

TEKOÄLYLLÄ TUOTETTUIJEN RAKENNUSAINEISTOJEN LAADUN ARVIOINTI

Hautamäki Veera

Opinnäytetyö

Maanmittaustekniikka
Insinööri (AMK)

2023

Maanmittaustekniikka
Insinööri (AMK)

Tekijä	Veera Hautamäki	Vuosi	2023
Ohjaaja	Teuvo Heimonen		
Toimeksiantaja	Maanmittauslaitos, Mikko Väisänen		
Työn nimi	Tekoälyllä tuotettujen rakennusaineistojen laadun arviointi		
Sivumäärä	46 + 13		

Tämän opinnäytetyön tarkoituksena oli selvittää, voiko tekoälyn avulla tuotettuja rakennusaineistoja hyödyntää maastotietokannan rakennusten ylläpidossa. Lisäksi kehitettiin menetelmä, jolla kartoittaja löytää tekoälyaineistosta maastotietokannan rakennusten päivittämisen kannalta kriittiset kohteet. Käytössä oli kaksi erilaista tekoälyn avulla tuotettua rakennusaineistoa: AI-lidar-aineisto ja AI-orto-aineisto. AI-lidar-aineisto oli tuotettu laserkeilausaineistosta ja AI-orto-aineisto ilmakuvilta tuotetuista tosiortokuvilta. Näitä aineistoja verrattiin maastotietokannan rakennuksiin ja tarkasteltiin ortokuvilla.

Opinnäytetyön aihetta käsiteltiin määrällisten menetelmien kautta. Aineistoille tehtiin ristiin vertailut päällekkäisyysanalyysillä, minkä jälkeen rakennuskohteet lajiteltiin eri luokkiin päällekkäisen pinta-alan mukaan. Tuloksia esiteltiin diagrammeina, taulukoina ja kuvioina.

Opinnäytetyön tulokset osoittivat, että tekoälyn avulla tuotettuja rakennusaineistoja voidaan hyödyntää maastotietokannan rakennusten ylläpidossa. Päällekkäisanalyysiä hyödyntäen tekoälyaineistoista voidaan löytää maastossa olevia rakennuskohteita, jotka puuttuvat maastotietokannasta sekä rakennuskohteita, joita ei enää maastossa ole ja jotka tulisi poistaa maastotietokannasta. Toimintatapa edellyttää, että tietty osa rakennuskohteista käydään yksitellen läpi, mutta koska niiden suhteellinen osuus on melko pieni, se on mahdollista. Tulevaisuudessa tilanne voi vielä parantua, koska tekoälyaineistoista on mahdollisesti tulossa yksi, yhdistetty aineisto.

Avainsanat Aineistot, rakennukset, tekoäly

Muita tietoja Työhön liittyy toimeksiantajalle toimitettu ohje päällekkäisyysanalyysin tekoon ja rakennusten luokitteluun

Degree Programme in Land Survey-
ing Engineering
Bachelor of Engineering

Author	Veera Hautamäki	Year	2023
Supervisor	Teuvo Heimonen		
Commissioned by	National Land Survey of Finland, Mikko Väisänen		
Title	Quality of Buildings Produced by Artificial Intelligence		
Number of pages	46 + 13		

The aim of thesis study was to find out whether building datasets produced with the help of artificial intelligence could be utilized in the maintenance of the topographic database.

The study was carried out using a quantitative method. Cross-comparisons were made between the data using an overlap analysis. After that, the building objects were sorted into different categories according to the common surface area of the buildings. The results were presented in the form of graphs, tables and figures. Also, a method by which the cartographer finds critical objects from artificial intelligence data was developed. The method can be used to update the topographic database. AI-lidar and AI-ortho-artificial intelligence data were used. The AI data was compared with the buildings in the topographic database and examined with orthophotos.

