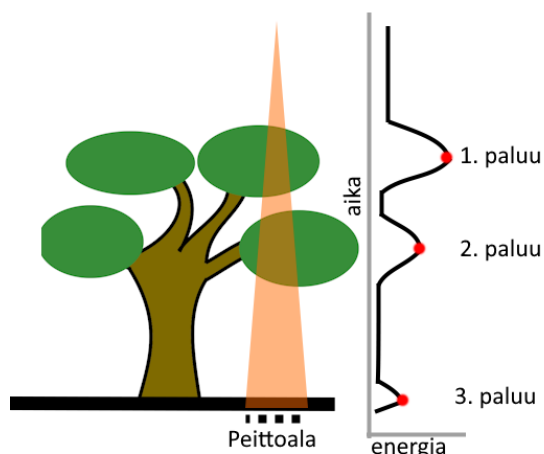
lähdeluetteloista, koska se todettiin tehokkaimmaksi tavaksi etsiä sopivaa aineistoa.

Artikkeliin valintaan vaikutti myös julkaisijan JUFO-luokitus. Suurin osa artikkeleista on vähintään JUFO-tason 1 julkaisijoilta.

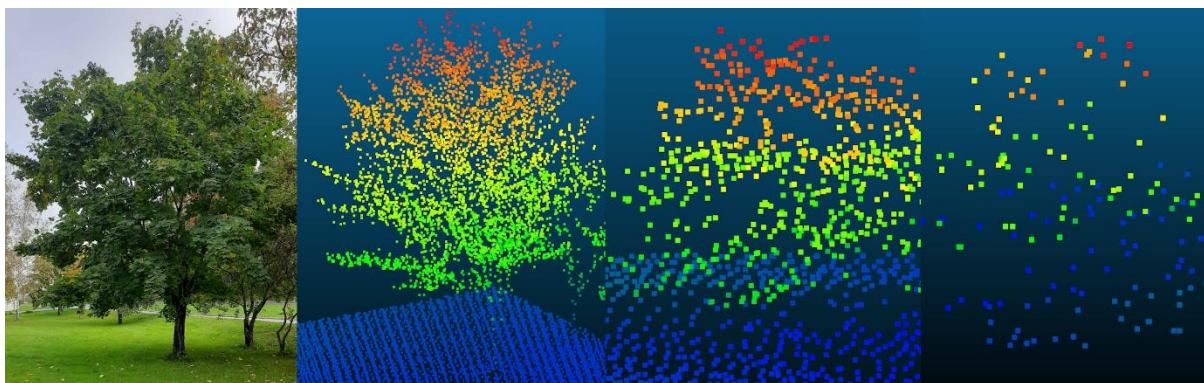
2 Laserkeilausaineiston kuvaus

Laserpulsseihin perustuvassa laserkeilauksessa kolmiulotteinen malli luodaan lähettämällä tuhansia laserpulsseja eri suuntiin. Pieni osa laserpulssien energiasta heijastuu takaisin keilaimelle. Keilaimessa on tarkka kello, joka mittaa ajan, joka laserpulssilla kuluu keilaimelle palaamiseen. Kuluneen ajan avulla voidaan laskea etäisyys kohteeseen, koska valonnopeus tunnetaan tarkasti. Näin muodostuu kolmiulotteisia pisteitä, joista lasersäde heijastui takaisin mittarille. Pisteistä käytetään myös nimeä paluukaiku. Pisteiden joukkoa kutsutaan pistepilveksi. Pistepilven pistetiheydellä tarkoitetaan pisteiden määrää neliometriä kohden. (Petrie ja Toth, 2008)

Laserkeilausaineistot voidaan jakaa kahteen luokkaan heijastuneen energian tallennustavan mukaan (Harding, 2008). Ensimmäisessä tavassa tallennetaan koko palautuneen energian aaltomuoto. Tätä menetelmää kutsutaan täyden aallonmuodon laserkeilaukseksi (engl. full waveform LiDAR, FWL). Toisessa tavassa tallennetaan vain aaltomuodon huiput, jolloin tieto vie vähemmän tallennustilaa, mutta informaatiota menetetään. Menetelmästä käytetään nimeä diskreetti laserkeilaus. Sen tulos on joukko pisteitä, jota kutsutaan pistepilveksi. Pistepilvien käsittely on yksinkertaisempaa kuin kokonaisten aaltojen, joten suurin osa laajoista aineistoista on diskreettiä. Tämä tutkielma keskittyy diskreetin datan hyödyntämiseen. Kuva 1 havainnollistaa diskreetin ja täyden aallonmuodon laserkeilauksen eroa. Kuvassa 2 on esimerkki vaahterapuun pistepilvestä eri pistetiheyksillä.



Kuva 1. Täyden aallonmuodon laserkeilauksessa tallennetaan koko kuvaaja, mutta diskreetissä vain punaisella merkatut huiput. Laserpulssi heijastuu ensin kaksi kertaa puusta ja sitten maasta.



Kuva 2. Sama vaahterapuu 30, 5 ja 0,5 pistetiheyksillä. Kolmenkymmenen pisteen aineisto saatiin Turun kaupungilta (Turun kaupunki, 2021). Viiden pisteen aineisto on maanmittauslaitokselta (Maanmittauslaitos, 2023). 0,5 pisteen aineisto tuotettiin poistamalla satunnaisesti 90 % pisteistä viiden pisteen aineistosta.

Jokainen pistepilven piste sisältää muun muassa koordinaatin ja intensiteetin. Intensiteetti kuvaa sitä, kuinka suuri osa pulssin energiasta palautui sensorille (Harding, 2008). Hyvin heijastavat pinnat tuottavat suuremman intensiteetin kuin huonosti heijastavat. Lisäksi pisteet voivat sisältää luokittelun, joka kertoo, minkälaisesta pisteestä on kyse. Luokittelu voi jakaa pisteet esimerkiksi rakennuksiin, kasvillisuuteen ja maaperään (Graham, 2008).

Yksi laserpulsssi voi tuottaa monta pistettä (paluuta), jos laserpulsssi menee osittain läpi esteestä, kuten puun latvustosta. Pisteisiin tallennetaan tietoa siitä, kuinka mones saman laserpulsssin paluu on kyseessä ja kuinka monta paluuta laserpulsssi tuotti yhteensä (Graham, 2008). Ensimmäiset paluut ovat yleensä korkeammalla kuin muut paluut. Esimerkiksi puun latvassa oleva piste on todennäköisesti ensimmäinen paluu ja maassa oleva paluu on viimeinen (Hyyppä ym., 2008). Ainoat paluut ovat sellaisia ensimmäisiä paluita, joiden jälkeen ei enää tule muita paluita.

Pisteiden määrään ja laatuun vaikuttaa pulssin peittoalue (engl. footprint), joka tarkoittaa pulssin kokoa maanpinnalla (Petrie ja Toth, 2008). Peittoalueeseen vaikuttaa lentokorkeus ja laserin hajaantuminen (engl. beam divergence). Laaja peittoalue huonontaa koordinaattien tarkkuutta, koska keilain ei voi tietää, mistä peittoalueen kohdasta heijastus tapahtui. Liian iso peittoalue aiheuttaa myös paljon ainoita paluita, koska energia jakautuu isommalle alueelle, jolloin on vaikeampi erottaa signaalia kohinasta (Goodwin ym., 2006). Peittoalue voidaan ilmoittaa myös sen säteen avulla. Monet tutkimukset käyttävät alle 0,3 m:n peittoalueita, kuten 0,25 m (Ørka ym., 2013; Torabzadeh ym., 2019), 0,15 m (Li ym., 2013) ja 0,14–0,28 m (Yu ym., 2017).

3 Luokittelualgoritmeja

Puulajien luokitteluun käytetyt algoritmit ovat kehittyneet sitä mukaa, kun uusia koneoppimisen menetelmiä on kehitetty. Viime vuosina on tehty paljon tutkimuksia, joissa käytetään syväoppivia neuroverkkoja (Mäyrä ym., 2021; Wang ym., 2024). Ennen neuroverkkojen laajaa käyttöönottoa tutkimuksissa käytettiin muita koneoppimisen malleja, kuten lineaarista erotteluanalyysia, satunnaismetsiä ja tukivektorikoneita (Michałowska & Rapiński, 2021). Koneoppimisen luokittelutehtävissä käytetään lähes aina ohjattua oppimista, jossa malli koulutetaan koulutusaineistolla. Koulutusaineistossa on sekä puuta kuvaava informaatio että todellinen puulaji. Oikea puulaji selvitetään esimerkiksi maastomittauksilla. Kouluttamisen jälkeen malli pystyy luokittelemaan puita, joita se ei ole aiemmin nähnyt.

Malli koulutetaan seuraavalla tavalla useimmissa tutkimuksissa:

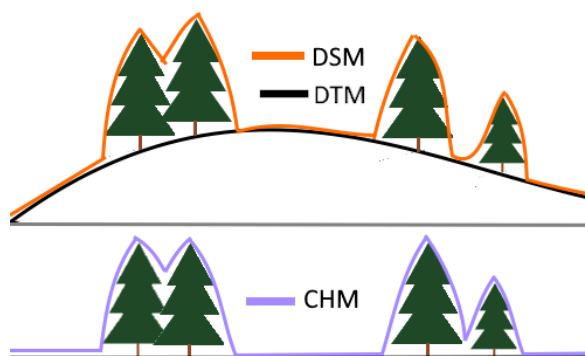
1. Määritetään tutkimusalue ja valitaan sen sisältä näytealueita, jotka edustavat koko aluetta.
2. Matkustetaan paikalle ja määritetään puiden sijainnit ja puulajit näytealueilta. Näin syntyy vertailuaineisto.
3. Laserkeilausaineistosta tunnistetaan yksittäiset puut ja yhdistetään ne vertailuaineiston puihin.
4. Malli koulutetaan koulutusaineistolla, jossa puihin on liitetty niiden todellinen laji.
5. Mallia testataan puilla, jotka eivät ole osa koulutusaineistoa.

