

KONEOPPIMISTEKNOLOGIAN HYÖDYNTÄMINEN MARJASATOJEN KARTOITTAMISESSA

Marjamasiina-hanke

Tiirola Josefi

Opinnäytetyö

Maaseutuelinkeinot
Agrologi (AMK)

2022

Maaseutuelinkeinot
Agrologi (AMK)

Tekijä	Josefi Tirola	Vuosi	2022
Ohjaaja	Kirsi Muuttonen		
Toimeksiantaja	Lapin ammattikorkeakoulu		
Työn nimi	Koneoppimisteknologian hyödyntäminen marjasatojen kartoittamisessa		
Sivu- ja liitesivumäärä	37		

Opinnäytetyön tavoitteena on selvittää, miten koneoppimisteknologiaa voidaan hyödyntää marjojen kartoittamisessa ja siten marjasatoennusteiden tuottamisessa. Opinnäytetyön tutkimuskysymys keskittyy kuvantunnistusteknologian peittymisongelman ratkaisemiseen, minkä avulla selvitetään, voidaanko kuvista laskettavalla marjojen määrällä päätellä marjojen tosiasiallinen määrä neliömetrin pinta-alalla.

Opinnäytetyö perustuu Lapin ammattikorkeakoulun Marjamasiina-hankkeeseen, jonka tavoitteena on ollut kehittää toimiva sovellus marjojen kartoittamiseksi. Hankkeessa on kehitetty koneoppimisalgoritmi paikantamaan ja luokittelemaan eri marjoja sille syötetystä kuva-aineistosta. Opinnäytetyössä hyödynnettävä aineisto on kerätty osana Marjamasiina-hanketta.

Tutkimuskysymys käsittelee koneoppimisalgoritmin tuottaman laskelman ja marjojen todellisen määrän korrelaatiota rajatulla alueella. Koneoppimisalgoritmi laskee marjojen määrän koealalta otetun kuvan perusteella, kun taas marjojen todellinen määrä on laskettu koealalta manuaalisesti käsin. Opinnäytetyö keskittyy mustikan ja puolukan tunnistamiseen ja laskemiseen.

Opinnäytetyön tuloksena osoitetaan korrelaatio, jonka avulla voidaan kehittää laskentamenetelmiä marjasatojen kartoittamiseksi hyödyntäen koneoppimisteknologiaa. Korrelaatioissa esiintyvää hajontaa selitetään erilaisilla marjoihin ja luontoon liittyvillä ominaisuuksilla, joiden ymmärrys on merkittävässä asemassa koneoppimisen pullonkaulaongelmia ratkaistaessa.

Opinnäytetyössä tuodaan esille tekijät, jotka vaikuttavat marjojen kartoittamiseen koneoppimisteknologian avulla. Näitä tekijöitä ovat esimerkiksi marjakasvien rakenteelliset erot, kuvatekniikan aiheuttama vaihtelu ja näköislajit.

Rural Industries
Agronomist

Author	Josefi Tiirola	Year	2022
Supervisor	Kirsi Muuttoranta		
Commissioned by	Lapland University of Applied Sciences		
Subject of thesis	Machine learning in inventory of berry yields		
Number of pages	37		

The target of this thesis is to consider how machine learning technology can be used in inventory of berries and thus in developing berry yield forecasts. The research question of the thesis focuses on solving the occlusion problem of object recognition technology, which is used to find out whether the actual number of berries per square meter can be deduced from the number of berries calculated from the images.

This thesis is based on the Berrymachine project in Lapland University of Applied Sciences, which goal has been to develop a functional application for inventory of berries. In the project, a machine learning algorithm has been developed to locate and classify different berries from the image data which have been fed to it. The material used in this thesis has been collected as part of the Berrymachine project.

The research question considers the correlation between calculation produced by the machine learning algorithm and the actual number of berries in a limited area. The machine learning algorithm calculates the number of berries based on the image taken from the test area, while the actual number of berries has been counted manually by hand from the test area. Thesis focuses on the recognition and counting of bilberries and lingonberries.

As a result of the thesis, a correlation is demonstrated, which can be used to develop calculation methods for inventory of berry yields using machine learning technology. The dispersion in the correlation is explained by various characteristics related to berries and nature, the understanding of which plays a significant role in solving bottleneck problems in machine learning.

In this thesis, the factors that influence the inventory of berries with the help of machine learning technology are brought to the fore. These factors include, for example, the structural differences of berry plants, the variation caused by image technology, and species of appearance.

Keywords Algorithm, berries, machine learning, occlusion problem

SISÄLLYS

1 JOHDANTO	5
2 MARJASATOJEN KARTOITTAMINEN	7
2.1 Mustikka ja puolukka.....	7
2.2 Marjojen keruu ja käyttö.....	7
2.3 Marjasatoennusteet	8
2.4 Marjat osana ekosysteemipalveluita ja kulttuuria.....	10
3 KONEOPPIMINEN JA KONENÄKÖ	11
3.1 Neuroverkko ja sen toiminta	11
3.2 Syväoppiminen ja neuroverkot.....	12
3.3 You Only Look Once -algoritmi	14
3.4 Peittymisongelma	15
3.5 Kuvantunnistuksen ja konenäönmerkitys.....	17
4 AINEISTO JA MENETELMÄT	18
4.1 Aineiston keruu	18
4.2 Aineiston käsittely koneoppimisalgoritmile	19
4.3 Marjat kartoittamisen ja käsittelyn kohteena	20
4.4 Näköislajit	22
4.5 Tutkimuskysymys	24
4.6 Tutkimusmenetelmä.....	25
5 TULOKSET JA JOHTOPÄÄTÖKSET	26
5.1 Korrelaatio laajalla aineistolla	26
5.2 Korrelaatio laatuaineistolla.....	28
5.2.1 Algoritmin koulutus.....	29
5.2.2 Marjojen ja niiden kasvuvaiheiden erot	30
6 POHDINTA	31
6.1 Hyödynnettävyys	31
6.2 Hyödynnettävyyden haasteet.....	32
6.3 Haasteiden pohdinta.....	33
6.4 Menetelmän soveltaminen ja jatkotutkimus	34
LÄHTEET	35

1 JOHDANTO

Marjat, marjojen poiminta ja marjojen hyödyntäminen ovat keskeinen osa suomalaista kulttuuria. Suomalais-ugrilaisessa kansanperinnössä puolukka on jopa yhdistetty naisen hedelmällisyyteen, mikä tulee esille Kalevala-teoksen 50. runossa. Loppukesän ja alkusyksyn keskusteluilmapiiriä leimaa monipuolinen keskustelu keräilytuotteiden sadosta ja sääolosuhteiden vaikutuksesta niiden muodostumiseen. Luonnonvarakeskus tuottaa vuosittain marjasatoennusteita vastaamaan suomalaisten mielenkiintoa marjasatojen muodostumista kohtaan (Luonnonvarakeskus 2022a).

Lapin ammattikorkeakoulun Marjamasiina-hankkeessa keskeinen tarkoitus on ollut automatisoida marjojen laskemista, sillä se on keskeinen tekijä marjasatoennusteiden tuottamisessa. Kuvien olennaisin ongelma on kuitenkin marjakasvuston peittävydessä, jossa osa marjoista jää varpukasvien taakse. Kuvista voidaan tosiasiallisesti laskea vain ne marjat, jotka eivät ole peittyneenä. Hankkeessa aineistoa on kerätty ja käsitelty mustikasta ja puolukasta. Automatisoitu laskenta toteutettaisiin mobiilikäyttöliittymän avulla.

Opinnäytetyössä hyödynnettävä lähdeaineisto painottuu yleisten koneoppimisteknologioiden menetelmien varaan. Globaalisti luonnonvarojen automatisoitu kartoitus on keskittynyt pääasiassa maa- ja metsätalouden laajoihin kokonaisuuksiin ja esimerkiksi hedelmätarhaamisen tehostamiseen. Koneoppimiseen ja mobiililaitteisiin liittyvää teknologiaa tutkitaan laajalti myös kaupallisesti, minkä vuoksi osa lähdeaineistosta on kaupallisilta toimijoilta ja teknologia-alan yrityksiltä.

Opinnäytetyön tutkimuskysymys käsittelee algoritmin tuottaman pintalaskentatuloksen ja marjojen todellisen määrän korrelaatiota. Jotta voitaisiin laskea marjojen määrä määrättyltä pinta-alalta hyödyntämällä koneoppimisteknologiaa ja konenäköä, kuvasta laskettavan määrän tulee korreloida myös todellisen marjojen määrän kanssa.

Tutkimuskysymyksen on tarkoitus myös havainnoida millä tavoin voidaan kehittää erilaisia menetelmiä, joiden avulla luonnonvaroja voidaan kartoittaa koneäön ja koneoppimisteknologian avulla. Kenttäkerroksessa kasvavien marjojen lajityypillinen variaatio havainnollistaa tutkimuksen sovellettavuutta monipuolisesti erilaisiin toimintaympäristöihin, sillä esimerkiksi mustikan ja puolukan välillä on huomattavia rakenteellisia eroja.

Opinnäytetyössä hyödynnettävä menetelmä on yksinkertainen, mutta ongelmakokonaisuuden hahmottaminen vaatii monipuolista ymmärrystä, miten automatisointi vaikuttaa luonnonvarojen kartoittamiseen ja minkälaisia pullonkauloja siinä esiintyy. Pullonkaulojen löytäminen vaatii asiantuntijuuden alkuperäisen menetelmän suhteesta automatisoituun menetelmään ja siihen, mitkä ovat niiden eksplisiittiset erot sekä vaikutus lopputulokseen. Esimerkiksi marjoja kartoittaessa kamera ei kykene siirtelemään varpukasveja, kuten ihminen. Tällöin kamera saa tallennettua vain osan tiedosta, kun ihminen voi toiminnallaan kerätä tietoa laajemmin.

Tutkimuskysymystä tarkastellaan kahdella eri aineistolla, joista toinen on laaja ja toinen suppeampi laatuaineisto. Kahdella eri aineistolla pyritään todentamaan tuloksen toistettavuus, sillä laajemmassa aineistossa on enemmän kuvaustekniikasta johtuvaa kohinaa. Suppeammassa aineistossa käsitellään kaksi eri aineistoa, jotka on rajattu saman laitteen, marjan ja kehitysvaiheen mukaisesti. Suppeamman aineiston tarkoitus on vähentää laitteesta ja eri kehitysvaiheista johtuvaa hajontaa. Aineisto on jaettu mustikan ja puolukan kolmeen eri kehitysvaiheeseen: kukkaan, raakileeseen ja marjaan.

Opinnäytetyön tavoitteena on selvittää, onko koneoppimisalgoritmin laskelmien marjojen ja marjojen todellisen määrän välillä yhteys. Opinnäytetyössä analysoidaan mitkä tekijät vaikuttavat tulokseen ja sen paikkansapitävyyteen. Jos pintalaskentatulokset korreloi marjojen todellisten määrän kanssa, voidaan marjojen kokonaismäärä päätellä kuvan avulla. Jos korrelaatiota ei voida todeta, kuvista ei voida laskea marjojen todellista määrää määrättyltä pinta-alalta.