The thesis was able to answer research questions. The results showed that building dataset produced with the help of artificial intelligence can be utilized when the buildings in the topographic database are updated. AI data can be used to find critical building objects which could be utilized in the maintenance of the topographic database. It is necessary to go through some of the building objects one by one but since there are only few of them it is possible. In the future, the situation may improve if a single, combined AI data is made available.

Keywords	Datasets, buildings, artificial intelligence
Special remarks	Thesis includes a guide for conducting overlap analysis and classifying buildings

SISÄLLYS

1 JOHDANTO	8
1.1 Tutkimuksen tarkoitus ja tavoite	9
1.2 Tutkimuskysymykset	10
1.3 Maanmittauslaitoksen organisaatio ja tehtävät	10
1.4 Peruspaikkatietopalvelut (PATI)	12
2 TEKOÄLYAINEISTOJEN KEHITTÄMISPROJEKTIT	13
3 AINEISTOT, MENETELMÄT JA ALUE	15
3.1 Aineistot	15
3.2 Menetelmät	16
3.3 Alue	18
4 TULOKSET	20
4.1 Aineistojen vertailu päällekkäisyysanalyysillä	20
4.1.1 MTK:n rakennukset vs AI-lidar-rakennukset	20
4.1.2 AI-lidar-rakennukset vs MTK:n rakennukset	23
4.1.3 MTK:n rakennukset vs AI-orto-rakennukset	25
4.1.4 AI-orto-rakennukset vs MTK:n rakennukset	27
4.1.5 Tekoälyaineistojen ristiin vertailu	29
4.1.6 Rakennusten päällekkäisyys verrattuna eri aineistojen kesken	30
4.2 Rakennusten jakautuminen 0–10 %-luokassa	31
4.2.1 MTK:n rakennukset vs AI-lidar-aineiston rakennukset	33
4.2.2 AI-lidar-rakennukset vs MTK:n rakennukset	34
4.2.3 MTK:n rakennukset vs AI-orto-rakennukset	35
4.2.4 AI-orto-rakennukset vs MTK:n rakennukset	36
4.3 Tekoälyaineistojen geometriat	37
4.4 Tulosten yhteenveto	39
5 JOHTOPÄÄTÖKSET	41
6 POHDINTA	42
LÄHTEET	44
LIITTEET	46

ALKUSANAT

Kiitän Maanmittauslaitoksen Peruspaikkatietopalveluiden (PATI) tulosityksikköä saamastani opinnäytetyön aiheesta ja mahdollisuudesta sen tekemiseen. Erityiskiitos opinnäytetyön ohjaajalle Mikko Väisäselle, joka asiantuntevasti ohjasi opinnäytetyön suunnittelussa ja työn tekemisessä sekä Pertti Nevalaiselle ohjauksesta ja kannustuksesta. Lisäksi kiitän Maanmittauslaitoksen työntekijöitä Jere Ranista vastauksista tekoälyaineistoihin liittyviin kysymyksiin, Iiro Förstiä avusta päällekkäisyysanalyysiin ja luokitteluun liittyvän menetelmän käytössä, Sampo Rahkoa Pudasjärven rakennusaineiston hankinnasta sekä Riikka Hyttistä kartoittajille tarkoitetun oppaan testaamisesta.

LYHENTEET JA KESKEISET KÄSITTEET

AI-lidar-aineisto

AI-lidar-aineisto on tässä opinnäytetyössä käytetty termi rakennusaineistolle, joka on tekoälyn avulla tuotettu laserkeilausaineistosta.

AI-orto-aineisto

AI-orto-aineisto on tässä opinnäytetyössä käytetty termi rakennusaineistolle, joka on tekoälyn avulla tosiortokuvilta tuotettu.

CSV

Comma Separated Values on tekstitiedosto, johon tallennetaan yksinkertaista taulukkomuotoista tietoa (Wikipedia 2023a).

DWG

Dwg-tiedosto koostuu kaksi- tai kolmiulotteisesta vektorigrafiikasta. Dwg-tiedostoja käytetään CAD (Computer Aided Design) -ohjelmistoissa. (Adobe 2023.)