Tämän luvun aliluvuissa käsitellään ensin sitä, miten yksittäiset puut erotellaan pistepilvestä, jotta niitä voidaan luokitella. Sen jälkeen esitellään piirteitä, joita puista voi laskea. Piirteet kuvaavat puun ominaisuuksia numeerisessa muodossa. Lopuksi esitellään neljä algoritmia: lineaarinen erotteluanalyysi, satunnaismetsä, tukivektorikone ja neuroverkot. Algoritmit esitellään tässä tutkielmassa hyvin lyhyesti. Tutkielma keskittyy tarkastelemaan, minkälaista tietoa malleille syötetään.

3.1 Yksittäisten puiden erottelu

Ennen puiden luokittelua täytyy tietää, mitkä pisteet kuuluvat yksittäiseen puuhun. Tähän käytetään algoritmeja, jotka tunnistavat puun latvuston rajat. Algoritmien toiminnan ymmärtämiseksi on kuitenkin tarpeellista tutustua siihen, miten latvuskorkeusmalli lasketaan.

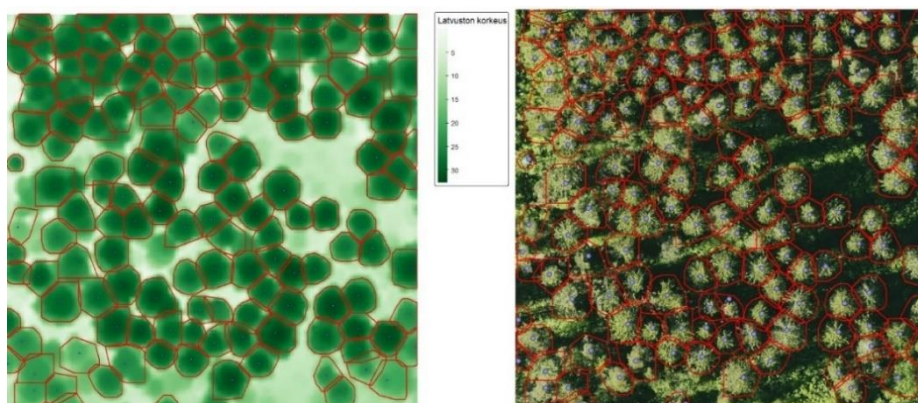
Latvuskorkeusmalli (engl. canopy height model, CHM) on puiden korkeutta kuvaava rasteriaineisto. Latvuskorkeusmalli saadaan vähentämällä korkeusmallista (engl. digital surface model, DSM) maastomalli (engl. digital terrain model, DTM) (Mielcarek ym., 2018). Korkeusmalli ja maastomalli ovat rasteriaineistoja. Korkeusmalli ja maastomalli voidaan luoda monella eri tavalla, mutta yksinkertaisimmassa tapauksessa korkeusmallin luomiseen käytetään ensimmäisiä paluusignaaleja ja maastomallin luomiseen käytetään viimeisiä paluusignaaleja. Ensimmäiset paluusignaalit heijastuvat useimmiten alueen korkeimmista kohdista – esimerkiksi puiden latvoista (Hyypä ym., 2008). Viimeiset paluusignaalit sen sijaan heijastuvat useimmiten maasta. Pisteet voidaan muuttaa rasteriaineistoksi esimerkiksi lineaarisesti interpoloimalla arvot sellaisiin kohtiin, joissa ei ole pisteitä (Mielcarek ym., 2018). Tuloksena muodostuu rasteriaineisto, jossa jokaisessa ruudussa on korkeusarvo. Rasteriaineiston ruudun sivun pituus voi olla esimerkiksi metri, ja ruutua kutsutaan myös pikseliksi. Koska latvuskorkeusmalli on korkeusmallin ja maastomallin erotus, se kuvaa puiden korkeuksia riippumatta maaston muodosta. Kuva 3 havainnollistaa latvuskorkeusmallin periaatetta.



Kuva 3. DSM, DTM ja CHM visualisoituna. Huomaa yhteys $CHM = DSM - DTM$.

Puiden erottelualgoritmeja on monenlaisia, mutta seuraavaksi esitellään Dalponton ja Coomesin (2016) kehittämä algoritmi, joka perustuu Hyypän tutkimusryhmän (2001) kehittämään algoritmiin. Algoritmin ensimmäisessä vaiheessa latvuskorkeusmallia pehmennetään matalataajuussuodattimella (engl. low pass filter) kohinan vähentämiseksi. Toisessa vaiheessa etsitään latvuskorkeusmallista lokaaliset maksimit, jotka vastaavat puiden

latvoja. Lokaaliset maksimit tunnistetaan ympyrän muotoisen liukuvan ikkunan avulla: pikseli merkitään lokaaliksi maksimiksi, jos se on korkeammalla kuin kaikki muut pikselit ympyrän muotoisessa ikkunassa. Latvojen tunnistamisen jälkeen algoritmi arvioi latvuston kokoa laajentamalla latvan pikselistä viereisiin pikseleihin, kunnes naapuripikselien ja latvan suhteellinen korkeusero on suurempi kuin algoritmille määritelty kynnyksisarvo. Lopuksi algoritmi hakee ensimmäiset paluusignaalit löydettyä latvuksen alueelta ja luo monikulmion, joka pitää sisällään kaikki paluusignaalit. Lokaalisen maksimin sijainti vastaa siis puun rungon sijaintia, ja monikulmio rajaa puun latvuston muodon. Puuhun kuuluu kaikki pisteet, jotka jäävät monikulmion sisälle. Dalponten algoritmia ovat käyttäneet esimerkiksi Mäyrä ym. (2021). Myös muissa puunerottelualgoritmeissa käytetään samankaltaisia menetelmiä, kuten lokaalisten maksimien etsimistä latvuskorkeusmallista.



Kuva 4. Esimerkki puiden erottelualgoritmin toiminnasta. Vasemmalla latvuskorkeusmalli (CHM) ja oikealla ortokuva samasta alueesta. Molempien päällä on algoritmin löytämät puut punaisella. Kuva tehtiin itse käyttäen maanmittauslaitoksen 5 pisteen laserkeilausaineistoa (Maanmittauslaitos, 2023), Turun kaupungin tarjoamaa ortokuvaa (Turun kaupunki, 2023) ja Mäyrä ym. (2021) tutkimuksen koodia pienin muokkauksin.

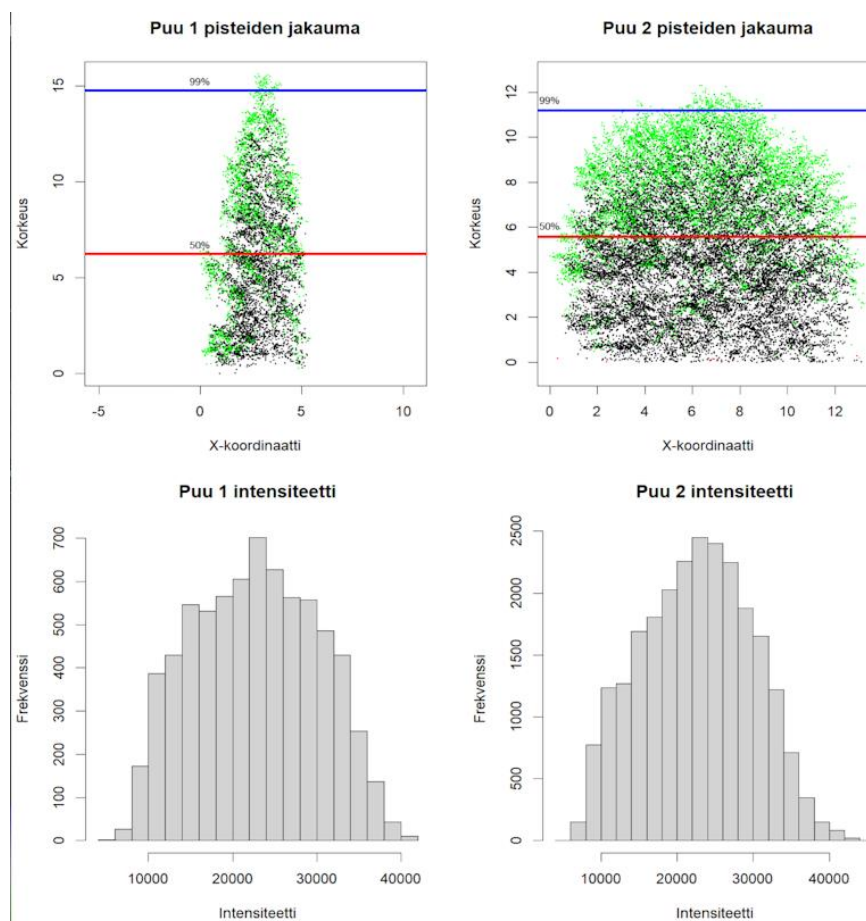
Kun yksittäiset puut on rajattu aineistosta, täytyy rajattuihin puihin yhdistää niiden todellinen puulaji maastomittausaineistosta. Näin muodostuu pareja, joissa on erotettu puu ja sitä vastaava puulaji. Puunerottelualgoritmin arvioimat puun sijainnit ovat arvioita eivätkä ne vastaa täysin oikeiden puiden sijainteja. Parien muodostamiseen on monia eri tapoja. Esimerkiksi Mäyrä ym. (2021) yhdistivät algoritmilla erotettuun puuhun lähimmän maastomittausdatan puun, joka on latvuston rajaavan monikulmion sisällä. Yu ym. (2017) käyttivät sen sijaan Hausdorff-etäisyyttä ja yhdistivät lähimmän koulutusdatan puun, joka on lähempänä kuin viiden metrin maksimietäisyys. Joissakin tutkimuksissa (Suratno ym. 2009; Wang ym. 2024) puiden erottelu jätetään kokonaan välistä ja puihin kuuluvat pisteet määritellään manuaalisesti, mikä parantaa luokittelutarkkuutta, mutta ei vastaa todellista sovellustilannetta.