2 MARJASATOJEN KARTOITTAMINEN

2.1 Mustikka ja puolukka

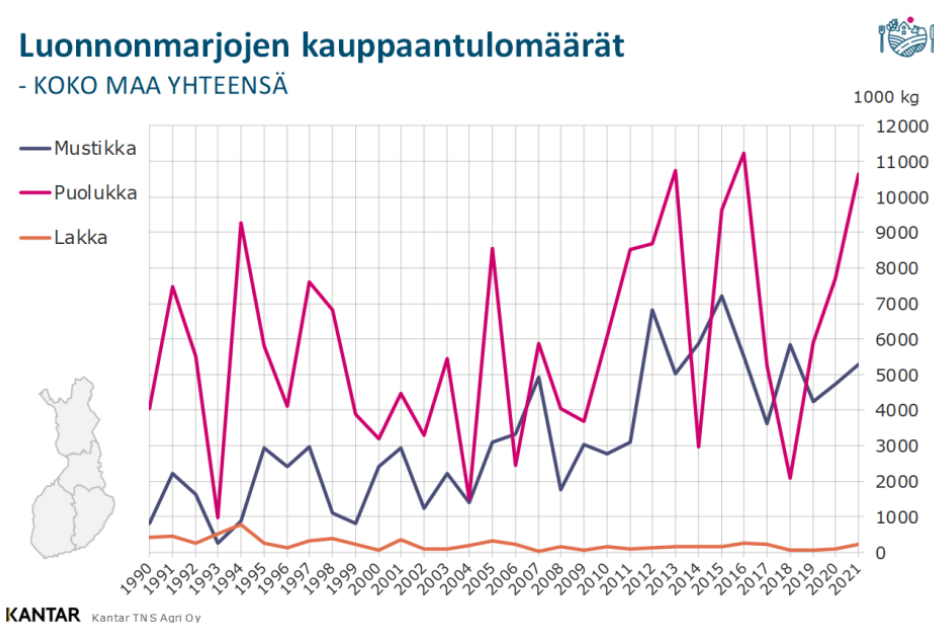
Aineistoa on kerätty mustikasta ja puolukasta, jotka ovat molemmat merkittäviä keräilykasveja. Mustikka on tuoreissa metsissä kasvava varpukasvi, joka aloittaa kukinnan toukokuussa ja kypsyttää sadon alkusyksystä (Ihamuotila 2011, 23). Mustikka kasvaa koko Suomessa sekä on monivuotinen (Arktiset aromit 2022a). Onnistunut ja runsas mustikkasato vaatii sopivat sääolosuhteet sekä onnistuneen hyönteispölytyksen, sillä mustikka on ristipölytteinen (Luonnonvarakeskus & 4H 2022a). Esimerkiksi edellisen syksyn lämpötila vaikuttaa kukka-aiheiden muodostumiseen ja talven lumikerros kasvin selviytymiseen kylmissä lämpötiloissa (Marjala 2015).

Puolukka on mäntymetsissä kasvava varpukasvi, jonka kukinta alkaa kesäkuun alussa ja sato kypsyy syksyllä (Ihamuotila 2011, 12). Puolukka kasvaa mustikan lailla koko Suomessa ja on yhtä lailla monivuotinen. Puolukan tunnusmerkistönä on sen vahapintaiset lehdet. (Arktiset aromit 2022b.) Puolukka kasvaa mustikkaa kuivemmassa ympäristössä, eikä ole yhtä herkkä säästä johtuville satovaihteluille. Puolukka kestää esimerkiksi kesän korkeita lämpötiloja hyvin (STT 2022). Lisäksi puolukalla myös itsepölytys on mahdollista, mutta hyötyy myöhäisen kukintavaiheen vuoksi myös hyönteispölytyksestä (Luonnonvarakeskus & 4H 2022b).

2.2 Marjojen keruu ja käyttö

Suomessa on aloitettu marjasatojen järjestelmällinen kartoittaminen vuonna 1997. Vuoteen 2014 asti kartoittamista hoiti Metsäntutkimuslaitos, joka yhdistyi vuonna 2015 Luonnonvarakeskukseen. (Salo 2015, 128.) Marjasatojen kartoituksen tarkoituksena on tuottaa satoennusteet, jotka palvelevat kansalaisia, marjanostajia ja jalostajia. Luonnonvarakeskus ylläpitää Marjahavainnot.fi-verkkosivua, jossa tiedotetaan 3–4 kertaa kesän aikana satoennusteista Luonnonvarakeskuksen omien verkkosivujen lisäksi. (Luonnonvarakeskus 2022a.)

Marjoja kerätään kotitalouksiin yksityiskäyttöön sekä kaupalliseen myyntiin. Vuonna 2021 suomalaisia luonnonmarjoja kerättiin ainakin miljoonia kiloja pelkästään kaupalliseen myyntiin. Mustikan kauppantulomäärä oli tuolloin noin 5,3 miljoonaa kiloa ja puolukan noin 10,7 miljoonaa kiloa. Luonnonmarjojen kauppantulomäärä vaihtelee vuosittain, mutta etenkin mustikan määrä on kasvanut 2010-luvulla (Kuvio 1). Vuonna 2021 mustikan poimintatulot olivat noin 9,7 miljoonaa euroa ja puolukan 9,4 miljoonaa euroa. (Ruokavirasto 2022, 13, 17.)



Kuvio 1. Luonnonmarjojen kauppantulomäärät 1990–2021 (Ruokavirasto 2022, 16.)

2.3 Marjasatoennusteet

Marjasatoennusteita ja -seurantaa tekee Suomessa Luonnonvarakeskus, joka on perustanut verkoston seuraamaan marjojen kypsyystilannetta eri puolella Suomea. Verkoston keskeisinä toimijoina ovat 4H-nuoret, jotka tekevät 4H-yhdistysten kautta tiivistä yhteistyötä Luonnonvarakeskuksen kanssa. (Luonnonvarakeskus 2022a.) Marjasatoennusteita toimitetaan mustikasta, puolukasta ja lakasta. Ennusteet perustuvat havaintometsiin ja koeruutuihin, jotka inventoidaan kasvukauden aikana Luonnonvarakeskuksen ohjeiden mukaisesti. (Luonnonvarakeskus 2022b, 3–5.)

Marjoja kartoittaessa perustetaan 1 x 1 metrin koeruutu, jonka alan marjat laskeetaan käsin. Koeruudut rakennetaan pysyviksi, sillä havainnot kirjataan samoista koeruuduista koko kasvukauden ajan. Koeruutujen tulee sijaita yhtenäisessä metsäkuviossa ja sisältää inventoitavaa marjakasvustoa. Luonnonvarakeskus ottaa vastaan tiedon koeruutukohtaisesti marjojen määrästä eri vaiheissa. Tiedot tulisi toimittaa marjojen ollessa kukinnassa, raakileena ja kypsänä Marjahavainnot.fi -palveluun (Luonnonvarakeskus 2022b, 3–5, 9.)

Marjasatoennusteet julkaistaan Marjahavainnot.fi -palvelussa ja Luonnonvarakeskuksen internetsivuilla. Marjahavainnot.fi -palvelussa on mahdollista nähdä marjasatoennusteet kartalla, jossa kuvataan marja eri kasvuvaiheessa ja esiintymän runsaus värein (Kuvio 2). Luonnonvarakeskuksen tiedotteissa kerrotaan yksityiskohtia kuluneen kesän marjasadon tilanteesta ja odotuksista. (Luonnonvarakeskus & 4H 2022c.)



Kuvio 2. Marjasatoennustekartta (Luonnonvarakeskus & 4H 2022d.)

2.4 Marjat osana ekosysteemipalveluita ja kulttuuria

Marjat ovat osa metsäekosysteemiä sekä ne luokitellaan osaksi metsän tuotantopalveluita (MTK 2017). Mitattavan taloudellisen arvon lisäksi marjoilla nähdään myös esimerkiksi sosiaalista ja terveydellistä arvoa. Marjat myös nähdään osaksi suomalaista ruoka- ja kansankulttuuria. (Vihervaara, Ahokumpu, Auvinen & Mononen 2017.)

Suomalaista luonto- ja marjakulttuuria leimaa jokamiehenoikeus, mikä takaa mahdollisuuden jokaiselle kerätä marjoja sellaisilta metsäalueilta, joilla voi myös vapaasti liikkua (Ympäristöministeriö 2021). Lapissa ja Kainuussa kotitalouksien marjanpoimintaa harjoitetaan alueellisesti eniten ja marjoja kerätäänkin noin 40 litraa kotitaloutta kohden, kun eteläisessä Suomessa määrä on vain noin kymmenen litraa kotitaloutta kohden (Kivelä 2021). Marjankeruun kulttuurinen merkitys on Suomessa erityinen, sillä lähes puolet suomalaisista aikuisista harjoittaa marjankeruuta jossain muodossa. Marjankeruu itsessään nähdään hyvinvointia lisäävänä toimintana, ja se on osa suomalaista metsäsuhdetta. Lisäksi marjankeruu voi olla sivutulonlähde ja se nähdään osaksi alkutuotantoa. (Hyvärinen 2021.)

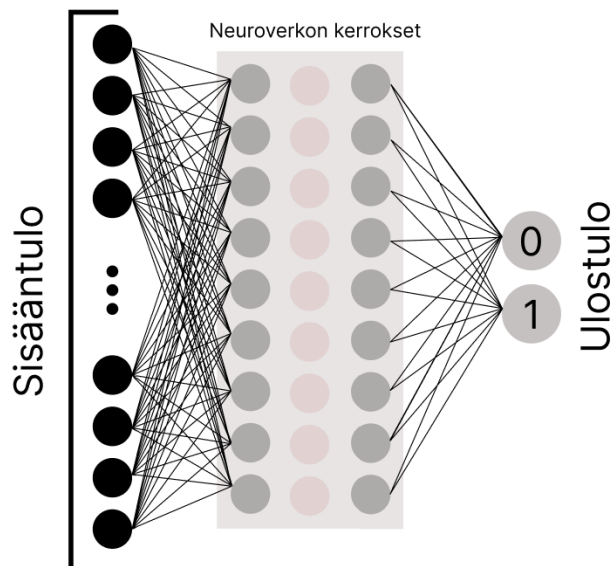
Marjojen asema huomioidaan myös metsätaloudessa entistä näkyvämmiin ja voimakkaammiin. Metsänomistajat kokevat marjat myös osaksi metsätalouden tuottamia hyötyjä, minkä vuoksi metsäsuunnittelussa voidaan ottaa huomioon metsäympäristössä kasvavat erilaiset marjat ja niiden kasvupaikat. Paikallisesti marjoihin ja niistä muodostuviin satoihin vaikuttavat yleistekijöiden lisäksi myös esimerkiksi metsänkehitysvaihe ja erilaiset metsätaloudelliset toimenpiteet. (Salo 2015, 143–145.)

Marjasatoennusteiden merkitys voidaan siis nähdä laaja-alaisesti eri väestöryhmissä ja niiden eri segmenteissä. Marjasatoennusteet myös kertovat muiden ekosysteemipalveluiden tilasta ja toimivuudesta metsäympäristössä, sillä esimerkiksi onnistunut pölytys ja kuluneen vuoden sää ovat avainasemassa marjasatojen onnistumiselle (Salo 2015, 129).