Ilmakuva

Ilmakuva on kameralla otettu valokuva maastosta, joka on otettu lentokoneesta. Ilmakuva voi olla mustavalkoinen, värillinen tai väärävärinen. (Maanmittauslaitos 2023c.)

Maastotietokanta (MTK)

Maastotietokanta on Maanmittauslaitoksen ylläpitämä maastoa kuvaava aineisto, joka kattaa koko Suomen. Maastotietokanta on avointa aineistoa, jota päivitetään ilmakuvia, keilausaineistoja ja muiden tiedontuottajien aineistoja hyväksi käyttäen. Tiet, rakennukset, hallintorajat, nimistö, maankäyttö, vedet ja korkeussuhteet ovat

sen tärkeimpiä kohderyhmiä. (Maanmittauslaitos 2023f.)

Ortokuva

Ortokuva on yksittäisistä ilmakuvista tehty yhdistelmäkuva, joka vastaa geometrialtaan karttaa (Maanmittauslaitos 2023e). Ortokuvia käytetään maastotietokannan ajantasaistuksessa.

RHR

Rakennus- ja huoneistorekisteri (Digi- ja väestötietovirasto 2023).

Shape

Shapefile on avoin vektoripohjainen tiedostomuoto, jonka sisältämä data koostuu useasta eri tiedostosta (Wikipedia 2023b).

Tosiortokuva

Tosiortokuva on ilmakuva, joka on muodostettu pintamallista. Sen kaikki kohteet on suoraan ylhäältä päin kuvattu. (Avoindata.fi 2023.)

1 JOHDANTO

Tämän opinnäytetyön tavoitteena on selvittää, voiko tekoälyn avulla tuotettuja rakennusaineistoja hyödyntää maastotietokannan rakennusten ylläpidossa. Keskeisenä tavoitteena on tutkia, ovatko tekoälyaineistot riittävän laadukkaita Maastotietokannan laatuvaatimusten näkökulmasta, jotta niitä voidaan hyödyntää rakennusten ajantasaistuksessa. Lisäksi tavoitteena on kehittää menetelmä, jolla kartoittaja löytää kriittiset kohteet tekoälyaineistoista.

Opinnäytetyön toimeksiantaja on Maanmittauslaitos ja Peruspaikkatietopalvelutulosyksikkö (PATI). Aihe valikoitui opiskelijan kiinnostuksen ja aiheen ajankohtaisuuden pohjalta.

Opinnäytetyössä käytetyt tekoälyaineistot on tuotettu Maanmittauslaitoksessa. Käytössä on kaksi eri aineistoa: AI-lidar-aineisto, joka on tuotettu laserkeilausaineistosta sekä AI-orto-aineisto, joka pohjautuu ilmakuvilta tuotettuihin tosiortokuviiin. Myöhemmin mahdollisesti tuotetaan yhdistetty tekoälyaineisto, joka tulee sisältämään sekä laserkeilausaineistosta että ilmakuvilta tuotettuja rakennuksia. Muina aineistoina tässä opinnäytetyössä käytetään maastotietokannan (MTK) rakennuksia sekä pyritään hyödyntämään Pudasjärven kunnan rakennusaineistoa sekä Digi- ja väestötietoviraston rakennus- ja huoneistotietorekisterin (RHR) tietoja. Rakennusaineistoja tarkastellaan 1: 25 000 karttalehden S4333 alueelta. Karttalehti S4333 sijoittuu Pudasjärven kunnan alueelle käsittäen noin puolet keskusta-alueesta sekä harvaan asuttua aluetta keskustasta länteen päin.

Maanmittauslaitoksessa tekoälyillä tuotettujen aineistojen hyödyntäminen maastotietokannan ajantasaistuksessa on ajankohtainen aihe. Vuosien 2021–2022 aikana toteutetussa ATMU2021/es-projektissa on tutkittu ja kehitetty menetelmiä, joita voidaan hyödyntää maastotietojen automaattisessa tuottamisessa. Vuonna 2023 käynnissä olevassa AI4TDB-projektissa jatketaan ATMU2031/es-projektissa kehitettyjen menetelmien kehittämistä ja hyödyntämistä. Tässä opinnäytetyössä pyritään tutkimaan tekoälyn avulla tuotettujen rakennusaineistojen hyödyntämistä peruspaikkatietotuotannon näkökulmasta.