3.2 Piirteiden laskeminen

Kun puuhun kuuluvat pisteet on selvitetty, niistä lasketaan piirteitä. Klassisen koneoppimisen menetelmät eivät voi käsitellä pistepilviä suoraan, vaan on laskettava numeerisia tunnuslukuja, eli piirteitä. Piirteet kuvaavat puun ominaisuuksia, kuten rakennetta ja intensiteettitietoja. Esimerkiksi kuusen ja männyn rakenne on erilainen. Kuusi on yleensä kartion muotoinen. Männyssä oksat alkavat ylempänä. Luokittelu onnistuu parhaiten, kun yhdistellään intensiteettiin ja puun rakenteeseen liittyviä tietoja (Suratno ym., 2009). Alla on esitelty esimerkkejä piirteistä, joita yksittäisestä puusta voi laskea.

- Puusta lasketaan korkeus, joka on suurempi kuin n % pisteiden korkeuksista (korkeuden prosenttipiste). Latvuston korkeuden arviointiin käytetään usein 99 % rajaa, koska se ei sisällä mahdollisia virhepisteitä. Voidaan esimerkiksi laskea korkeusrajat 10–90 % kymmenen prosentin välein, josta tulee yhdeksän piirrettä. Korkeusjakauma kuvaa puun pystysuuntaista rakennetta. Vastaava luku voidaan laskea myös intensiteetille. Tällöin lasketaan intensiteetti, joka on suurempi kuin n % pisteiden intensiteeteistä. (Torabzadeh ym., 2019; Yu ym., 2017; Shi ym., 2018; Dalponte ym., 2012).
- Pisteet voidaan jakaa ryhmiin: ensimmäiset paluut, ainoat paluut ja muut paluut. Kullekin ryhmälle voidaan laskea erikseen arvoja, jotka kuvaavat puun rakennetta (Kim ym., 2009; Li ym., 2010). Näitä arvoja voivat olla esimerkiksi: ensimmäisten ja yhden paluun pisteiden osuus, pisteiden korkeuksien keskiarvo ja keskihajonta, pisteiden intensiteettien keskiarvo ja keskihajonta (Suratno ym., 2009).
- Latvuston pinta-ala (Li ym., 2010), säde ja tilavuus (Yu ym., 2017; Shi ym., 2018). Nämä voidaan laskea monella eri algoritmilla, mutta esimerkiksi Yu ym. (2017) laskee seuraavalla tavalla. Pinta-ala laskettiin muodostamalla 2D-konvekssi kuori, eli pienin monikulmio, joka peittää kaikki pisteet xy-tasossa. Säde lasketaan sovittamalla ympyrä tähän monikulmioon. Tilavuus sen sijaan on pisteistä lasketun 3D-konvekssi kuoren tilavuus, eli pienin kolmiulotteinen kappale, joka pitää sisällään kaikki pisteet.
- Puu voidaan jakaa pystysuunnassa lohkoihin, joista lasketaan piirteet erikseen. Esimerkiksi kymmenen metrin korkuinen puu voidaan jakaa kymmeneen yhden metrin lohkoon. Lohkoista voidaan laskea muun muassa pistetiheys (Torabzadeh ym., 2019) tai pisteiden suhteellinen osuus verrattuna kaikkiin pisteisiin (Yu ym., 2017).

- Pisteiden intensiteeteistä voidaan laskea erilaisia tilastollisia tunnuslukuja, kuten minimi, maksimi, keskiarvo, keskihajonta, vinous, vaihteluväli ja kurtoosi (Yu ym., 2017; Shi ym., 2018; Li ym., 2010).
- Pisteiden korkeuksista voidaan laskea niin sanottu ”canopy relief ratio”, joka kuvaa pisteiden jakautumista korkeussuunnassa. Se lasketaan kaavalla: $((\text{mean}(z) - \text{min}(z)) / (\text{max}(z) - \text{min}(z)))$. (Marrs ja Ni-Meister, 2019)



Kuva 5. Kaksi puuta x,z-koordinaatistossa ja samojen puiden intensiteetin histogrammit. Viivat kuvaavat 50 ja 99 prosenttipisteitä. Ensimmäiset palaut on värjätty vihreällä. Laserkeilausaineisto on Turun kaupungilta.

Kaikkia piirteitä ei kuitenkaan kannata käyttää luokittelussa (Shi ym., 2018), sillä silloin ongelmaksi muodostuu dimensiokirous (engl. curse of dimensionality). Liian moniulotteinen data suhteessa koulutusdataan johtaa usein mallin ylisovittamiseen ja/tai huonoon suorituskykyyn (Hughes 1968). Lisäksi suuri määrä piirteitä tekee mallista raskaamman suorittaa. Seuraavaksi esitellään joitain menetelmiä, joilla piirteistä voidaan valita ne piirteet, jotka vaikuttavat eniten lopputulokseen. Esimerkiksi vahvasti korreloivia piirteitä ei kannata

ottaa mukaan malliin, sillä ne lisäävät mallin monimutkaisuutta eivätkä lisää tarkkuutta, koska piirteet eivät tarjoa mitään uutta informaatiota (Shi ym., 2018).

Piirteiden valintaan käytetään monenlaisia menetelmiä, kuten korrelaatioanalyysi, PCA, MDA ja SFFS, jotka esitellään lyhyesti seuraavaksi. Korrelaatioanalyysissä lasketaan kaikkien piirteiden korreloiminen keskenään, jolloin syntyy korrelaatiomatriisi (Shi ym., 2018).

Kahdesta vahvasti korreloivista piirteistä valitaan se piirre, joka saa parhaan tuloksen muista testeistä – esim. MDA:sta (Shi ym., 2018). MDA (mean decrease error) indeksi on satunnaismetsäalgoritmissa käytetty mitta. Se kertoo, kuinka paljon keskineliövirhe laskee, kun muuttujaa lisätään malliin (Breiman, 2001). Sen avulla satunnaismetsäalgoritmi pystyy järjestämään muuttujat tärkeysjärjestykseen. PCA (principal component analysis) on sen sijaan menetelmä, joka tiivistää piirteet pienempään määrään piirteitä, joissa on lähes kaikki informaatio kuin alkuperäisissä piirteissä (Mustonen, 1995; Kim ym., 2009). SFFS (sequential forward floating selection) on menetelmä, jossa piirteitä lisätään ja poistetaan riippuen siitä, kuinka paljon ne parantavat tarkkuutta (Pudil ym., 1994; Torabzadeh ym., 2019).

Piirteidenvalinta-algoritmeilla tutkimukset ovat löytäneet piirteitä, jotka pystyvät erottelmaan puulajit huomattavasti toisia piirteitä paremmin. Marrsin ja Ni-Meisterin (2019) tutkimuksessa tärkeimpiin piirteisiin kuuluivat: ensimmäisten paluiden osuus, pistetiheys ylimmässä lohossa kymmenestä lohokosta, 50. ja 100. korkeuden prosenttipiste. Yu:n ym. tutkimuksessa (2017) tärkeimpiin piirteisiin kuului muun muassa: intensiteetin minimi, paluiden osuus alle kahden metrin korkeudessa, pisteiden osuus viidennessä ja yhdeksännessä lohossa ja korkeuden 90. prosenttipiste.

3.3 Laserkeilausaineiston ja spektrikuvien yhdistäminen

Spektrikuvat eroavat tavallisista valokuvista siinä, että spektrikuvat tallentavat aallonpituuksia, joita ihmissilmä ei näe. Puut voidaan erottaa niiden tuottaman spektrin perusteella. Spektrikuvat voidaan luokitella multi- ja hyperspektrikuviin. Multispektrikuvat tallentavat yleensä noin 3–8 aallonpituuskaistaa, kun taas hyperspektrikuvat voivat tallentaa satoja aallonpituuskaistoja (Ballanti ym., 2016). Aallonpituuskaistat muodostavat spektrin, joka vaihtelee puulajin mukaan. Hyperspektrikuvat soveltuvat kuitenkin huomattavasti paremmin puulajien erotteluun kuin multispektrikuvat niiden suuren kaistamäärän vuoksi (Clark ym., 2005).