3 KONEOPPIMINEN JA KONENÄKÖ

3.1 Neuroverkko ja sen toiminta

Koneoppiminen on yksi tekoälyn alalaji, jossa tietokone koulutetaan suorittamaan ennalta määritellyt prosesseja. Näitä ennalta määriteltyjä prosesseja kutsutaan algoritmeiksi, joiden tarkoituksena on ratkaista jokin ongelma tai suorittaa tehtävä. Koneoppiminen perustuu synteettisesti luotuun neuroverkkoon, jonka toiminta perustuu hakuprosessiin. Neuroverkko pyrkii päättämään vaihtoehtoista parhaan tarkastelemalla syötetyn aineiston eri ominaisuuksien suhteita (Kuvio 3). Neuroverkon toiminta perustuu sen kouluttamiseen valmiiksi käsitellyllä aineistolla. (Kelleher 2020, 15–16, 19.)



Kuvio 3. Neuroverkon rakenne

Konenäöllä tarkoitetaan prosessia, jossa kuva-aineistoa pyritään hyödyntämään erilaisten näköön perustuvien prosessien suorittamisessa, esimerkiksi kasvontunnistuksessa. Konenäkö on runsaasti tutkittu syväoppimisen alue, sillä konenäön on vaikea saavuttaa samanlaista kyvykkyyttä kuin ihmisen tai eläimen luontainen näkökyky. (Bengio, Courville & Goodfellow 2016, 440.) Hankkeessa hyödynnetään You Only Look Once -algoritmia, jonka toiminta periaate nojaa kohteentunnistukseen (Divvala, Farhadi, Girshick & Redmon 2016, 1).

3.2 Syväoppiminen ja neuroverkot

Koneoppiminen perustuu kouluttamiseen ja päättelyyn. Koneoppimisessa on kyse neuroverkosta, joka pyrkii etsimään parhaimman vaihtoehdon useiden joukosta. Koneoppimista hyödynnetään erityisesti silloin kun algoritmin luominen ihmistyöllä olisi liian monimutkaista ja ajallisesti haastavaa. (Bengio ym. 2016, 96.) Kun useita eri neuronikerroksia yhdistetään verkoksi, kutsutaan teknologiaa syväoppimiseksi. Syväoppiminen mahdollistaa tietokoneelle itsensä kouluttamisen aineiston perusteella, jolloin aineisto iteroi käytettävissä olevaa neuroverkkoa, jolloin algoritmi koulutautuu. (Kelleher 2020, 19–22, 64.)

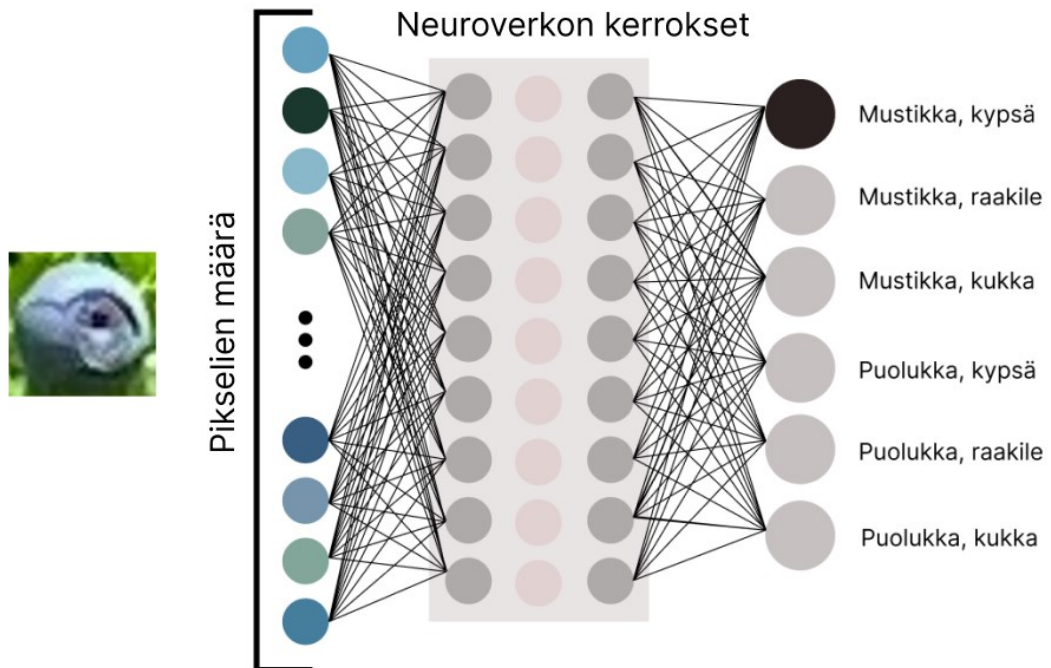
Neuroverkkojen toimintaperiaate mukaillee ihmisarvojen rakennetta ja on siten tehokas laskennallinen malli. Neuroverkkojen toiminta perustuu neuroneihin, jotka lähettävät tietoa eteenpäin sisääntulosignaalin ollessa tarpeeksi vahva. (Kelleher 2020, 64–66.)

Neuroverkon koulutuksessa tulee ottaa huomioon mahdollisimman laaja aineisto, sillä se vähentää virhetulkintojen määrää. Laajempi aineisto kouluttaa neuroverkon rakennetta monipuolisemmaksi, jolloin neuroverkko ei tee yksinkertaisettuja päätelmiä. Koulutuksessa tulee ottaa huomioon neuroverkolle asetettu tehtävä ja tehtävää tulee arvioida tarkasti, millä voidaan rajata halutut piirteet ja määrittellä niiden merkitsevyys. (Kelleher 2020, 25, 30–31.)

Marjamasiina-hankkeessa koneoppimisalgoritmia on koulutettu monipuolisella kuva-aineistolla, jossa marjoja on kuvattu esimerkiksi mahdollisimman monipuolisista kuvakulmista. Koneoppimisteknologiaa hyödyntämällä pyritään tilanteeseen, jossa käyttäjät keräävät kuva-aineistoa ja siitä voidaan nopeasti sekä luotettavasti analysoida marjojen määrä ennalta määritetyllä pinta-alalla (Kelleher 2020, 21). Kun neuroverkon koulutus on valmis, voidaan kouluttaminen keskeyttää ja jättää neuroverkon rakenne pysyväksi. Tämän jälkeen neuroverkkoa voidaan hyödyntää tehtävän suorittamisessa. (Kelleher 2020, 21–22.)

Neuroverkon koulutuksen päättämisen ajoittaminen on olennaista koulutuksen onnistumisen kannalta, sillä tunnistuksia suorittava neuroverkko voi myös ylikouluttautua. Ylikouluttautuminen näkyy väärin tunnistuksien kasvavana määränä. Ylikouluttautuminen käytännössä luo neuroverkon, joka löytää liikaa kohteita sille annetusta aineistosta ja nämä tunnistukset ovat virheellisiä. (Bengio ym. 2016, 107–108.)

Marjamasina-hankkeessa 640 x 640 pikselin kokoisia kuvia lähetetään neuroverkkoon, joka on koulutettu ihmisten käsittelemällä aineistolla tunnistamaan erilaisia marjoja ja niiden kasvuvaiheita. Neuroverkko antaa tuloksen syötteestä perustuen pikselien sisältöön ja niiden välisiin suhteisiin (Kuvio 4).

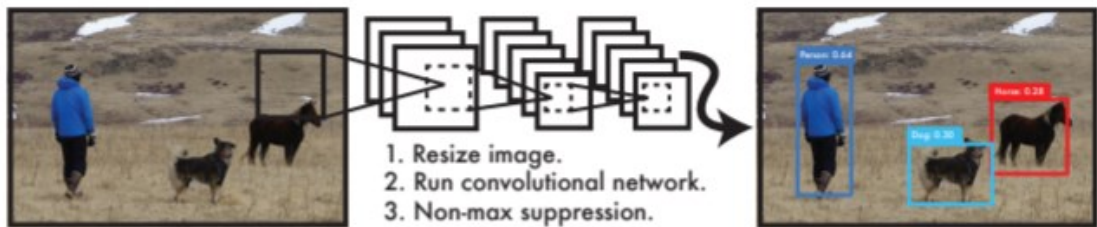


Kuvio 4. Marjoja käsittelevän neuroverkon yksinkertaistus

Neuroverkkoon lähetetyn kuvan pikselit saavat numeraalisen arvon, joka perustuu kuvan pikselien väriin. Tässä vaiheessa kuvan sisältämää informaatiota suodatetaan neuroverkolle käyttökelpoiseksi, sillä neuroverkolla ei luonnollista tapaa erotella eri värejä tai värien eri asteita ihmisaivojen tavoin. (Kananen & Puolitaival 2019, 81–82.)

3.3 You Only Look Once -algoritmi

You Only Look Once -algoritmi eli YOLO on kuvantunnistukseen kehitetty tehokas ja jatkuvasti kehittyvä koneoppimisteknologia. YOLO perustuu kohteen paikantamiseen ja luokitteluun kuva-aineistosta. YOLO on nopea sekä sen teho perustuu sen tapaan käsitellä kuva-aineistoa skaalaamalla kuva-aineistoa pienempiin osiin ennen sen käsittelyä (Kuvio 5). YOLO on myös luotettava kuva-aineiston skaalauksen vuoksi, sillä se ottaa hyvin huomioon kuva-aineistossa esiintyvät taustatekijät. (Divvala ym. 2016, 1–2.)



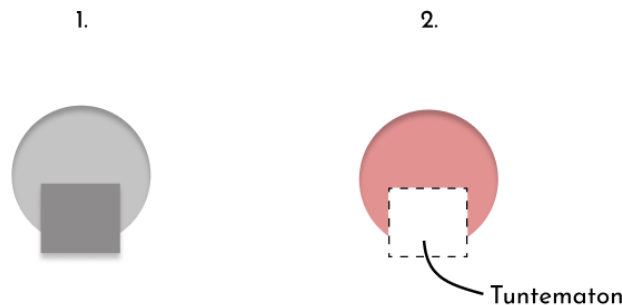
Kuvio 5. YOLO:n kuva-aineiston skaalaus (Divvala ym. 2016, 1.)

YOLO perustuu konvoluutioneuroverkkoon (Divvala ym. 2016, 2). Konvoluutioneuroverkot ovat erityisesti kuvantunnistuksessa hyödynnetty teknologia. Konvoluutioneuroverkko koostuu useista neuroverkkojen kerroksista, joiden tehtävänä on käsitellä kuvan piirteitä aluksi pienemmissä yksiköissä, minkä jälkeen yhdistelemällä aikaisempia löydöksiä kootaan suuremmaksi kuvaksi neuroverkon myöhempien kerroksien analysoitavaksi. (Kelleher 2020, 145.)

Konvoluutioneuroverkot perustuvat teknologiaan, jonka tavoitteena on jäljitellä silmien ja aivojen tapaa käsitellä visuaalista materiaalia (Kananen & Puolitaival 2019, 149). Konvoluutioneuroverkot ovatkin poikkitieteellisen tutkimustyön tulos, jossa on pyritty yhdistämään ymmärrystä neurotieteestä ja tietotekniikasta. Konvoluutioneuroverkon tapaa käsitellä tietoa kutsutaan myös syväoppimiseksi. (Mitchell 2019, 70–71.)