1.1 Tutkimuksen tarkoitus ja tavoite

Tämän opinnäytetyön tarkoituksena on kehittää menetelmä, jonka avulla tekoälyaineistoista voidaan löytää maastotietokannan rakennusten ylläpidossa käsiteltävät rakennuskohteet. Maastotietokannan rakennuksia päivitetään ilmakuvilta sekä kunnilta saatujen rakennusaineistojen avulla. Kansallisen ilmakuvausohjelman mukaisesti vuosittain ilmakuvataan noin 110 000 neliökilometriä, joten alueet ilmakuvataan kolmen vuoden sykleissä (Maanmittauslaitos 2023b). Tämän vuoksi myös ilmakuviin perustuvaa ajantasaistusta tehdään kolmen vuoden kiertoon perustuen. Lisäksi rakennuksia päivitetään maastotietokantaan kunnilta saatujen aineistojen avulla vähintään kerran vuodessa (Maanmittauslaitos 2023g).

Kehitettävän menetelmän avulla pyritään vaikuttamaan maastotietokannan rakennusaineiston laatuun parantamalla rakennusten täydellisyyttä ja sijaintitarkkuutta. Menetelmän avulla tarkastettavat rakennuskohteet voitaisiin mahdollisesti löytää nopeammin ja kattavammin, mikä vähentäisi maastotietokannan päivittämiseen käytettävää aikaa sekä pitäisi rakennusaineiston laatua parempana.

Opinnäytetyössä tutkitaan tekoälyn avulla tuotettujen rakennusaineistojen laatua suhteessa todelliseen tilanteeseen maastossa aineistojen keruuhetkellä. Opinnäytetyössä pyritään selvittämään, onko tekoälyaineistojen avulla mahdollista löytää rakennuskohteita, jotka puuttuvat maastotietokannasta tai kohteita, jotka maastotietokannasta pitäisi poistaa. Lisäksi selvitetään, kuinka suuri osa puuttuvista ja ylimääräisistä kohteista on todellisuudessa olemassa olevia rakennuksia.

Tekoälyn avulla tuotettuja rakennusaineistoja on kaksi: AI-lidar-aineisto ja AI-orto-aineisto. AI-lidar-aineisto on tuotettu laserkeilausdatasta ja AI-orto-aineisto tosiortokuvilta. Tekoälyaineistot ovat Maanmittauslaitoksessa tuotettuja. Molempia aineistoja verrataan maastotietokannan rakennuksiin, millä saadaan selville aineistojen välisiä eroja.

Tavoitteena on myös tuottaa menetelmä, jonka avulla kartoittaja pystyy tekemään rakennusaineistoille päällekkäisyysanalyysin ja rakennusten luokittelun

sekä ohje menetelmän käyttöön. Ohjeessa (Liite 1.) kuvataan menetelmä, jolla aineistoja on tässä opinnäytetyössä vertailtu. Ohjeen avulla pystytään valikoimaan eri aineistosta ne kohteet, joista voisi olla hyötyä maastotietokannan rakennusten päivittämisessä.

1.2 Tutkimuskysymykset

Tämän opinnäytetyön tutkimusongelma on tekoälyn avulla tuotettujen rakennusaineistojen hyödyntäminen maastotietokannan rakennusten ylläpidossa. Ongelmaa lähestytään seuraavien tutkimuskysymysten kautta:

- Kuinka laadukkaita tekoälyllä tuotetut rakennusaineistot ovat
Maastotietokannan laatuvaatimusten näkökulmasta? Löytyykö niiden avulla sellaisia rakennuksia, jotka tulisi lisätä maastotietokantaan tai poistaa maastotietokannasta?
- Millä tavalla tekoälyllä tuotettuja rakennusaineistoja voidaan hyödyntää maastotietokannan rakennusten ylläpidossa?