Sekä multi- että hyperspektrikuvien avulla voidaan luokitella puulajeja ilman muita aineistoja (Ballanti ym., 2016). Luokittelutarkkuus kuitenkin kasvaa, jos luokittelussa käytetään sekä spektrikuvia että laserkeilausaineistoa (Mäyrä ym., 2021; Torabzadeh ym., 2019). Aineistojen käyttäminen yhdessä on usein myös kustannustehokasta. Tämä johtuu siitä, että spektrikuvien hankkiminen on halvempaa kuin laserkeilausaineiston (Dalponte ym., 2012). Kustannuksia säästyy etenkin, kun alueesta on jo olemassa spektrikuvia tai pienen pistetiheyden laserkeilausaineistoja.

Aineistojen yhteiskäyttö on tehokasta, sillä laserkeilausaineisto sisältää tiedon korkeuksista, joita spektrikuva ei voi tallentaa. Dalponten tutkimusryhmän tutkimuksessa (2012) laserkeilausaineistosta laskettujen piirteiden lisääminen luokittelumalliin lisäsi tarkkuutta noin 11–13 prosenttiyksikköä riippuen käytetystä luokittelumallista. Piirteet liittyivät pisteiden suhteelliseen jakautumiseen eri korkeuksilla. Tutkimuksessa testattiin kahta eri aineistoa, joiden pistetiheydet ovat noin 0.48 ja 8.6 neliometriä kohden. Myös harvemman aineiston käyttö paransi tarkkuutta noin 7–10 prosenttiyksikköä, mikä osoittaa, että myös pieni pistetiheys riittää spektrikuvien tueksi.

Spektrikuvia täytyy kuitenkin esikäsitellä. Koko aluetta ei usein voida kuvata yhteen hyperspektri tai multispektri kuvaan. Kuvat otetaan eri kellonaikaan ja siksi kirkkaus voi vaihdella paljon eri kuvien välillä. Ongelmaa yritetään ratkaista normalisoimalla jokaisen aallonpituuden arvo jakamalla se kyseisen pisteen kaikkien aallonpituuksien summalla (Dalponte ym., 2016). Näin saadaan suhteellinen aallonpituus koko spektristä.

Kuten laserkeilausaineistosta laskettujen piirteiden kohdalla, myös hyperspektridatasta pitää valita tärkeimmät piirteet, jotta vältetään ylisovittaminen (Ballanti ym., 2016). Piirteet voidaan sen jälkeen syöttää mallille yhdessä laserkeilausaineistosta laskettujen piirteiden kanssa. Spektridatasta lasketut piirteet ovat käytännössä tiettyjen tärkeiden spektrikaistojen arvot puuhun kuuluvissa pikseleissä.

3.4 Lineaarinen erotteluanalyysi

Lineaarinen erotteluanalyysi (engl. Linear discriminant analysis, LDA) on luokittelumenetelmä, jossa jokainen piirre on ulottuvuus moniulotteisessa avaruudessa ja havainnot ovat pisteitä (Tharwat ym., 2017). Puulajien luokittelussa havainnot ovat yksittäisiä puita ja niiden piirteet määrittävät niiden sijainnin. LDA etsii hypertasoja, jotka erottelevat

luokat toisistaan mahdollisimman hyvin ja minimoivat luokan sisäisen hajonnan (Tharwat ym., 2017).

LDA olettaa, että jokaisella luokalla on sama piirteiden kovarianssimatriisi ja että piirteet seuraavat multinormaalijakaumaa (Mustonen, 1995). Multinormaalijakauma tarkoittaa sitä, että kaikki piirteiden lineaariset yhdistelmät noudattavat normaalijakaumaa. Nämä oletukset eivät usein täyty puista lasketuissa piirteissä, mikä voi huonontaa mallin tarkkuutta (Immitzer ym., 2012). Toinen LDA:n rajoitus on, että se osaa erotella luokkia vain lineaaristen riippuvuuksien avulla (Tharwat ym., 2017).

Esimerkiksi Suratno ym. (2009) laskivat 14 piirrettä jokaisesta puusta ja käyttivät LDA:ta puulajien luokitteluun. LDA luokitteli 65 % puista oikein tutkimuksen testialueella. Tutkimuksessa käytettiin hyvin harvaa aineistoa ($0.44/m^2$), mikä selittää osittain huonoa tarkkuutta.

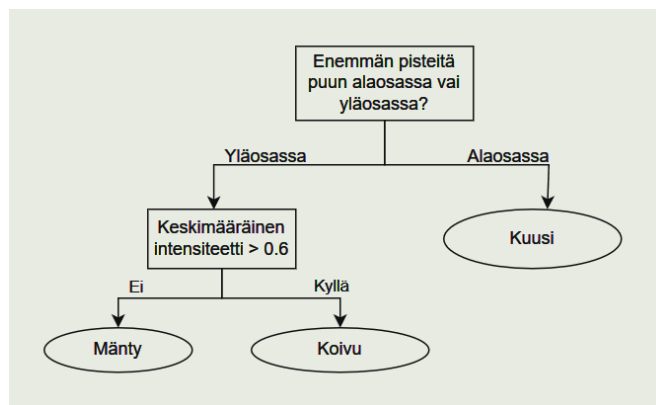
Lineaarista erotteluanalyysiä käytettiin eniten vuosina 2008–2013 (Zhong ym., 2024). Sen jälkeen satunnaismetsien ja tukivektorikoneiden suosio ylitti sen suosion.

3.5 Satunnaismetsä

Satunnaismetsä (engl. Random Forest) on yksi käytetyimmistä algoritmeista puulajien tunnistuksessa (Michałowska & Rapiński, 2021). Suosio johtuu sen helppokäyttöisyydestä ja kyvystä tunnistaa ne muuttujat, jotka erottelevat puulajit parhaiten (Yu ym., 2017).

Breimanin (2001) kehittämä satunnaismetsäalgoritmi perustuu päätöspuihin. Kuvassa 6 on havainnollistava esimerkki yksinkertaistetusta päätöspuusta, joka päättelee, onko kyseessä kuusi, mänty vai koivu. Satunnaismetsäalgoritmi luo monta matalaa päätöspuuta, joista jokainen määrittää puun lajin erikseen. Lopulliseksi puulajiksi valitaan se puulaji, joka esiintyy useimmiten päätöspuiden lopputuloksena. Päätöspuiden määrä on mallille syötettävä parametri. Kun yksittäinen päätöspuu luodaan, sille syötetään bootstrap-otos koulutusaineistosta. Päätöspuu pyrkii luomaan puuhun haaroja niin, että luokat voidaan erotella niiden avulla mahdollisimman hyvin. Jokaisessa haarassa valitaan satunnainen otos piirteitä, joiden perusteella jako tehdään. Algoritmille voi syöttää parametrina sen, kuinka montaa satunnaista piirrettä käytetään puun solmuissa havaintojen erotteluun. Satunnaistamisen vuoksi jokainen päätöspuu on erilainen

Satunnaismetsäalgoritmia käytetään esimerkiksi Yu:n tutkimusryhmän tutkimuksessa (2017). Tutkimuksessa satunnaismetsää käytettiin ensin tärkeimpien piirteiden löytämiseen ja sitten luokitteluun. Satunnaismetsä valitsi 15 tärkeintä piirrettä 145 piirteestä. Luokitteluun käytettiin mallia, jossa on 200 päätöspuuta. Jokaisessa haarassa päätöspuu käytti neljää satunnaisesti valittua piirrettä havaintojen erotteluun.



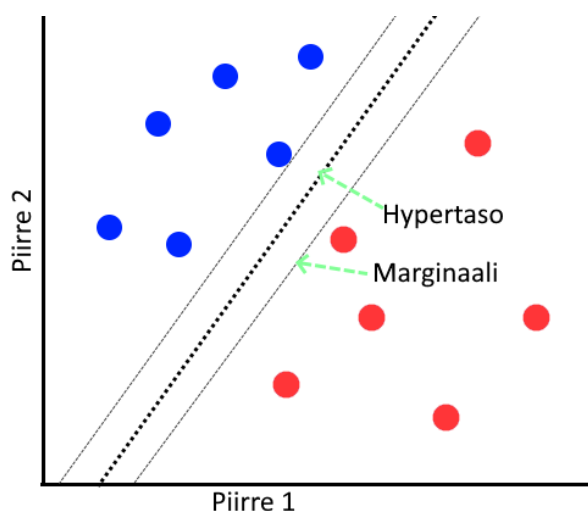
Kuva 6. Yksinkertainen päätöspuu, joka päättelö piirteistä, onko kyse kuusesta, männystä vai koivusta

3.6 Tukivektorikone

Tukivektorikone on alun perin kahden luokan erotteluun kehitetty menetelmä, jossa havainnot ovat pisteitä moniulotteisessa avaruudessa (Cortes ja Vapnik, 1995). Havainnot erotetaan sovittamalla niiden väliin hypertaso. Hypertaso sovitetaan niin, että se maksimoi pisteiden etäisyyden tasosta ja minimoi väärällä puolella olevat pisteet. Kuvassa 7 on havainnollistettu yksinkertaista kahden luokan erottelutilannetta kahdella piirteellä. Tukivektorikone pystyy jakamaan havainnot myös epälineaarisisessa tilanteessa kernelien avulla (Boser ym., 1992). Kernelit lisäävät ulottuvuuksia dataan niin, että luokat voidaan erotella lineaarisesti hypertasolla