3.4 Peittymisongelma

Vaikka koneoppimisteknologialla onkin mahdollista käsitellä kuva-aineistoa menestyneesti, konenäkö aiheuttaa perustavanlaatuisia ongelmia todellisen maailman ja aineiston välisessä suhteessa. Käytettävä aineiston sisältämä informaatio on 2D-kuvan muodossa, minkä vuoksi myös koneoppimisalgoritmi tulkitsee sitä vain ja ainoastaan 2D-kuvana. Konenäkö ei siis näe tai tulkitse objektien taakse, jolloin suoritukseen jää tuntematon alue (Kuvio 6). Jos tuntematon alue peittää vain osan objektista, voidaan päätellä tuntemattoman alueen sisältö.



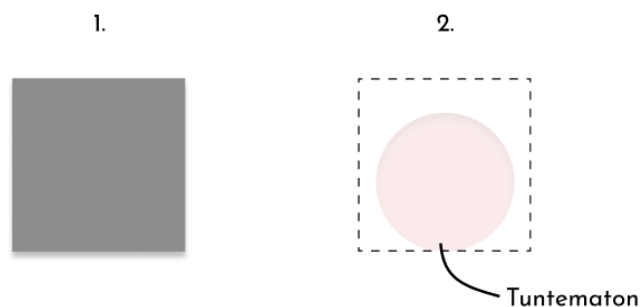
Kuvio 6. Tuntematon alue syntyy edellä olevan objektin taakse

Peittymisongelman filosofia perustuu ihmisen ja koneoppimisalgoritmin erilaiseen tapaan tulkita kuvainformaatiota. Ihmisillä on kyky löytää kuvasta olennainen sekä liittää kuvamateriaalia jonkinlaiseen kontekstiin, mikä helpottaa materiaalin tulkintaa. Koneoppimisalgoritmeilla vastaavaa kykyä ei ole, vaan se tulkitsee kuvainformaatiota mekaanisesti ja tarkkailemalla pikselien eri riippuvuussuhteita. Vieläkään ei kyetä täysin päättämään, miten ihmisaivojen kyky prosessoida monipuolista visuaalista materiaalia on kehittynyt ja miten se käytännössä toimii. (Mitchell 2019, 67–69.)

Toisaalta koneoppimisalgoritmin ja ihmisaivojen erilainen tapa käsitellä kuvamateriaalia voi olla myös etu. Koneoppimisalgoritmia voidaan kouluttaa tulkitsemaan kuvamateriaalista sellaisia ominaisuuksia tai piirteitä, joita ihminen ei kykene analysoimaan tai tunnistamaan. Esimerkiksi koneoppimisalgoritmi voi tunnistaa kasvokuvasta viitteitä perinnöllisiin sairauksiin (Neittaanmäki & Siukonen 2019, 76).

Peittymisongelma on kuitenkin hyvinkin konventionaalinen teknologiaan liittyvä pulma, jonka haasteellisuutta ei voida sivuuttaa objektien sekasortoisessa maailmassa. Peittymisongelma esiintyy erityisesti urbaanien ympäristöjen ulkopuolella, joissa luonnollista vaihtelevuutta on enemmän. Peittymisongelmaa voidaan ehkäistä neuroverkon kouluttamisen aikana tarjoamalla mahdollisimman monipuolista aineistoa koulutukseen. Tällä tavoin ehkäistään esimerkiksi se, että neuroverkko kouluttaisi tunnistamaan vain yhden osan kokonaisesta kohteesta. (Solawetz 2020.)

Pääasiassa peittymisongelmaa käsitellään siten, että osa havainnoitavasta kohteesta on peittynyt, kuten kuviossa 7 on esitetty. Peittymisongelma kuitenkin voi aiheuttaa virhetulkintoja myös silloin, kun kohde on kokonaan peittynyt, sillä se ei poista takana sijaitsevien objektien faktista olemassaoloa. Tämä on erityisesti ongelma niissä tapauksissa, joissa koneoppimisalgoritmin tehtävänä on laskea tai etsiä kuvamateriaalista jotain, joka on kokonaan peittynyt (Kuvio 7).



Kuvio 7. Tuntematon alue jää edellä olevan objektin taakse kokonaan

Toistaiseksi ei ole olemassa yleismaailmallista ratkaisua peittymisongelman ratkaisemiseksi. Peittymisongelmaa kasvien marjoilla on pyritty ratkaisemaan esimerkiksi 3D-mallinnuksen ja sen soveltamisen avulla (Jiang, Li, Ni & Takeda 2021, 299).

3.5 Kuvantunnistuksen ja konenäönmerkitys

Jo 1960-luvulla alkoi nykymuotoisen kuvantunnistuksen kehittäminen. Alkumetreillä oli selvää, että ihmisen luontaista kykyä käsitellä visuaalista materiaalia on vaikeaa jäljitellä. Keskeisimpänä ongelmana on holistisen maailman monimutkaisuus eli äärimmäisen runsas variaatio maailmankaikkeudessa. (Mitchell 2019, 69–70.) Käänteentekevin kuvantunnistussovellus oli Fei-Fei Lin 2000-luvulla kehittelemä ImageNet ja sen kehittyminen. ImageNet vahvisti käsitystä massadatan (*big data*) merkityksestä kuvantunnistuksen kehittämisessä. (Neit-taanmäki & Siukonen 2019, 77.)

Erilaisten kohteiden tunnistamiseen ja laskemiseen hyödynnettävissä sovelluk-sissa ja niihin liittyvässä tutkimuksessa esiintyy kartoittavien tutkimuksien kes-keinen ongelma eli suurten tietomäärien monotoninen käsittely. Manuaalinen objektien laskeminen ihmiseltä on suhteessa monin tavoin hitaampaa ja työ-läämpää kuin koneelta, jolle ihminen on toimittanut ainoastaan materiaalin, josta objektien määrä halutaan laskea ja objektit luokitella. Marjasatoennusteiden tapauksessa tämä tarkoittaa kuvien ottamista koeruuduista sen sijaan, että koe-ruutujen marjat laskettaisiin käsin. Myös asiantuntijoiden näkemykset puoltavat oletusta siitä, että konenäöllä ja tekoälyllä korvataan muutoin työläitä ja mono-tonisia työtehtäviä (Mäntylä 2018).

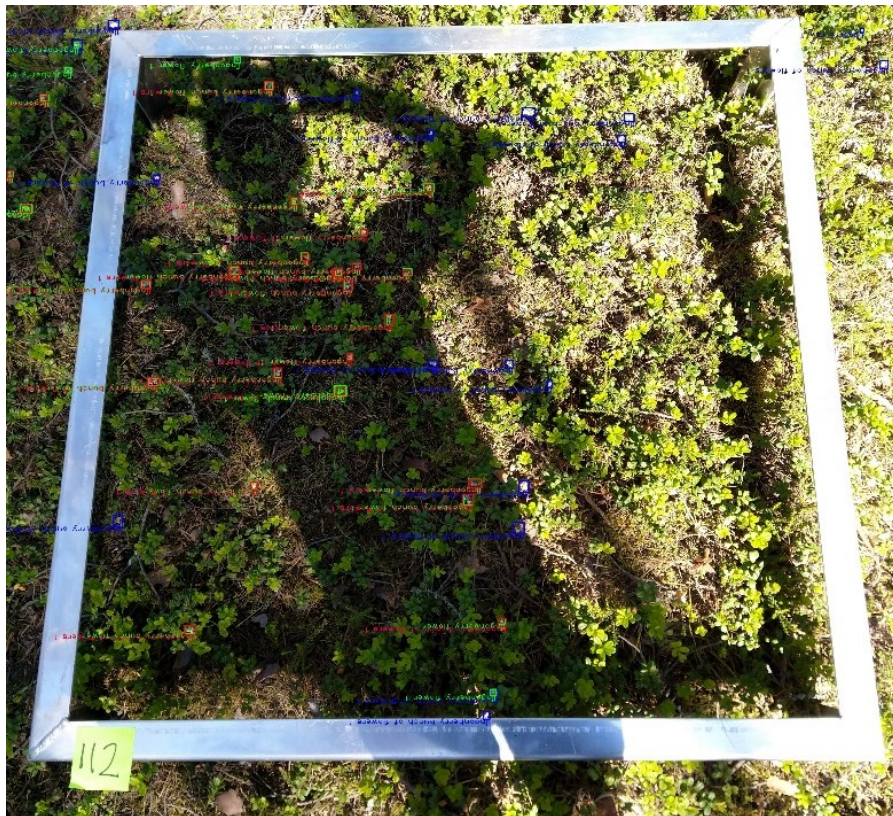
Euroopan parlamentti kutsuu tekoälyn rinnalla kehittyvää maataloutta älykkääk-si maanviljelyksi ja sen tarkoituksena on tukea Euroopan Unionin ruoantuotan-onjärjestelmän kehittymistä kestävämpään suuntaan. (Euroopan parlamentti 2021.) Euroopan Unionin tavoitteet kulkevat hyvin käsikädessä kehittyvän tek-nologian kanssa, sillä esimerkiksi Alankomaissa on kehitetty rikkaruohoja kitke-vä robotti, joka vähentää huomattavasti torjunta-aineiden käyttöä hyödyntämällä työssään konenäköä ja tekoälyä (Virtanen 2020). Kestävyyden lisäksi teknolo-gisen kehityksen tavoitteena on lisätä maatalouden tuottavuutta Euroopan unionin alueella (Euroopan parlamentti 2021).

4 AINEISTO JA MENETELMÄT

4.1 Aineiston keruu

Tutkimuksessa hyödynnetään vuoden 2021 kesällä kerättyä aineistoa, joka sisältää useita erilaisia kuvia marjakoealoilta. Aineistoa on kerätty järjestelmäkameralla sekä erilaisten puhelimien avulla. Aineistoa on kerätty Lapin ammattikorkeakoulun useissa luonnonvara-alan hankkeissa, jotka ovat liittyneet marjoihin tai niiden satojen kartoittamiseen. Järjestelmäkameran malli on Canon EOS 6D ja puhelimien mallit iPhone 11, iPhone 12, iPhone 6s, Nokia 3.4, Nokia 7+ ja Samsung a40. Marjoja ovat kuvanneet sekä laskeneet useat ihmiset, jotka lisäävät aineistoon luonnollista hajontaa. Koealoja valittiin satunnaisesti, eikä niiden määrää ole dokumentoitu.

Koealoilla marjat on laskettu yhden neliömetrin kokoiselta pinta-alalta. Apuvälineenä on käytetty joko narua tai metallista laskentakehikkoa, jotta jokainen koeala on ollut samankokoinen ja muotoinen (Kuvio 8).



Kuvio 8. Marjakoeala ja metallinen kehikko

4.2 Aineiston käsittely koneoppimisalgoritmille

Marja-aineistossa on käsitelty kuvia mustikoista ja puolukoista niiden eri kasvuvaiheissa. Jokaiseen laskentakehikkoon on kiinnitetty muistilappu, jossa on ilmoitettu pinta-alalla olevien mustikoiden tai puolukoiden määrä. Marjakoealat on valittu siten, että niissä on jompikumpi marja valtalajina.