Opinnäytetyössä kuvataan, kuinka maastotietokannan rakennuksia päivitetään tällä hetkellä, kuinka usein rakennuksia ajantasaistetaan ja mistä aineistoja saadaan. Tämän opinnäytetyön tulosten pohjalta pohditaan keinoja, kuinka maastotietokannan rakennusten ajantasaistusta voitaisiin kehittää. Tutkimuskysymyksiä tarkastellaan peruspaikkatietotuotannon näkökulmasta ja siitä, kuinka kartoittaja tulevaisuudessa voisi hyödyntää tekoälyn avulla tuotettuja aineistoja maastotietokannan päivityksessä.

1.3 Maanmittauslaitoksen organisaatio ja tehtävät

Maanmittauslaitos on valtion virasto, joka on toiminut jo yli 200 vuoden ajan. Sillä on toimipaikkoja 36 paikkakunnalla eri puolilla Suomea ja niissä työskentelee noin 1800 työntekijää. Maanmittauslaitoksen toimintaa ohjaa maa- ja metsätalousministeriö. Maanmittauslaitoksen tehtävänä on ylläpitää rekistereitä kiinteistöjen ja osakehuoneistojen tiedoista sekä pitää huolta kiinnityksistä ja omistuso-

keuksien rekisteröinneistä. Näillä toimilla se turvaa maanomistusta ja luototusjärjestelmää. Lisäksi paikkatiedon tutkimus ja soveltaminen ovat yksi Maanmittauslaitoksen tärkeistä tehtävistä. Asiakkaalle näkyvimpiä Maanmittauslaitoksen tehtäviä ovat lohkomiset, tilusjärjestelyt ja muut maanmittaustoimitukset sekä erilaisien kartta-aineistojen tuottaminen ja niiden käytön edistäminen. (Maanmittauslaitos 2023i.)

Maanmittauslaitoksen organisaation johdossa on pääjohtaja ja sen rinnalla työskentelevä keskushallinto. Pääjohtajasta seuraavana tulee neljä toimintayksikköä, jotka ovat tuotanto, yleishallinto, tietohallinto ja paikkatietokeskus. (Maanmittauslaitos 2023i.)

Tuotannon toimintayksikkö koostuu kahdeksasta tulosityksiköstä. Ne ovat: Uudet kiinteistöt ja maanmittaus (UMA), Lunastus ja arviointi (LUNA), Kiinteistörakenteen kehittämispalvelut (KIPA), Huoneistojen omistuksen palvelut (HUPA), Kiinteistöjen omistuksen palvelut (KOPA), Rekisteripalvelut (REKI), Peruspaikkatietopalvelut (PATI) ja Tietopalvelut (TIETO). Tuotannon tulosityksiköt vastaavat maanomistuksen, huoneistojen, paikannuksen- ja peruspaikkatietopalveluiden tuottamisesta asiakkaille. Paikannuspalvelua tuotetaan myös sisäisiin tarpeisiin. Palvelukokonaisuuksiin kuuluvat kiinteistönmuodostamistehtävät, kiinteistö-, lainhuuto- ja kiinnitysrekisterin ylläpito, paikantamisen perusta yhdessä paikkatietokeskuksen kanssa sekä maastotietokannan ylläpito. (Maanmittauslaitos 2023i.)

Yleishallinnon alla toimii kolme tulosityksikköä: Henkilöstö, Talous ja Viestintä. Niiden tehtävänä on vastata Maanmittauslaitoksen sisäisistä palveluista ja toimia sisäisenä tukipalveluyksikkönä. Yleishallinto mahdollistaa Maanmittauslaitoksen sujuvan arjen sekä korkealuokkaiset ja luotettavat palvelut asiakkaille. (Maanmittauslaitos 2023n.)