Tukivektorikonetta voi myös käyttää tilanteeseen, jossa on monta luokiteltavaa luokkaa. Silloin tarvitaan monta kahden luokan tukivektorikonetta. Tähän on kaksi strategiaa: ”yksi vastaan kaikki” (engl. one against all) tai ”yksi vastaan yksi” (engl. one against one). Ensimmäiseksi mainitussa kutakin luokkaa vertaillaan muiden luokkien kanssa niin, että muut luokat kuuluvat yhteiseen luokkaan. Toisessa strategiassa vertaillaan kaikkia mahdollisia luokkien pareja toisiinsa. Puulajintunnistuksessa ”yksi vastaan yksi” strategia voi mahdollisesti toimia paremmin puulajinluokittelussa, sillä eri puulajeissa on usein eri määrä havaintoja. (Torabzadeh ym., 2019)

Tukivektorikonetta käytti esimerkiksi Torabzadeh ym. (2019). He päättivät käyttää juuri tukivektorimallia, koska se toimii suurella määrällä piirteitä, on vakaa, ei jää jumiin lokaaleihin maksimeihin ja toimii myös pienellä määrällä koulutusdataa. Tukivektorikoneelle syötettiin sekä laserkeilausaineistosta että hyperspektrikuvista laskettuja piirteitä. Tutkimuksessa kokeiltiin erilaisia piirteiden yhdistelmiä, mutta parhaaksi osoittautui piirrejoukko, jossa oli 82 piirrettä.



Kuva 7. Tukivektorikone kahden luokan tapauksessa, kun piirteitä on kaksi. Hypertaso on viiva, kun piirteitä on vain kaksi.

3.7 Neuroverkot

Erityisesti viime vuosina puiden luokittelu tehdään yhä useammin neuroverkoilla. Neuroverkoille syötetään tiettyyn puuhun kuuluvat pisteet ja tuloksena syntyy luokittelu (Liu ym., 2022). Puuhun kuuluvat pisteet saadaan joko manuaalisesti erottelemalla (Wang ym., 2024) tai puiden erottelualgoritmeilla (Liu ym., 2022). Neuroverkot vaativat paljon oikein luokiteltua koulutusdataa (Taher ym., 2025). Koulutusdatan avulla verkko oppii itse tunnistamaan puulajeille tyypillisiä piirteitä. Verkon toimivuutta testataan testiaineistolla, jossa on eri puita kuin koulutusaineistossa. Testiaineistolla voidaan testata, kuinka hyvin verkko toimii uusilla puilla, joita se ei ole aiemmin nähnyt. Neuroverkkoratkaisut voidaan jakaa pistepilviin ja 2D-profiileihin perustuviin menetelmiin (Taher ym., 2025). Pistepilviin perustuvissa menetelmissä neuroverkolle syötetään puuhun kuuluvat pisteet. Profiiliratkaisuissa neuroverkolle syötetään eri kulmista otettuja 2D-projektioita puun pisteistä.

Taher ym. (2025) tutkimuksessa vertaillaan eri menetelmien toimivuutta Espoossa sijaitsevalla testialueella. Tutkimuksen mukaan pistepilvi- ja profiilimalleissa käytetään eri neuroverkkoarkkitehtuureja. Pistepilvimalleissa käytetään arkkitehtuureja, kuten PointNet, PointNet++, Point2Vec, Dynamic Graph CNN ja pistemuunnin mallit. Profiilimalleissa käytetään arkkitehtuureja, kuten YOLOv8, DenseNet ja ConvNeXt-T.

Profiilimalleja ei käsitellä sen tarkemmin tässä tutkielmassa. Seuraavaksi käsitellään pistepilvimalleja. Neuroverkkomallit eroavat satunnaismetsistä ja tukivektorikoneista siinä, että niille voi syöttää puun pisteet niistä laskettujen piirteiden sijaan. Neuroverkoille ei silti kannata syöttää pisteitä ilman esikäsitelyä. Esikäsitely auttaa mallia keskittymään olennaiseen. Esikäsitelytapoihin kuuluu esimerkiksi virhepisteiden poistaminen (Liu ym., 2022), puupisteiden erottelu maapisteistä (Liu ym., 2022), pisteiden koordinaattien normalisointi (Marrs ja Ni-Meister, 2019) ja syötteen koon yhtenäistäminen (Wang ym., 2024; Marrs ja Ni-Meister, 2019).

Laserkeilauksessa voi syntyä virhepisteitä, jotka johtuvat keilaimen rajoituksista, sääolosuhteista tai esimerkiksi lentävistä linnuista, jotka heijastavat lasersäteitä (Petrie ja Toth, 2008). Suhteettoman korkealla tai matalalla olevat pisteet suhteessa maan korkeuteen voidaan poistaa asettamalla korkeusraja, jonka ylä- tai alapuolelta poistetaan kaikki pisteet (Liu ym., 2022). Muista pisteistä erillään olevat pisteet voidaan poistaa selvittämällä niiden pisteiden määrä, jotka ovat tietyn säteen etäisyydellä käsitellystä pisteestä (Liu ym., 2022). Jos pisteiden määrä alittaa asetetun kynnyksarvon, piste poistetaan. Virhepisteiden lisäksi pitää erotella maahan ja puihin kuuluvat pisteet, koska mallille syötetään vain puun pisteet (Liu ym., 2022).

Puiden pisteiden koordinaatit voidaan standardisoida niin, että puun keskusta on koordinaatiston origo (Wang ym., 2024). Pisteiden koordinaatit voidaan myös skaalata miinus yhdestä yhteen, jotta puun koko ei vaikuta luokitteluun (Marrs ja Ni-Meister, 2019). Näin varmistetaan, että malli ei opi luokittelemaan puita niiden maantieteellisten sijaintien perusteella. Sen sijaan mallin pitäisi tutkia puun sisäistä rakennetta.

Kolmas mahdollinen esikäsitelytapa koskee syötteen kokoa. Monissa pistepilven neuroverkkoratkaisuissa syötteen on oltava aina samankokoinen (Marrs ja Ni-Meister, 2019). Siksi verkon suorituskykyä pystyy parantamaan lisäämällä tai vähentämällä tietyn puun pisteitä riippuen siitä onko niitä vähemmän tai enemmän kuin suurimmassa osassa puita (Wang ym., 2024). Pisteitä voi vähentää poistamalla satunnaisesti valittuja pisteitä ja lisätä

luomalla kopioita pisteistä (Wang ym., 2024). Pisteiden vähentämiseen on monia eri algoritmeja, joihin voi tutustua Liu ym. (2022) tutkimuksessa. Nykyään on myös ratkaisuja, joissa syötteen koolla ei ole väliä (Taher ym., 2025).

Neuroverkot vaativat paljon koulutusaineistoa toimiakseen hyvin ja liian pieni koulutusaineisto johtaa helposti ylisovittamiseen. Ylisovittamista voidaan ehkäistä laajentamalla koulutusaineistoa tekemällä kopioita olemassa olevista puista siten, että kopioita muokataan hieman satunnaisuuden avulla (Mäyrä ym., 2021; Wang ym., 2024). Pisteitä voidaan esimerkiksi liikutella muutamia senttejä satunnaisiin suuntiin tai puuta voidaan kääntää rungon suuntaisesti niin, että se osoittaa eri suuntaan. On tärkeää, että muutokset eivät ole liian isoja, jotta ne eivät anna väärää tietoa neuroverkolle.

Wang tutkimusryhmineen (2024) käyttivät neuroverkkoon pohjautuvaa mallia puulajin tunnistukseen pienen tiheyden laserkeilausaineistosta. Heidän neuroverkkonsa perustuu muunninmalliin, jolla on laajempi konteksti kuin useimmilla muilla neuroverkkomalleilla. He esikäsittelivät aineistoa aiemmin mainituilla tavoilla. Heidän mallinsa on kaksihaarainen, jossa ensimmäiselle haaralle syötetään vain pisteiden intensiteetit ja toiselle haaralle syötetään pisteiden sijainnit. Näiden ominaisuuksien erillinen käsittely parantaa heidän mukaansa luokittelukykyä. Intensiteetistä kerättyjä piirteitä annetaan kuitenkin myös toisen haaran syötteeksi mallin eri vaiheissa. Heidän käyttämässään laserkeilausaineistossa on intensiteetti kolmelta eri aallonpituusalueelta, mikä auttaa luokittelussa.

4 Tutkimusten tuloksia

Laserkeilausaineistoa käytetään puuntunnistuksessa, koska se kuvaa puiden kolmiulotteista rakennetta. Laserkeilaus antaa myös tietoa pintojen heijastamiskyvystä. Laserkeilausaineiston käsittelymenetelmällä on suuri merkitys luokittelun tarkkuuteen, mutta menetelmien vertailu on haastavaa, koska tarkkuuteen vaikuttaa moni tekijä.

Tässä luvussa käsitellään tutkimusten tuloksia ja tekijöitä, jotka vaikuttavat luokittelutarkkuuteen keskittyen erityisesti pistetiheyden vaikutukseen.