Aineisto on kerätty Excel-taulukoihin seulomalla ne kuvat, joissa on yhden neliömetrinkokoinen kehikko kokonaan näkyvissä, marjojen todellisen määrän ilmoittava post-it-lappu näkyvissä ja koneoppimisalgoritmin antama tulos marjojen määrästä. Koneoppimisalgoritmin tuloksista huomioidaan vain se marja ja sen kehitysvaihe, jota on ollut kuvassa eniten. Laajassa aineistossa ei ole otettu huomioon kuvausteknologiaan tai -tilanteeseen liittyviä tekijöitä, joiden seurauksena voi syntyä hajontaa kuvan tulkinnan ollessa hankalaa suhteessa todellisuuteen. Laaja aineisto sisältää 1241 kuvaa.

Ennen kuin aineistoa on voitu hyödyntää koneoppimisalgoritmeilla, se on täytynyt käsitellä manuaalisesti ihmiskatseen avulla eli annotoida. Marja-aineiston annotointi on suoritettu Label Studio -työkalun avulla, jossa marjat on luokiteltu lajin ja kasvuvaiheen mukaisesti (Kuvio 9). Kasvuvaiheita on määriteltä kolme eli kukinta-, raakile- ja marjavaihe.



Kuvio 9. Label Studio -työkalu

Label Studio -työkalun avulla kuva-aineisto on käsitelty rajaamalla kuvista marjat kasvuvaiheen mukaisesti, jolloin työkalu on luonut kuviin linkitettävän tiedoston marjojen sijainneista kuvissa. Aineistoon siis luodaan erillinen tiedosto, joka sisältää ihmiskatseen manuaalisesti havaitsemat marjat ja niiden avulla koulutetaan koneoppimisalgoritmia.

Aineistossa käsiteltiin vain marjat ja esimerkiksi muiden kasviosien tunnistamista ei ole koneoppimisalgoritmille koulutettu (Kuvio 10). Koneoppimisalgoritmi siis tunnistaa mustikan ja puolukan marjoja ja niiden näköislajeja, mutta ei esimerkiksi mustikan tai puolukan varpua.



Kuvio 10. Annotoituja raakilevaiheen mustikoita

4.3 Marjat kartoittamisen ja käsittelyn kohteena

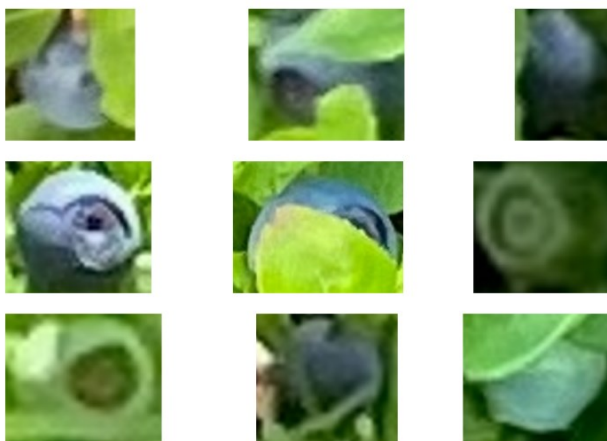
Marja-aineisto sisältää muutamia erityispiirteitä, jotka ovat vaikuttaneet aineiston keruuseen ja käsittelyyn. Huomattavin näistä erityispiirteistä on mustikan ja puolukan rakenteellinen eroavaisuus, joka on lisännyt haastetta aineistoa käsiteltäessä.

Mustikka on aineistossa havaittavissa yksittäisinä marjoina, kun taas puolukka esiintyy pääasiassa ryppäinä (Kuvio 11). Aluksi aineistossa puolukka käsiteltiin erikseen yksittäisinä sekä ryppäinä. Koneoppimisalgoritmin kouluttamisen kannalta aineiston käsittelyä oli järkevä yksinkertaistaa, jolloin päädyttiin merkitsemään ryppäitä ja yksittäin esiintyviä puolukoita samalla annotaatiolla.



Kuvio 11. Marjojen rakenteelliset erot

Marjojen osalta ei ole myöskään yksiselitteistä mihin määritellään eri kasvuvaiheiden rajat. Aineistoa käsiteltäessä tulee pyrkiä yksiselitteisyyteen, mutta joissain tapauksissa kyse on tulkinnasta ja jokainen aineistoa käsittelevä joutuu itse tekemään päätöksen rajatapauksen kasvuvaiheen luokittelusta. Luonnon monitulkintaisuus voi lisätä epätarkkuuksia koneoppimisalgoritmin antamissa tuloksissa ja lisäksi kuvausteknologiaan liittyvät haasteet lisäävät variaatiota (Kuvio 12).



Kuvio 12. Mustikoita eri kasvuvaiheissa, kuvakulmissa ja valaistuksissa

4.4 Näköislajit

Mustikalla ja puolukalla on useita eri näköislajeja, jotka voivat tunnistautua koneoppimisalgoritmillla väärin. Marjamasinassa käytettävä koneoppimisalgoritmi ei kykene tekemään tulkintoja, vaan toteuttaa mekaanisesti sille asetettuja tehtäviä. Näköislajit voivat aiheuttaa virheellisen tulkinnan löydöksestä, jos niiden ulkonäkö on tarpeeksi lähellä koneoppimisalgoritmille opetettua ulkonäköä marjasta sen jossain kasvuvaiheessa.

Kypsien marjojen näköislajit tunnetaan hyvin ja niitä ovat esimerkiksi puolukalla mauttomampi sianpuolukka (Arktiset aromit 2022a) ja mustikalla juolukka (Arktiset aromit 2022b). Aineiston käsittelyssä huomio kiinnittyi siihen, että raakilekasvuvaiheen mustikka ja puolukka voivat ulkonäöllisesti muistuttaa esimerkiksi variksenmarjaa raakilekasvuvaiheessa tai jopa kypsymätöntä katajanmarjaa, jolloin tulos voi olla väärä positiivinen (*false positive*) (Kuvio 13). Näköislajien virheellistä tunnistamista mustikaksi tai puolukaksi kutsutaan vääräksi positiiviseksi, joka on tuloksien luokitteluun liittyvä vinouma (Google Developers 2022).



Kuvio 13. Variksenmarja (vasemmalla) ja katajanmarja

Tuloksissa esiintyviä vääriä positiivisia lisää esimerkiksi se, että kuva-aineisto sisältää runsaasti kuvausteknillisiä vaihtelevuuksia. Vaihtelevuus voi liittyä esimerkiksi laitteistoon, sääolosuhteisiin tai valotukseen. Vaihtelevuus lisää koulutuksessa käytettävän kuva-aineiston väri- ja kokoeroja, jolloin koneoppimisalgoritmi voi havaita näköislajin vääräksi positiiviseksi. Suoritusongelmat eivät liity pelkästään muihin marjoihin, vain joissain olosuhteissa raakilevaiheiden marjoja voi olla haastavaa havaita lehtien seasta (Kuvio 14).

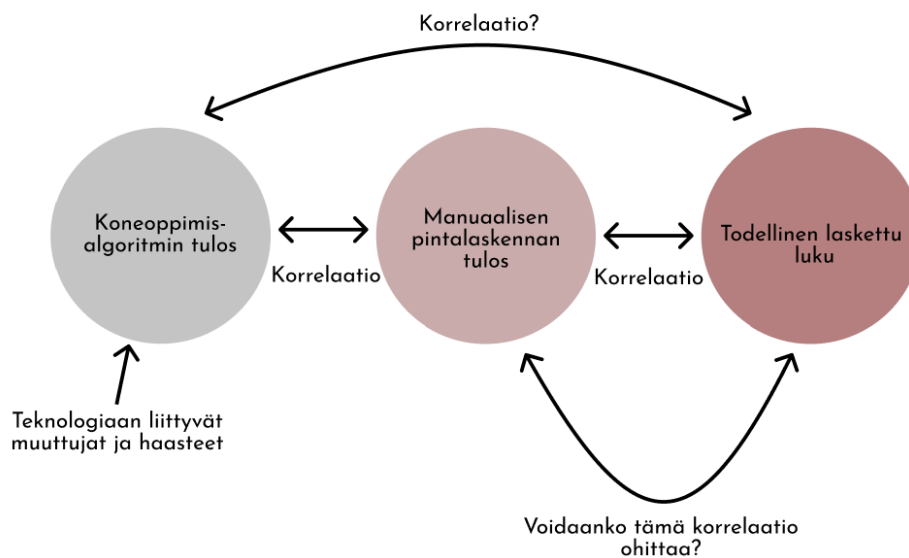


Kuvio 14. Vaikeasti havaittavia raakilevaiheen puolukoita

Väärät positiiviset tulokset voivat aiheuttaa ongelmia, jos algoritmin ensisijainen tarkoitus on laskemisen sijasta keskittyä esimerkiksi luokitteluun erilaisia asioita. Marjojen osalta ongelma on ilmeinen, sillä esimerkiksi sinisiä marjoja tunnistamaan koulutettu koneoppimisalgoritmi voisi luokitella myrkyllisen sudenmarjan mustikaksi (Luontoportti 2022). Tämänkaltaisen ongelman voitaisiin korjata kouluttamalla koneoppimisalgoritmille koko kasvin tunnistaminen pelkästään marjojen sijasta.

4.5 Tutkimuskysymys

Kehittämisen prosessin aikana on syntynyt kolme erilaista laskentatulosta: ihmisen luonnossa manuaalisesti laskema tulos, ihmisen kuvasta manuaalisesti laskema tulos ja koneen kuvasta manuaalisesti laskema tulos. Kuviossa 15 esitetään aineistosta saatavia laskematuloksia ja niiden suhde tutkimuskysymykseen, joka tarkastelee näiden välistä korrelaatiota.



Kuvio 15. Tutkimuskysymyksen asetelma

Käytännössä nykyaikaisen teknologian avulla voidaan saavuttaa riittävä korrelaatio ihmisen manuaalisesti kuvasta laskeman tuloksen ja koneoppimisalgoritmin antaman tuloksen välillä, minkä vuoksi tutkimuskysymys keskittyy koneoppimisalgoritmin ja marjojen todellisen määrän välisen korrelaation tarkasteluun. Koneoppimisalgoritmi jäljittelee ihmisen manuaalisesti kuvasta laskeman tuloksen taktiikkaa, jolloin näiden kahden tuloksen välinen korrelaatio on riippuvainen hankkeessa kehitetyn koneoppimisteknologian ja sen algoritmin edistymisestä.

4.6 Tutkimusmenetelmä

Opinnäytetyössä lasketaan korrelaatio kuva-aineistosta, josta on saatavilla koneoppimisalgoritmin pintalaskentatulokset ja ihmisten manuaalisesti laskelmaa tulosten todellisesta määrästä. Näiden tulosten keskinäistä riippuvuutta tutkitaan Pearsonin korrelaatiokertoimella, joka voi mahdollisesti osoittaa yhteyden, jossa peittyneet kasvusto käyttäytyy säännönmukaisesti suhteessa niihin marjoihin, jotka on mahdollista havaita 2D-kuvasta.

Korrelaatiota tutkitaan Excel-työkalun avulla, jonka sisäänrakennetut ominaisuudet mahdollistavat sen vaivattomasti. Riippuvuus tutkitaan laitteittain, sekä marjoilla kasvuvaiheittain. Jaottelulla on tarkoitus selvittää, minkälaista vaihtelua esiintyy, mistä ominaisuuksista vaihtelu mahdollisesti johtuu ja mitkä tekijät antavat parhaan mahdollisen lopputuloksen.