Tietohallinto muodostuu kahdesta tulosityksiköstä: Digiasiantuntija- ja -menetelmäpalveluista sekä Digitalisaatoratkaisuista. Tietohallinto kehittää ja ylläpitää laadukkaita ja turvallisia IT-palveluita Maanmittauslaitoksessa. (Maanmittauslaitos 2023j.)

Paikkatietokeskuksen toimintayksikön alla on neljä tulosityksikköä. Ne ovat Geodesia ja geodynamiikka, Geoinformatiikka ja kartografia, Kaukokartoitus ja fotogrammetria sekä Navigointi ja paikannus. Paikkatietokeskus tekee paikkatietoalan tutkimus- ja asiantuntijatyötä ja tarjoaa luotettavaa tietoa yhteiskunnan hyväksi. (Maanmittauslaitos 2023m.)

1.4 Peruspaikkatietopalvelut (PATI)

Peruspaikkatietopalvelut (PATI) -tulosityksikössä huolehditaan peruspaikkatiedoista koko Suomen alueella. Peruspaikkatiedot kerätään ja pidetään ajantasaisina yhteisesti sovittujen ohjeiden mukaan. Peruspaikkatiedot, joita peruspaikkatietopalveluissa tuotetaan, ovat paikannuksen perustiedot, ilmakuvaukset ja laserkeilausaineistot, maastotietokannan tiedot sekä korkeusmalli. Lisäksi kerätyistä aineistoista tuotetaan erilaisia karttatuotteita. (Maanmittauslaitos 2023g.)

Paikannuksen tietoihin kuuluvat kiintopisterekisterin tiedot sekä FinnRef-tukiasemien avulla kerätty havaintodata (Maanmittauslaitos 2023g). Ilmakuvaukset ja laserkeilausaineistot kerätään Kansallisen laserkeilaus- ja ilmakuvauksen ohjelman (KALLIO) mukaisesti. Käytännössä Suomi on jaettu alueisiin, jotka laserkeilataan kuuden vuoden sykleissä ja ilmakuvataan kolmen vuoden sykleissä pohjoisinta Lappia lukuun ottamatta. (Maanmittauslaitos 2023b.) Maastotietokannan maastotietoja kuten tiet, rakennukset, vedet ja maaston kuviotiedot kerätään ja päivitetään muun muassa ilmakuvien ja erilaisten ulkopuolisten aineistojen avulla (Maanmittauslaitos 2023f). Myös paikannimet ovat peruspaikkatietoja (Maanmittauslaitos 2023g). Korkeusmalli (KM2) kuvaa maanpinnan korkeutta ja sen ruutukoko on 2 m x 2 m. Se tuotetaan laserkeilausaineistosta, jonka pistetiheys on vähintään 0,5 pistettä neliömetrille. (Maanmittauslaitos 2023a.)

2 TEKÖÄLYAINEISTOJEN KEHITTÄMISPROJEKTIT

Maanmittauslaitoksessa toteutettiin vuosien 2021–2022 aikana Advanced Technology for National Map Updating (ATMU2021/es) -projekti, jossa tutkittiin ja kehitettiin menetelmiä, joilla voidaan automaattisesti tuottaa maastotietoja. Projektissa kehitetyillä menetelmillä tunnistetaan automaattisesti rakennuksia, tiestöä ja uomaverkostoa laserkeilaus- ja ilmakuva-aineistoista. Lisäksi kehitettiin koneoppimismenetelmiä tiestö- ja rakennuskohteiden automaattiseen muutostulkintaan. Projektissa tuotettiin myös tosiortokuvia, joita käytettiin mallien lähtöaineistoina ja tuotettiin syväoppimismallien kouluttamiseen tarkoitettua opetusaineistoa. (Savola & Zhu 2023, 2.)