4.1 Luokittelutarkkuudet ja niihin vaikuttavat tekijät

Tässä luvussa käsitellään, miten hyvin eri tutkimuksissa on onnistuttu luokitelemaan puulajeja laserkeilausaineiston avulla. Tutkimusten tulosten vertailu on haastavaa, sillä niissä on käsitelty erilaisia aineistoja ja luokiteltu erilaisia puulajeja erilaisilta alueilta. Pieni määrä lajeja on helpompi luokitella, sillä sattuma vaikuttaa sitä vahvemmin mitä pienempi osuus kullakin luokalla on kaikista näytteistä. Tätä tasapainottamaan on kehitetty Cohenin kappa κ (Cohen, 1960), joka saa arvoja välillä $-1-1$. Suuremmat arvot tarkoittavat parempaa luokittelutarkkuutta. Kappa mittaa kuinka paljon parempi luokittelija on verrattuna satunnaisesti saavutettavaan tulokseen. Cohenin kappa lasketaan seuraavalla kaavalla sekaannusmatriisista (engl. confusion matrix), jossa on luokitteluiden frekvenssit.

$$\frac{\sum_i^m X_{ii} - \frac{1}{N} \cdot \sum_i^m X_{i+} \cdot X_{+i}}{N - \frac{1}{N} \cdot \sum_i^m X_{i+} \cdot X_{+i}}$$

Jossa: N on näytteiden määrä, m on eri luokkien määrä, X_{ii} on matriisin diagonaalien arvo ja X_{i+} sekä X_{+i} ovat matriisien rivien ja sarakkeiden summa.

Taulukkoon 1 on listattu valittujen tutkimusten pistetiheys, käytetyt menetelmät, paras tarkkuus ja lajien määrä. Taulukkoon valittiin tutkimuksia, joissa käytetyn diskreetin laserkeilausaineiston tiheys on alle 20 neliometriä kohden ja sovellusalue on havupuuvaltainen.

Taulukko 1 Tutkimusten tulokset ja käytetyt menetelmät ja pistetiheys

Menetelmistä käytetään lyhenteitä SVM (tukivektorikone), RF (satunnaismetsä), SPE (yhdistäminen spektrikuvien kanssa), NE (neuroverkot) ja LDA (lineaarinen erotteluanalyysi). Tarkkuudella tarkoitetaan oikein luokiteltujen puiden osuutta kaikista puista. Myös Kappa-arvo (κ) on ilmoitettu, jos se laskettiin tutkimuksessa.

Lähde	Pistetiheys	SVM	RF	SPE	NE	LDA	Paras tarkkuus	Lajeja
Suratno ym., 2009	0,44					x	63,6 (κ 56)	4
Korpela ym., 2010	6–8		x			x	90,8 (κ 84,4)	3
Dalponete ym., 2012	0,48 ja 8,6	x	x	x			83,0 (κ 77,2)	7
Ørka ym., 2013	7,2		x	x			87,0 (κ 78)	3
Yu ym., 2014	1,5		x				62,1	3
Lin ja Hyyppä, 2016	10	x					92,5	4
Mäyrä ym., 2021	10,2	x	x	x	x		87	4
Reisi Gahrouei ym., 2024	2,5	x	x	x	x		78 (κ 72)	9
Wang ym., 2024	9				x		83,1	6

Tutkimusten menetelmien toimivuutta on kuitenkin vaikea vertailla, sillä niiden tutkimusalueet ja laserkeilausaineisto ovat erilaisia (Taher ym., 2025). Tutkimusalueella voi olla esimerkiksi paljon nuorta tai vanhaa puuta. Borealisissa havumetsissä puuston tiheys ja monimutkaisuus ovat pienempiä kuin lauhkean vyöhykkeen metsissä (Heinzel ja Koch, 2012). Joissain tutkimuksissa luokitellaan samankaltaisia puita, kun taas toisissa puut eivät muistuta toisiaan (Taher ym., 2025).

Myös mittauksen vuodenajalla on vaikutusta. Erityisesti lehtipuut muuttuvat paljon vuoden mittaan. Puun lehtien vaiheita ovat esimerkiksi silmujen puhkeaminen, lehtien kasvu, kukinta, lakastuminen ja lepokausi. Nämä eri vaiheet muuttavat puun latvuston rakennetta ja saatuja intensiteetti-arvoja (Shi ym., 2018). Monessa tutkimuksessa on myös todettu, että on hyödyllistä yhdistää aineistoja eri ajoilta niin, että toisessa aineistossa puissa on lehdet (leaf-on) ja toisessa ei (leaf-off) (Shi ym., 2018; Kim ym., 2009). Lehtitilanteen vaikutuksesta on erilaisia näkemyksiä. Osassa tutkimuksissa tilastollisesti merkittävää eroa ei ole (Shi ym., 2018) mutta toisissa leaf-off data on tuottanut parempia tuloksia (Kim ym., 2009; Imangholiloo ym., 2020). Leaf-off datan paremmat tulokset saattavat johtua siitä, että havupuut on helpompi erottaa lehtipuista, koska havupuut eivät pudota neulasiaan (Imangholiloo ym., 2020). Tällöin havupuista tulee enemmän paluita ja intensiteetti-arvoista voidaan päätellä, tuleeko paluu rungosta vai lehdistä.

Tutkimukset käyttävät eri tavalla kerättyä laserkeilausaineistoa. Peittoalue vaikuttaa siihen, kuinka syvälle latvustoon pulssit yltävät ja laserkeilaimen käyttämä aallonpituus vaikuttaa intensiteettiarvoihin. Lisäksi intensiteettiarvot voivat olla kalibroituja niin, että ne ottavat huomioon lentokorkeuden, laserpulssin kulman ja sään. Säteen pituus hajauttaa lasersädettä, vinosti osuva laserpulssi voi tuottaa erilaisen intensiteetin ja sää voi vaikuttaa ilmassa oleviin partikkeleihin, jotka heijastavat osan lasersäteestä. (Harding, 2008)

Lisäksi tutkimukset kouluttavat ja testaavat malliaan eri määrällä aineistoa. Osassa tutkimuksista on laaja koulutusaineisto ja testiaineisto (Wang ym., 2024; Mäyrä ym., 2021; Li ym., 2013), jotka on erotettu toisistaan tarkasti. Monissa aineisto on kuitenkin pieni, jolloin mallia testataan k-kertaisella ristiinvalidoinnilla (Suratno ym., 2009; Lin ja Hyyppä, 2016; Zhang ym., 2016). Siinä aineisto jaetaan k osaan, ja joka kierroksella yksi osa jätetään koulutuksen ulkopuolelle (Zhang ym., 2016). Mallin suorituskykyä testataan poisjätetyllä osalla. Koulutusaineisto ja testiaineisto vaikuttavat tuloksen yleistettävyyteen paljon, sillä pieni otoskoko edustaa vain pientä osaa metsän puista, jolloin malli saattaa toimia hyvin vain tietynlaisilla puilla. Suuri otoskoko parantaa mallin yleistettävyyttä erilaisiin puihin ja metsiin.

Luokittelutarkkuuteen vaikuttaa siis ainakin luokiteltavat puulajit, tutkimusalue, mittauksen vuodenaika, laserkeilaimen asetukset, koulutusaineiston koko ja mallin testaustapa. Tämän vuoksi tutkimuksia ei voi laittaa paremmuusjärjestykseen niiden luokittelutarkkuuden perusteella.

4.2 Pistetiheyden vaikutus tarkkuuteen

Kuten aiemmassa luvussa kerrottiin, luokittelutarkkuuteen vaikuttaa moni asia. Tässä luvussa keskitytään erityisesti pistetiheyden vaikutukseen. Pistetiheyden ja luokittelutarkkuuden välillä on havaittu positiivinen lineaarinen korrelaatio (Li ym., 2013). Li:n tutkimuksessa (2013) puita luokiteltiin eri pistetiheyksillä. Alkuperäistä tiheää aineistoa harvennettiin poistamalla pisteitä ja harvennettu aineisto käsiteltiin samalla tavalla kuin tiheä aineisto. Tiheällä aineistolla (90 pistettä / m²) saavutettiin tarkkuus 77.5 %. Kaikista eniten harvennetulla aineistolla (2 pistettä / m²) tarkkuus oli noin 50 %. Muilla pistetiheyksillä lasketut tarkkuudet sijoittuivat niiden väliin niin, että tuloksiin voitiin sovittaa suora, jonka R²-luku on 0.88.

Toisaalta kaikki menetelmät eivät toimi yhtä hyvin harvalla aineistolla: esimerkiksi neuroverkkojen tapauksessa verkkoarkkitehtuuria pitää hiukan muokata, jotta se ottaisi huomioon pienen syötteen (Wang ym., 2024). Todella tiheä pistepilvi ei myöskään välttämättä paranna tarkkuutta, vaan lisää tarpeetonta tietoa (Liu ym., 2021). Liu ym. (2021) totesivat, että heidän neuroverkkonsa luokittelutarkkuus ei kasvanut, kun yksittäisen puun pisteiden määrä ylitti noin 2000 pisteen rajan. Taher ym. (2025) tutkimuksessa syväoppivat neuroverkot toimivat paremmin kuin klassiset koneoppimismenetelmät.

Pistetiheys vaikuttaa siihen, kuinka hyvin pistepilvi kuvaa todellisuutta. Pienempi pistetiheys johtaa siihen, että pisteet osuvat harvemmin puiden latvoihin, eli korkeimpiin kohtiin (Imangholiloo ym., 2020). Pienen pistetiheyden aineistot siis aliarvioivat puiden korkeuksia. Puun latvan sijaintia ja korkeutta käytetään yksittäisten puiden erotteluun pistepilvestä (Dalponte ym., 2016). Huonosti erotellut puut laskevat puulajiluokittelun tarkkuutta (Taher ym., 2025).