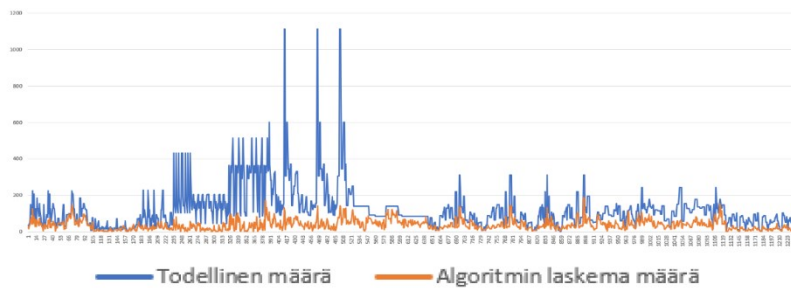
Lisäksi korrelaatiota tutkitaan suppeammalla laatuaineistolla, jossa korrelaation tutkiminen rajataan tarkemmin. Laatuaineistossa ei tutkita pelkästään laitteittain tai marjojen kasvuvaiheittain korrelaatiota, vaan rajataan tutkimus koskemaan yhdellä laitteella tapahtuvaa korrelaatiota marjalla ja sen kasvuvaiheella. Laatuaineiston tuloksia verrataan laajemman aineiston tuloksiin. Tämän tarkoituksena on osoittaa tutkimuksen toistettavuutta tai sen puutetta, sillä laajemman aineiston koko voi vääristää tuloksia. Laatuaineisto on valittu satunnaisotannalla.

Tutkimusaineistosta on karsittu pois ne kuvat, sekä niiden mahdolliset laskentatulokset, jos niistä puuttui joko algoritmin antama pintalaskentatulokset tai ihmisten manuaalisesti laskelmaa marjojen todellinen määrä. Nämä kuvat ovat olleet lähtökohtaisesti läheltä otettuja kuvia, joilla koneoppimisalgoritmia on koulutettu tunnistamaan marjojen ominaisuuksia. Tutkimuskysymyksen ja -menetelmän kannalta näillä kuvilla ei ole merkitystä, sillä niistä ei voitaisi muutenkaan laskea määriteltyä pintalaskentatulosta.

5 TULOKSET JA JOHTOPÄÄTÖKSET

5.1 Korrelaatio laajalla aineistolla

Laajan aineiston korrelaatiota mitattiin Pearsonin korrelaatiokertoimen avulla. Laajassa aineistossa otettiin huomioon kaikki kuvat, jotka olisivat myös käytännössä kelvollisia marjojen kartoittamiseksi. Pearsonin korrelaatiokertoimen antama tulos on noin 0,52. Tulos kertoo, ettei merkittävää korrelaatiota ole havaittavissa, vaikka otos on suuri. Keskimääräisesti hajonta jakautui tasaisesti, vaikka hajontakuviosta on nähtävissä esimerkiksi suuret erot laskentatulosten välillä niillä aloilla, joissa oli määrällisesti paljon marjoja (Kuvio 16).



Kuvio 16. Esiintyvä hajonta

Eri laitteiden välillä on nähtävissä huomattavia eroja korrelaatioissa. Esimerkiksi iPhone 11, iPhone 12 ja järjestelmäkamera saivat huomattavasti parempia tuloksia kuin Nokian puhelimet ja Samsung a40 (Taulukko 1).

Taulukko 1. Eri laitteiden tulomomenttikorrelaatiokertoimet

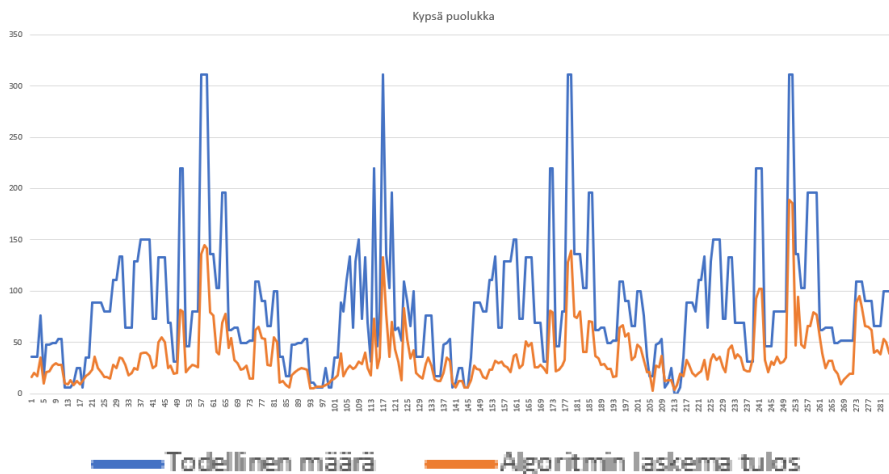
Laite	Pearsonin tulomomenttikorrelaatiokerroin	Otoskoko (yht. 1241)
iPhone 11	0,76	68
iPhone 12	0,66	275
iPhone 6s	0,01	52
Nokia 3.4	-0,39	14
Nokia 7 Plus	0,54	297
Samsung a40	0,54	230
Canon	0,63	305

Kun hajontaa tarkastelee marjojen lajien ja kehitysasteiden mukaan, voidaan huomata ero kehitysasteiden välillä. Mustikan osalta korrelaatio lisääntyy marjasadon kypsyessä, puolukan osalta korrelaatio on tasaisempi, mutta mustikkaa parempi (Taulukko 2).

Taulukko 2. Pearsonin tulomomenttikorrelaatiokerroin marjan ja sen kehitysvaiheen mukaan

Marja ja kehitysvaihe	Pearsonin tulomomenttikorrelaatiokerroin	Otoskoko (yht. 1241)
Mustikka, kukka	0,32	217
Mustikka, raaka	0,59	178
Mustikka, kypsä	0,76	126
Puolukka, kukka	0,69	298
Puolukka, raaka	0,72	137
Puolukka, kypsä	0,86	285

Kokonaiskuvan kannalta mustikan kukka on tulosta heikentävä tekijä, sillä sen korrelaatio on suhteessa muihin kehitysvaiheisiin ja puolukkaan heikko. Selittävä tekijä voi olla esimerkiksi marjojen rakenteessa, sillä puolukan marjat kasvavat ryppäinä ja mustikka yksittäisinä marjoina. Etenkin kypsän puolukan osalta korrelaatio on selkeästi havaittavissa ja hajonta vähäistä (Kuvio 17).



Kuvio 17. Kypsän puolukan hajonta

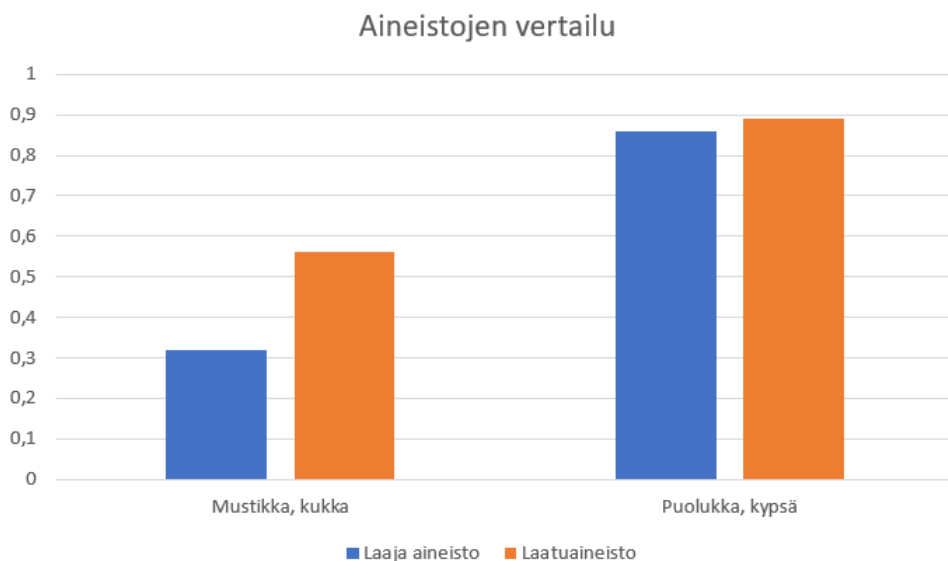
5.2 Korrelaatio laatuaineistolla

Laatuaineiston tarkoituksena on todentaa laajan aineiston tuloksia suppeammalla otannalla ja lisätä tutkimuksen toistettavuutta. Laatuaineistoon valitsin iPhone 12:lla kuvatut kypsät puolukat ja Nokia 7 Plussalla kuvatut mustikan kukat. Laatuaineistolla suoritettu tutkimus korrelaatiosta vahvistaa laajan aineiston tulosten paikkansapitävyyttä (Taulukko 3).

Taulukko 3. Pearsonin tulomomenttikorrelaatiokerroin laatuaineistossa

Laite	Marja ja kehitysvaihe	Pearsonin tulomomenttikorrelaatiokerroin	Otoskoko
iPhone 12	Puolukka, kypsä	0,89	82
Nokia 7 Plus	Mustikka, kukka	0,56	50

iPhone 12:ssa on 12 megapikselin kaksoiskamera (Apple 2022). Nokia 7 Plus:ssa 12:n ja 13 megapikselin kaksoiskamera, joten laitteiden välillä ei ole merkittävää teknistä eroa (Nokia 2022). Hajontaa kokonaistuloksiin voi kuitenkin selittää esimerkiksi, miten usein ja missä valaistuksessa kutakin puhelinta on käytetty tiedon keräämiseen. Laitekohtaiset tulokset kuitenkin noudattelevat marjakohtaisten tulosten suuntaa, joissa kypsä puolukka on huomattavasti paremmin korreloiva kuin kukintavaiheen mustikka (Kuvio 18).

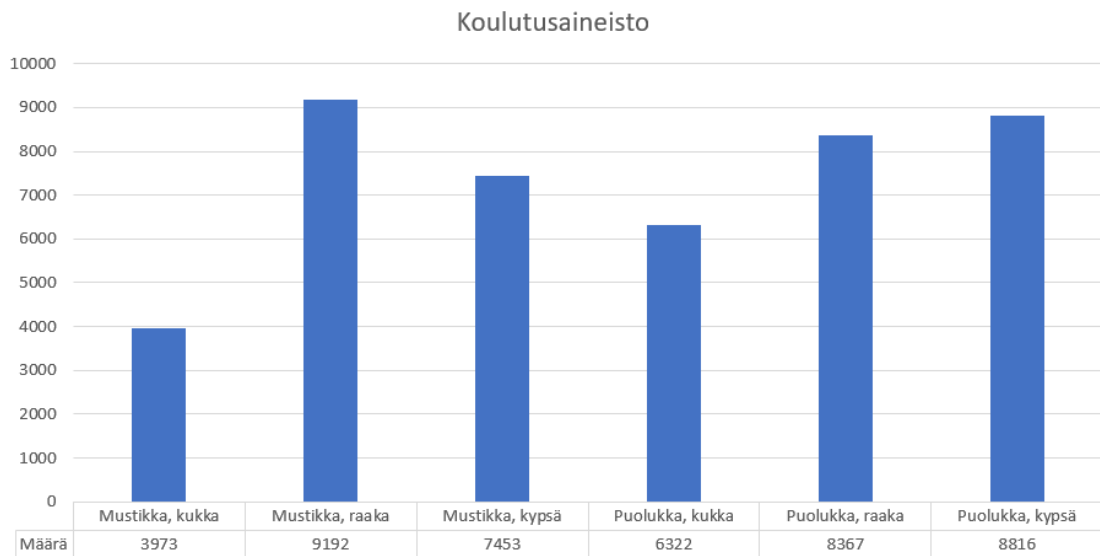


Kuvio 18. Aineistojen vertailu

5.2.1 Algoritmin koulutus

Kuvasta saatu pintalaskentatulokset antaa suuntaa antavan arvion marjojen määrästä koealalla, mutta tarkkaa tulosta sillä ei välttämättä kyetä selvittämään. Mahdollisuus selvittää pintalaskentatuloksesta arvio marjojen määrästä riippuu marjasta ja sen kasvuvaiheesta. Vaihtelua marjojen ja niiden kasvuvaiheiden välillä selittävät muutamat eri tekijät.

Marjamasiinassa käytettyä algoritmia on koulutettu 44123 kuvalla, jotka on annotoinnin aikana rajattu koneellisesti isommista kuvista. Yhdessä kuvassa on vain yksi marja tai marjarypäs. Algoritmia ei ole koulutettu tasaisella määrällä kuvia jokaisesta marjasta ja sen kasvuvaiheesta, vaan koulutuksessa on esiintynyt vaihtelua (Kuvio 19).



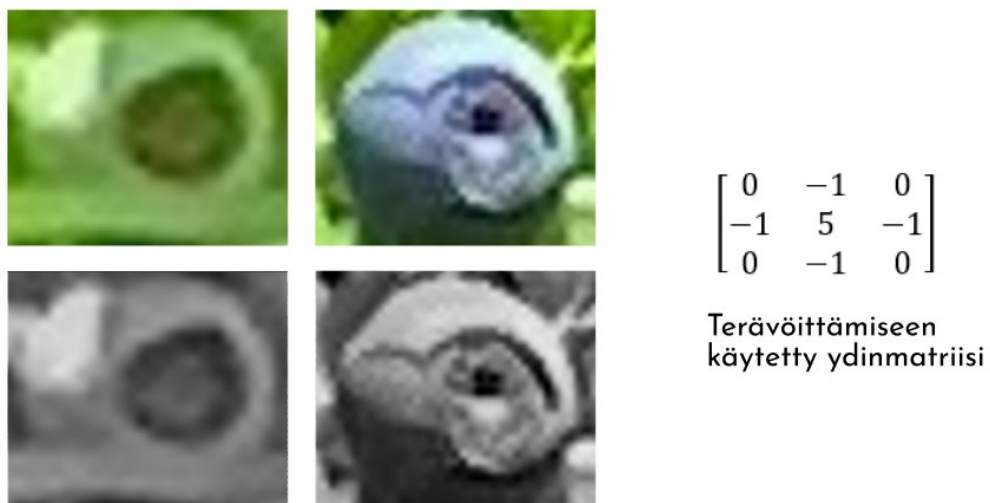
Kuvio 19. Koulutusaineistossa käytetyt kuvat

Koulutusaineistossa esiintyvä vaihtelu on hyvin samanlaista kuin marjojen eri kasvuvaiheilla lasketulla korrelaatiolla. Ainoa poikkeus on raaka mustikka, jota on koulutettu enemmän kuin kypsää mustikkaa.

5.2.2 Marjojen ja niiden kasvuvaiheiden erot

Marjat ja niiden eri kasvuvaiheet näyttäytyvät graafisesti hyvin eri tavoin. Kuten luvussa 3.3 tuli esille, esimerkiksi raakilevaiheen puolukka on niin väriltään kuin muodoltaankin hyvin samankaltainen kuin puolukkakasvin lehti. Tämä vaikeuttaa marjojen koneellista tunnistamista verrattuna esimerkiksi kypsään puolukkaan, jonka väri on merkittävästi erilainen verrattuna puolukkakasvin lehteen. Koneoppimisalgoritmia on koulutettu ja testattu myös erillisellä ruska-aineistolla, jotta voitiin parantaa tunnistuskykyä myös marjojen ympäristöjen muuttuessa.

Vaikka algoritmia on koulutettu enemmän raa'alla kuin kypsällä mustikalla, oli korrelaatio todelliseen marjojen määrään suurempi kypsällä mustikalla. Tämä toisintaa havaintoa siitä, että marjojen erilaisuus suhteessa ympäristöön lisää tunnistuksen tarkkuutta. Tämä selittynee kuvantunnistusalgoritmin hyödyntämällä menetelmällä, jossa kuvaa käsitellään erityispiirteiden etsimiseksi ja objektien löytämiseksi siten kuvasta (Kananen & Puolitaival 2019, 152–153). Raa'an mustikan vähäinen värinmuutos suhteessa mustikan lehteen heikentää sen löydettävyyttä, vaikka kuvaa käsiteltäisiinkin erityispiirteiden etsimiseksi. Kuviossa 20 esitetään Image Kernels -sivuston avulla terävöitetty mustikka (Powell 2022).



Kuvio 20. Raaka ja kypsä mustikka terävöitettynä

6 POHDINTA

6.1 Hyödynnettävyys

Opinnäytetyössä on kyetty osoittamaan korrelaatio marjojen pintalaskentatuloksen ja todellisen määrän välille. Tämä antaa mahdollisuuden kehittää menetelmän, jonka avulla voidaan päätellä marjojen todellinen määrä marjakoealasta otetuista kuvista. Peittymisongelman ratkaiseminen yksinkertaisesti mittaamalla todellisen maailman ja digitaalisesti tallennetun tietojen välistä yhteyttä antaa mahdollisuuden ratkaista ongelma myös muissa tapauksissa, joissa luonnonvarojen esiintyminen on jollain tavalla säännömukaista.

Konenäön ja luonnonvarojen yhteensovittaminen vaatii syvällisen ymmärryksen, jossa kyetään näkemään maailma niiden rajoitteiden mukaan, mitä konenäkö ja koneoppiminen omalle tiedonkäsittelylleen asettaa. Jos näitä ongelmia ei kyetä ratkaisemaan, eivät tulokset vastaa mitattavaa todellisuutta. Opinnäytetyössä olen oppinut laajasti soveltamaan monialaista osaamista, jossa yhdistyy luonnonvara-alan asiantuntijuus ja tietotekninen ymmärrys.

Kuten aikaisemmin luvussa 2.3 on tuotu esille, marjasatoennusteita tuottavat vapaaehtoiset henkilöt sekä 4H-nuoret. Marjamasiina-hankkeen tavoitteena ollut mobiilikäyttöliittymä havaintojen keräämiseen madaltaa etenkin nuorilla ihmisillä toimintaan mukaan lähtemistä, sillä mobiilikäyttöliittymään voidaan sulauttaa esimerkiksi käyttöohjeita ja havainnollistamista, mikä helpottaa tietojen keräämistä. Mobiilikäyttöliittymän suunnittelussa tulee ottaa huomioon kohde-ryhmä, joka on pääasiassa lapset ja kansalaiset.

Mobiilikäyttöliittymän avulla suoritettu tiedonkerääminen mahdollistaa myös reaaliaikaisen tietojen lähettämisen eteenpäin, jolloin käyttäjän ei tarvitse erikseen rekisteröidä tietoja eteenpäin, kuten nykyään Marjahavainnot.fi-palvelun kautta. Myös tietojen automaattinen lähettäminen mobiilikäyttöliittymästä madaltaa kynnystä ryhtyä marjahavaintojen kerääjäksi, sillä se lisää vaivattomuutta.

6.2 Hyödynnettävyyden haasteet

Korrelaatio pintalaskentatuloksen ja marjojen todellisen määrän välillä osoittaa, että on mahdollista päätellä todellinen marjojen määrä, kun koneoppimisalgoritmi on laskenut kuvasta löydettävät marjat. Määrä ei ole yhtä tarkka, mitä ihmisen käsin laskema määrä ja vaihtelua esiintyy runsaasti (Kuvio 21). Hyödynnettävyys vaatisi, että kuvia kasvupaikalta otettaisiin mahdollisimman paljon ja niiden tuloksia yhdisteltäisiin keskenään.



Oikea määrä: 140
Kuvasta laskettu: 87



Oikea määrä: 140
Kuvasta laskettu: 85



Oikea määrä: 140
Kuvasta laskettu: 7

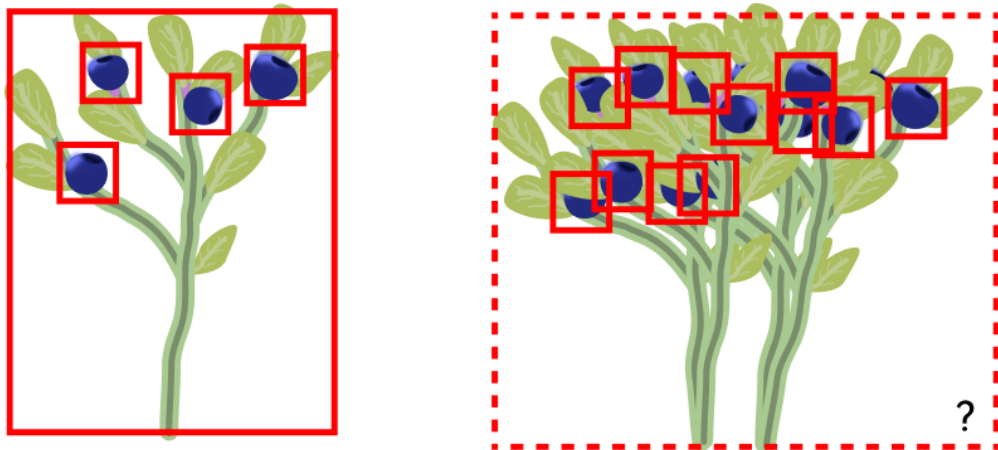
Kuvio 21. Saman koealuruudukon eri tuloksia raa'alla mustikalla

Ottamalla huomioon eri kasvuvaiheiden ja marjojen väliset erot tunnistettavuudessa, on mahdollista kehittää jonkinlainen laskentamalli, jolla voidaan päätellä alalla olevien marjojen määrä suuntaa antavasti. Vaikka marjojen määrää ei voida tarkasti päätelläkään, on keino huomattavasti nopeampi kartoittamaan määrätyn alan marjojen esiintyvyyttä ja modernien mobiililaitteiden avulla voidaan myös tallentaa osaksi kuvaa aika ja tarkka sijainti automatisoidusti.

Marjojen määrää arvioivassa laskentamallissa tulee ottaa huomioon tuloksissa esiintyvä vaihtelu, sekä miten käyttäjät tulisivat sovellusta käyttämään. Esimerkiksi eri laitteiden välinen ero tulee ottaa huomioon marjojen ja niiden kasvuvaiheiden eron lisäksi.