5 Pohdinta

Tässä luvussa tarkastellaan pienen otoskoon ongelmaa tutkimuksissa, mallien yleistettävyyttä ja käytännön sovelluskohteita puulajitiedolle.

Tutkimukset keskittyvät tietyn alueen puihin ja mallien kalibrointi on edelleen tarpeellista. Koulutusdata pitää luoda aluekohtaisesti. Tämä tarkoittaa, että on edelleen tarpeellista käydä tekemässä manuaalisia mittauksia metsässä, jos metsä ei vastaa ominaisuuksiltaan koulutusdatan metsää. Tutkimusten tekemistä nopeuttaisi ja helpottaisi, jos tutkijoilla olisi käytössään avoimia aineistoja, joissa on maastomittauksen tiedot mukana. Yhteiset aineistot mahdollistaisivat tulosten vertailun, eikä tutkijoiden tarvitsisi käyttää aikaa maastomittausten tekemiseen. Avoimet aineistot ratkaisisivat myös pienen otoskoon ongelman. Monessa tutkimuksessa (Lin ja Hyypä, 2016; Imangholiloo ym., 2020) otoskoko on vain muutaman sadan puun luokkaa tai vähemmän. Esimerkki avoimesta aineistosta on Puliti ym. (2025) kokoama FOR-species20K tietoaaineisto, jossa on 20000 puuta 33:sta eri puulajista. Tosin tämän tietoaaineiston yleistettävyyys on kyseenalaistettu Taher ym. (2025) tutkimuksessa, jossa testialueen pistepilvellä koulutettu malli toimi huomattavasti paremmin kuin vain FOR-species20K aineistolla koulutettu malli.

Toisaalta on myös hyvä, että tutkimuksissa on erilaiset ympäristöt, jossa puita luokitellaan. Näin saadaan lisää tietoa erilaisista metsätyypeistä. Maastomittausten tekemistä voisi kuitenkin automatisoida jollain tavalla, kuten esimerkiksi käyttämällä maassa olevaa liikkuvaa laserkeilainta, jonka pistetiheys on hyvin suuri. Tiheän pistepilven voisi sen jälkeen syöttää hyväksi todetulle mallille, joka osaa hyödyntää suurta pistetiheyttä. Näin tehdessä ihmisten tarvitsisi vain viedä laserkeilain paikalle ja puista kerättäisiin tiedot automaattisesti. Tämä menetelmä tosin vaatisi, että olisi olemassa tarpeeksi hyvä malli, että se voisi korvata ihmisen. Ajatus olisi siis, että paikallisella pienen kantaman laserkeilaimella voisi tuottaa maastomittausdataa mallien kouluttamiseen, jotta voitaisiin käsitellä suuria alueita.

Tutkielman perusteella harvasta laserkeilausaineistosta on eniten hyötyä, kun sitä käyttää yhdessä hyperspektrikuvien kanssa. Neuroverkkoteknologiat vaikuttavat lupaavimmalta teknologialta muun muassa sen vuoksi, että niitä varten ei tarvitse laskea piirteitä. On myös osoitettu, että ne toimivat hyvin myös harvalla laserkeilausaineistolla (Wang ym. 2024; Taher ym., 2025) Suurin osa uudesta tutkimuksesta käyttää neuroverkkoja. Laserkeilausaineiston pistetiheyden vaikutuksesta luokittelutarkkuuteen tarvitaan silti enemmän tutkimusta.

Pistetiheyden vaikutusta voi tutkia esimerkiksi harventamalla tiheää laserkeilausaineistoa poistamalla pisteitä (Taher ym., 2025; Li ym., 2013). Tosin on huomioitava, että pieni pistetiheys on usein seurausta lentokorkeuden nostamisesta, joka vaikuttaa myös muun muassa peittoalueen kokoon, joka vuorostaan vaikuttaa pisteiden tarkkuuteen (Taher ym., 2025).

Olisi tärkeää saada mallit toimimaan myös pienen pistetiheyden aineistolla, koska se alentaisi metsäinventaarion hintaa. Hinnan aleneminen toisi hyödylliset laserkeilauksesta lasketut tiedot useammalle metsän omistajalle. Tiedot voitaisiin jakaa metsänomistajille puukarttojen avulla. Puukartat sisältävät puiden sijainnit ja puihin liittyviä ominaisuuksia, kuten rinnankorkeusläpimitan (Nevalainen ym., 2020). Ominaisuuksiin voisi lisätä myös puulajin, jolloin metsän puulajeja pystyisi tarkastelemaan visuaalisesti. Puukartat voivat olla hyödyllisiä myös metsäkoneiden käyttäjille, koska niiden avulla osa prosessista voidaan automatisoida (Nevalainen ym., 2020). Alueen puulajitietojen avulla voitaisiin mahdollisesti varmistaa harvennettavan alueen monimuotoisuus jättämällä ekologisesti tärkeät puut metsään kasvamaan.

Tulevaisuudessa tutkimuksen tulisi keskittyä yhä enemmän yleisen mallin luomiseen, joka toimisi erilaisissa metsissä. Yksi lupaava lähestymistapa on siirto-oppiminen, jossa malli koulutetaan ensin laajalla eri alueilta kerätyllä aineistolla ja kalibroidaan sitten tavoitealueelta kerätyllä pienellä koulutusaineistolla. Toinen mahdollinen trendi on laserkeilaimet, jotka keilaavat maastoa monella eri aallonpituudella. Tällainen laserkeilausaineisto vastaa käytännössä multispektrikuvia, mutta spektritiedot ovat integroitua laserkeilausaineistoon. Esimerkiksi Taher ym. (2025) tutkimus keskittyy juuri multispektriaineistoon, joka sisältää dataa kolmelta eri aallonpituusalueelta. Tulevaisuudessa parhaat tulokset saadaan siis mahdollisesti yhdistämällä suuret koulutusaineistot, siirto-oppiminen, syväoppivat neuroverkot ja multispektriaineistot.

Yhteenveto

Tässä tutkielmassa käsiteltiin käytännönläheisesti, miten puulajeja voidaan luokitella laserkeilausaineiston avulla. Tutkielman tutkimuskysymykset olivat

TK1: Mitä algoritmeja voi käyttää puulajin luokitteluun laserkeilausaineistoa käytettäessä?

TK2: Miten pistetiheys vaikuttaa luokittelutarkkuuteen?

Luokitteluun käytettävät koneoppimisalgoritmit voidaan jakaa syväoppiviin neuroverkkoihin ja klassisiin koneoppimismenetelmiin. Neuroverkkoratkaisuissa riittää vähäinen pistepilven esikäsittely ennen mallille syöttämistä. Muut menetelmät vaativat sen sijaan piirteiden laskemista ja merkittävimpien piirteiden valintaa. Mahdollisia piirteitä ovat esimerkiksi pisteiden korkeusien prosenttipisteet, ensimmäisten ja ainoiden paluiden osuus, latvuston pinta-ala sekä pisteiden intensiteetistä lasketut tilastolliset tunnusluvut. Molempia lähestymistapoja käytettäessä pitää kuitenkin ensin erottaa yksittäiset puut pistepilvestä. Yksi tällainen algoritmi esitellään Dalponten ym. (2016) tutkimuksessa. Puulajintunnistukseen käytetyt koneoppimismenetelmät ovat seuranneet koneoppimisen kehitystä. Uudemmat tutkimukset keskittyvät syväoppiviin neuroverkkoihin. Vanhat tutkimukset käyttävät sen sijaan menetelmiä, kuten lineaarinen erotteluanalyysi, satunnaismetsä ja tukivektorikone. Kaikkia kuvattuja menetelmiä on käytetty myös harvoilla pistepilvillä ja niillä on saavutettu hyviä tarkkuuksia.

Pistetiheydellä on selkeä vaikutus luokittelutarkkuuteen. Suurempi pistetiheys tuottaa lähtökohtaisesti tarkempia tuloksia, mutta myös pienen tiheyden ($<20/m^2$) aineistolla on saavutettu korkeita luokittelutarkkuuksia, esimerkiksi: 87 % (Mäyrä ym., 2021), 83.1 % (Wang ym., 2024) ja 87 % (Ørka ym., 2013).

Alan haasteet liittyvät tulosten yleistettävyyteen, sillä luokittelutarkkuuteen vaikuttavat ainakin luokiteltavat puulajit, tutkimusalue, mittauksen vuodenaika, laserkeilaimen asetukset, koulutusaineiston koko ja mallin testaustapa. Tämän vuoksi tutkijoiden pitäisi yhä enemmän keskittyä mallinsa monipuoliseen testaukseen erilaisissa ympäristöissä ja mahdollisimman suurella otoksella puita.