6.3 Haasteiden pohdinta

Näköislajien ja marjojen rakenteisiin liittyviä ongelmia voisi periaatteessa ratkoa lisäämällä algoritmin kouluttamista muilla kasvinosilla. Tällöin algoritmi pystyisi vahvistamaan marjojen luokittelua myös muiden esiintyvien ominaisuuksien perusteella. Ratkaisu ei kuitenkaan olisi täysin aukoton, sillä kenttäkerroksessa kasvavilla kasveilla olisi todennäköisesti entistä suurempia haasteita peittymisongelman suhteen. Jos algoritmi tunnistaisi marjojen lisäksi varvun, peittymisongelma olisi todennäköisesti suurempi monesta eri perspektiivistä (Kuvio 22).



Kuvio 22. Peittymisongelman tunnistushaaste

Pelkästään koulutusprosessi vaatisi, että algoritmi opetetaan tunnistamaan yksittäinen varpu yksittäisenä varpuna, sillä ihmissilmälläkin on lähes mahdotonta kyetä kenttäkerroksesta erottamaan yksittäisiä varpuja annotointiprosessia varten. Jos kokonaisen varvun sijasta pohdittaisiin lehtien tunnistamista marjan lisäksi ja sen liittäminen marjojen luokitteluprosessiin, voisi hankaluuksia esiintyä mustikka- ja puolukkakasvustojen vaihtumisvyöhykkeillä. Todennäköisesti tunnistuksesta saatu hyöty olisi pienempi kuin siitä syntynyt haitta. Tämä myös hidastaisi algoritmin toimintaa, sillä prosessoitavaa olisi aikaisempaa enemmän.

6.4 Menetelmän soveltaminen ja jatkotutkimus

Opinnäytetyössä esiteltyä menetelmää voidaan soveltaa myös muiden luonnossa esiintyvien peittymisongelmien tarkasteluun. Mahdollisesti löytyvät korrelaatiot mahdollistavat erilaisten laskentamallien luomisen, joiden avulla voidaan laskea ja päätellä myös erilaisten luonnollisten objektien takana olevia asioita.

Kuvantunnistus ja koneoppiminen ovat toimineet mahdollistajana monien ongelmien ratkaisemisessa, minkä pulmallisuus on tähän asti ollut teknologisen kehityksen saavuttamattomissa. Satelliittien ja lentoalusten tarjoama kuvamateriaali luonnosta ja luonnonvaroista yhdistettynä tehokkaaseen prosessointiin tekoälyllä lisää laajamittaisesti ymmärrystä niin luonnon kuin ilmastokin kokonaisvaltaisesta kehityksestä. Esimerkiksi Kiinassa kuvantunnistusta on sovellettu villieläinten inventoinneissa, joiden päämääränä on suojella biologista monimuotoisuutta ja ehkäistä villieläimiin kohdistuvaa salametsästystä (Liao ym. 2020, 364).

Kuvantunnistukseen koneoppimisalgoritmeilla liittyy eettisiä tekijöitä, jotka tulee huomioida jatkotutkimuksessa. Marjojen kuvantunnistuksessa eettinen kestävyys on hyvällä tasolla, sillä se ei pyri yksilöimään tietoa, joka liittyyisi mahdollisesti ihmisiin. Myös peittymisongelma marjoilla tai muilla luonnonvaroilla sisältää aineistoa, joka lähtökohtaisesti ei ole arkaluontoista ja sen tutkimus on pitkälti matemaattista ja luonnonvara-alan asiantuntijuuden hyödyntämistä. Suomessa marjat ovat myös jokamiehenoikeuden alaisia, jolloin eettisyydessä ei tule ristiriitaa yksityisomaisuuden kanssa.

Peittymisongelman esiintymistä luonnonvaroja kartoittaessa voisi tulevaisuudessa tutkia esimerkiksi kaivannaisteollisuudessa, uhanalaisten luontotyyppien kartoituksessa tai muilla kasveilla. Haastavinta peittymisongelman tutkimisessa on aineiston saatavuus, sillä korrelaatio voidaan todeta kahden aineiston välille, joista toinen on lähtökohtaisesti manuaalisesti tuotettu ja kerätty.

LÄHDELUETTELO

Apple 2022. iPhone 12 – Tekniset tiedot. Viitattu 11.8.2022
https://support.apple.com/kb/SP830?locale=fi_FI.

Arktiset Aromit ry 2022a. Mustikka. Viitattu 23.3.2022
<https://www.arktisetaromit.fi/fi/marjat/luonnonmarjat/mustikka/>.

Arktiset Aromit ry 2022b. Puolukka. Viitattu 23.3.2022
<https://www.arktisetaromit.fi/fi/marjat/luonnonmarjat/puolukka/>.

Bengio, Y., Courville, A. & Goodfellow, I. 2016. Deep learning. Lontoo: The MIT Press.

Divvala, S., Farhadi, A., Girshick, R. & Redmon, J. 2016. You Only Look Once: Unified, Real-Time Object Detection. Viitattu 27.3.2022
<https://arxiv.org/pdf/1506.02640v5.pdf>.

Euroopan parlamentti 2021. Mitä tekoäly on ja mihin sitä käytetään? Viitattu 9.8.2022
<https://www.europarl.europa.eu/news/fi/headlines/society/20200827STO85804/mita-tekoaly-on-ja-mihin-sita-kaytetaan>.

Google Developers 2022. Classification: True vs. False and Positive vs. Negative. Viitattu 23.3.2022 <https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>.

Hyvärinen, P. 2021. Keräilytalouksia ja plantaasimaisuuksia metsissä ja metsien raunioilla. Viitattu 18.10.2022
<https://www.versuslehti.fi/tiededebatti/kerailytalouksia-ja-plantaasimaisuuksia-metsissa-ja-metsien-raunioilla/>.

Ihamuotila, R. 2011. 300 kasvia Suomen luonnossa. Helsinki: Otava.

Jiang, H., Li, C., Ni, X., & Takeda F. 2021. Three-dimensional photogrammetry with deep learning: instance segmentation to extract berry fruit harvestability traits. ISPRS Journal of Photogrammetry and Remote Sensing vol 169, 364-376. Viitattu 3.4.2021 <https://doi.org/10.1016/j.isprsjprs.2020.11.010>.

Kananen, H. & Puolitaival, H. 2019. Tekoäly. Bisneksen uudet työkalut. Helsinki: Alma Talent.

Kelleher, D. 2020. Syväoppiminen. Kuinka tekoäly toimii. Helsinki: Terra Cognita.

Kivelä, S. 2021. Keräilyllä on myönteinen vaikutus hyvinvointiin – Kerro, millaisia kokemuksia sinulla on sienten tai marjojen keräämisestä? Viitattu 18.10.2022 <https://yle.fi/aihe/artikkeli/2021/08/13/kerailylla-on-myonteinen-vaikutus-hyvinvointiin-kerro-millaisia-kokemuksia>.

Liao, X., Peng, J., Shao, Q., Sun, Z., Wang, D., Ye, H., & Yue, H. 2020. Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS Journal of Photogrammetry and Remote Sensing* vol 171, 297-309. Viitattu 2.12.2021 <https://doi.org/10.1016/j.isprsjprs.2020.08.026>.

Luonnonvarakeskus 2022a. Luonnonmarjojen satoseurannat ja ennusteet – kuvaus. Viitattu 10.10.2022 <https://www.luke.fi/fi/seurannat/luonnonmarjojen-satoseurannat-ja-ennusteet/luonnonmarjojen-satoseurannat-ja-ennusteet-kuvaus>.

Luonnonvarakeskus 2022b. Luonnonmarjojen satohavainnot. Viitattu 10.10.2022 https://marjahavainnot.fi/assets/info/Havaintometsan_perustaminen_v2.pdf.

Luonnonvarakeskus & 4H 2022a. Mustikka (*Vaccinium myrtillus*). Viitattu 10.10.2022 <https://marjahavainnot.fi/#/mustikka>.

Luonnonvarakeskus & 4H 2022b. Puolukka (*Vaccinium vitis-idaea*). Viitattu 10.10.2022 <https://marjahavainnot.fi/#/puolukka>.

Luonnonvarakeskus & 4H 2022c. Etusivu. Viitattu 10.10.2022 <https://marjahavainnot.fi/#/>.

Luonnonvarakeskus & 4H 2022d. Kartta. Viitattu 23.11.2022 <https://marjahavainnot.fi/#/kartta>.

Luontoportti 2022. Sudenmarja. Viitattu 26.3.2022 <https://luontoportti.com/t/437/sudenmarja>.

Marjala, M. 2015. Marjasadon ennustaminen on raakaa työtä. Viitattu 13.10.2022 <https://yle.fi/uutiset/3-8063270>

Mitchell, M. 2019. *Artificial Intelligence. A Guide for Thinking Humans*. New York: Picador.

MTK 2017. Ekosysteemipalvelut. Viitattu 18.10.2022 <https://www.mtk.fi/-/ekosysteemipalvelut>.

Mäntylä, J-M. 2018. Nuori, älä ole huolissasi – tekoäly tuhoaa vain tylsät työt. Viitattu 6.8.2022 <https://yle.fi/uutiset/3-10235593>.

Neittaanmäki, P. & Siukkonen, T. 2019. *Mitä tulisi tietää tekoälystä*. Jyväskylä: Docendo.

Nokia 2022. Nokia 7 Plus. Viitattu 11.8.2022 https://www.nokia.com/phones/en_int/nokia-7-plus?sku=11B2NB21A08.

Powell, V. 2022. Image Kernels. Viitattu 18.10.2022 <https://setosa.io/ev/image-kernels/>.

Ruokavirasto 2022. Marsi 2021. Luonnonmarjojen ja –sienten kauppaantulo-määrät vuonna 2021. Viitattu 13.10.2022
<https://www.ruokavirasto.fi/globalassets/viljelijat/tuet-ja-rahoitus/marsi-2021-raportti.pdf>.

Salo, K. 2015. Metsä: Monikäyttö ja ekosysteemipalvelut. Helsinki: Luonnonva-rakeskus LUKE.

Solawetz, J. 2020. Occlusion Techniques in Computer Vision. Viitattu 3.4.2022
<https://blog.roboflow.com/occlusion-computer-vision/>.

STT 2022. Puolukkasato on tänä syksynä keskimääräistä runsaampi – "Marjat alkavat olla pikkuhiljaa valmiita poimittavaksi". Viitattu 13.10.2022
<https://yle.fi/uutiset/3-12605764>.

Vihervaara, P., Ahokumpu, A., Auvinen, A. & Mononen, L. 2017. Marjojen ja sienten arvo. Viitattu 18.10.2022
<https://www.luonnontila.fi/ekosysteemipalvelut/ekosysteemipalvelut/tuotantopalvelut/marjat-ja-sienet/arvo>.

Virtanen, S. 2020. Tekoäly kitkee rikkaruohot pellosta. Viitattu 9.8.2022
<https://www.tekniikkatalous.fi/uutiset/tekoaly-kitkee-rikkaruohot-pellosta/a75dca3d-56d2-4940-8eca-33f834345eaa>.

Ympäristöministeriö. 2017. Usein kysytyt kysymykset jokamiehen oikeuksista. Viitattu 18.10.2022 [https://www.ymparisto.fi/fi-FI/Luonto/Usein_kysytyt_kysymykset_jokamiehen_oikeu\(17111\)](https://www.ymparisto.fi/fi-FI/Luonto/Usein_kysytyt_kysymykset_jokamiehen_oikeu(17111)).