Lähteet

- Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers. *Remote Sensing*, 8(6), 445. <https://doi.org/10.3390/rs8060445>
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152. <https://doi.org/10.1145/130385.130401>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96(3–4), 375–398. <https://doi.org/10.1016/j.rse.2005.03.009>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Dalponte, M., Bruzzone, L., & Gianelle, D. (2012). Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*, 123, 258–270. <https://doi.org/10.1016/j.rse.2012.03.013>
- Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, 7(10), 1236–1245. <https://doi.org/10.1111/2041-210X.12575>
- Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103(2), 140–152. <https://doi.org/10.1016/j.rse.2006.03.003>
- Graham, L. (2008). Management of LiDAR Data. Teoksessa J. Shan & C. K. Toth (Toim.), *Topographic Laser Ranging and Scanning* (1. p., s. 295–306). CRC Press. <https://doi.org/10.1201/9781420051438-10>
- Harding, D. (2008). Pulsed Laser Altimeter Ranging Techniques and Implications for Terrain Mapping. Teoksessa J. Shan & C. K. Toth (Toim.), *Topographic Laser Ranging and Scanning* (1. p., s. 173–194). CRC Press. <https://doi.org/10.1201/9781420051438-5>

- Heinzel, J., & Koch, B. (2011). Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 152–160. <https://doi.org/10.1016/j.jag.2010.09.010>
- Heinzel, J., & Koch, B. (2012). Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. *International Journal of Applied Earth Observation and Geoinformation*, *18*, 101–110. <https://doi.org/10.1016/j.jag.2012.01.025>
- Hyypä, J., Kelle, O., Lehtikainen, M., & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(5), 969–975. <https://doi.org/10.1109/36.921414>
- Hyypä, J., Hyypä, H., Yu, X., Kaartinen, H., Kukko, A., & Holopainen, M. (2008). Forest Inventory Using Small-Footprint Airborne LiDAR. Teoksessa J. Shan & C. K. Toth (Toim.), *Topographic Laser Ranging and Scanning* (1. p., s. 335–370). CRC Press. <https://doi.org/10.1201/9781420051438-12>
- Imangholiloo, M., Saarinen, N., Holopainen, M., Yu, X., Hyypä, J., & Vastaranta, M. (2020). Using Leaf-Off and Leaf-On Multispectral Airborne Laser Scanning Data to Characterize Seedling Stands. *Remote Sensing*, *12*(20), 3328. <https://doi.org/10.3390/rs12203328>
- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, *4*(9), 2661–2693. <https://doi.org/10.3390/rs4092661>
- Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sensing of Environment*, *130*, 245–253. <https://doi.org/10.1016/j.rse.2012.11.024>
- Kim, S., McGaughey, R. J., Andersen, H.-E., & Schreuder, G. (2009). Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, *113*(8), 1575–1586. <https://doi.org/10.1016/j.rse.2009.03.017>
- Korpela, I., Ørka, H., Maltamo, M., Tokola, T., & Hyypä, J. (2010). Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica*, *44*(2). <https://doi.org/10.14214/sf.156>

- Li, J., Hu, B., & Noland, T. L. (2013). Classification of tree species based on structural features derived from high density LiDAR data. *Agricultural and Forest Meteorology*, 171–172, 104–114. <https://doi.org/10.1016/j.agrformet.2012.11.012>
- Li, J., Hu, B., Sohn, G., & Jing, L. (2010). Individual tree species classification using structure features from high density airborne lidar data. *2010 IEEE International Geoscience and Remote Sensing Symposium*, 2099–2102. <https://doi.org/10.1109/IGARSS.2010.5651629>
- Lin, Y., & Hyypä, J. (2016). A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 46, 45–55. <https://doi.org/10.1016/j.jag.2015.11.010>
- Liu, B., Chen, S., Huang, H., & Tian, X. (2022). Tree Species Classification of Backpack Laser Scanning Data Using the PointNet++ Point Cloud Deep Learning Method. *Remote Sensing*, 14(15), 3809. <https://doi.org/10.3390/rs14153809>
- Liu, M., Han, Z., Chen, Y., Liu, Z., & Han, Y. (2021). Tree species classification of LiDAR data based on 3D deep learning. *Measurement*, 177, 109301. <https://doi.org/10.1016/j.measurement.2021.109301>
- Maanmittauslaitos. (2023). Laserkeilausaineisto 5 p [paikkatietoaineisto]. Avoin data. <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/laserkeilausaineisto-5-p>
- Marrs, J., & Ni-Meister, W. (2019). Machine Learning Techniques for Tree Species Classification Using Co-Registered LiDAR and Hyperspectral Data. *Remote Sensing*, 11(7), 819. <https://doi.org/10.3390/rs11070819>
- Michałowska, M., & Rapiński, J. (2021). A Review of Tree Species Classification Based on Airborne LiDAR Data and Applied Classifiers. *Remote Sensing*, 13(3), 353. <https://doi.org/10.3390/rs13030353>
- Mielcarek, M., Stereńczak, K., & Khosravipour, A. (2018). Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation. *International journal of applied earth observation and geoinformation*, 71, 132–143.
- Mustonen, S. (1995). *Tilastolliset monimuuttujamenetelmät*. Survo Systems Helsinki. <http://survo.fi/mustonen/monim.pdf>
- Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., & Vihervaara, P. (2021). Tree species classification from airborne hyperspectral and LiDAR data using 3D

- convolutional neural networks. *Remote Sensing of Environment*, 256, 112322. <https://doi.org/10.1016/j.rse.2021.112322>
- Nevalainen, P., Li, Q., Melkas, T., Riekkki, K., Westerlund, T., & Heikkonen, J. (2020). Navigation and Mapping in Forest Environment Using Sparse Point Clouds. *Remote Sensing*, 12(24), 4088. <https://doi.org/10.3390/rs12244088>
- Petrie, G., & Toth, C. K. (2008). Introduction to Laser Ranging, Profiling, and Scanning. Teoksessa J. Shan & C. K. Toth (Toim.), *Topographic Laser Ranging and Scanning* (1. p., s. 1–28). CRC Press. <https://doi.org/10.1201/9781420051438-1>
- Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11), 1119–1125. [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
- Reisi Gahrouei, O., Côté, J.-F., Bournival, P., Giguère, P., & Béland, M. (2024). Comparison of Deep and Machine Learning Approaches for Quebec Tree Species Classification Using a Combination of Multispectral and LiDAR Data. *Canadian Journal of Remote Sensing*, 50(1), 2359433. <https://doi.org/10.1080/07038992.2024.2359433>
- Ruotsalainen, S., Himanen, K., Viherä-Aarnio, A., Aarnio, L., Haapanen, M., Luoranen, J., Matala, J., Riikonen, J., Uotila, K., & Ylioja, T. (2022). *Puulajivalikoiman monipuolistaminen metsänviljelyssä: Synteesiraportti*. <https://jukuri.luke.fi/handle/10024/551684>
- Sharma, M., & Parton, J. (2007). Height–diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management*, 249(3), 187–198. <https://doi.org/10.1016/j.foreco.2007.05.006>
- Shi, Y., Wang, T., Skidmore, A. K., & Heurich, M. (2018). Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137, 163–174. <https://doi.org/10.1016/j.isprsjprs.2018.02.002>
- Suratno, A., Seielstad, C., & Queen, L. (2009). Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 683–693. <https://doi.org/10.1016/j.isprsjprs.2009.07.001>
- Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 30(2), 169–190. <https://doi.org/10.3233/AIC-170729>
- Torabzadeh, H., Leiterer, R., Hueni, A., Schaepman, M. E., & Morsdorf, F. (2019). Tree species classification in a temperate mixed forest using a combination of imaging

- spectroscopy and airborne laser scanning. *Agricultural and Forest Meteorology*, 279, 107744. <https://doi.org/10.1016/j.agrformet.2019.107744>
- Turun kaupunki. (2021). 3D-laserkeilausaineisto [Paikkatietoaineisto]. Avoin data. <https://www.avoindata.fi/data/fi/dataset/turun-kaupungin-kaupunkitietomalli/resource/15932cbe-13eb-44b7-88b0-bcb858a5d111>
- Turun kaupunki. (2023). Turun seudun ilmakehä 2023 [Paikkatietoaineisto]. Avoin data. <https://www.avoindata.fi/data/fi/dataset/turun-seudun-ilmakeha-2023>
- Wang, L., Lu, D., Xu, L., Robinson, D. T., Tan, W., Xie, Q., Guan, H., Chapman, M. A., & Li, J. (2024). Individual tree species classification using low-density airborne multispectral LiDAR data via attribute-aware cross-branch transformer. *Remote Sensing of Environment*, 315, 114456. <https://doi.org/10.1016/j.rse.2024.114456>
- Yu, X., Hyypä, J., Litkey, P., Kaartinen, H., Vastaranta, M., & Holopainen, M. (2017). Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning. *Remote Sensing*, 9(2), 108. <https://doi.org/10.3390/rs9020108>
- Yu, X., Litkey, P., Hyypä, J., Holopainen, M., & Vastaranta, M. (2014). Assessment of low density full-waveform airborne laser scanning for individual tree detection and tree species classification. *Forests*, 5(5), 1011–1031.
- Zhang, Z., Kazakova, A., Moskal, L., & Styers, D. (2016). Object-Based Tree Species Classification in Urban Ecosystems Using LiDAR and Hyperspectral Data. *Forests*, 7(6), 122. <https://doi.org/10.3390/f7060122>
- Zhong, L., Dai, Z., Fang, P., Cao, Y., & Wang, L. (2024). A Review: Tree Species Classification Based on Remote Sensing Data and Classic Deep Learning-Based Methods. *Forests*, 15(5), 852. <https://doi.org/10.3390/f15050852>
- Ørka, H. O., Dalponte, M., Gobakken, T., Næsset, E., & Ene, L. T. (2013). Characterizing forest species composition using multiple remote sensing data sources and inventory approaches. *Scandinavian Journal of Forest Research*, 28(7), 677–688. <https://doi.org/10.1080/02827581.2013.793386>