Rakennusten automaattisen tunnistamisen osalta, Maanmittauslaitoksen asiantuntijoiden arvion mukaan, ATMU2021/es-projektissa tuotettu menetelmä tunnistaa rakennukset jopa 97,9 % tarkkuudella. Myös rakennusten muutosten automaattisesta tunnistamisesta saatiin hyviä tuloksia. Mukautettu Changer-malli löysi arvion mukaan noin 61 % uusista rakennuksista ja NestNet-neuroverkko-malli 91 % uusista rakennuksista. (Savola & Zhu 2023, 8.)

Tässä opinnäytetyössä tutkitaan ATMU2021/es-projektissa kehitetyillä menetelmillä tuotettuja tekoälyaineistoja verrattuna maastotietokannan rakennuksiin. Koska projektin tulokset vaikuttavat hyviltä, on mielenkiintoista selvittää, kuinka hyvin tekoälyllä tuotetut rakennusaineistot vastaavat maastotietokannan rakennuksia ja onko tekoälyaineistoihin tunnistettu rakennusten muutoskohteita opinnäytetyöhön valikoituneelta alueelta.

Maanmittauslaitoksen vuoden 2023 Enhance the accuracy of topographic database by using the AI (AI4TDB) -projekti jatkaa tekoälyteknologian ja -sovellusten kehittämistä muun muassa maastotietokannan rakennusvektorien tarkkuuden ja ajantasaisuuden parantamisen parissa. Lisäksi projektissa selvitetään, kuinka paljon automatisoinnilla voidaan parantaa tuottavuutta maastotiedon tuotannossa ja voidaanko esimerkiksi kaikki KALLIO (Kansallinen ilmakuvauus- ja laser-

keilaus) -ohjelmassa tuotettava aineisto hyödyntää maastotietokannan ajantasaistuksessa. (Maanmittauslaitos 2023h). Koska AI4TDB-projekti on vielä käynnissä, sen tuloksia ei ole vielä saatavilla.

3 AINEISTOT, MENETELMÄT JA ALUE

3.1 Aineistot

Opinnäytetyössä vertaillaan kahta eri tekoälyllä tuotettua rakennusaineistoa maastotietokannan rakennuksiin. Maastotietokannan rakennukset ovat ajantasaistettu ilmakuviin perustuvan ajantasaistusprosessin mukaisesti vuosien 2022 ja 2023 aikana. Ne ladataan QGIS-sovellukseen rajapinnan kautta. Karttalehden S4333 alueella on yhteensä 3738 maastotietokannan rakennusta.

Tekoälyn avulla tuotetut rakennusaineistot, joita tässä opinnäytetyössä kutsutaan AI-lidar- ja AI-orto-aineistoiksi, on tuotettu Maanmittauslaitoksessa. AI-lidar-aineiston rakennukset on muodostettu laserkeilausaineistosta ja AI-orto-aineiston rakennukset tosiortokuvilta. Opinnäytetyössä vertaillaan myös tekoälyaineistoja toisiinsa, mutta keskitytään enemmän tekoälyaineistojen ja MTK:n rakennusten vertailujen tuloksiin, koska Maanmittauslaitoksessa tuotetaan mahdollisesti myöhemmin yhdistelmäaineisto. Mahdollinen yhdistelmäaineisto tullee sisältämään sekä laserkeilausaineistosta että tosiortokuvilta tekoälyn avulla tuotettuja rakennuksia. Karttalehden S4333 alueella on AI-lidar-aineistossa 2886 ja AI-orto-aineistossa 3043 rakennuskohdetta. Aineistot ovat käytettävissä shape-formaattissa.

Maanmittauslaitoksen maastotietokohteet-tiedonkeruuohjeen mukaan maastotietokantaan tallennetaan kaikki pinta-alaltaan vähintään 10 neliömetrin suuruiset rakennukset. Pienimmistä rakennuksista voidaan tallentaa käytöltään erityisen merkittävät tai ympäristöstään erityisen hyvin erottuvat rakennukset sekä kaikki lentoesteiksi luettavat rakennukset. (Maanmittauslaitos 2023d, 78.) MTK:n rakennukset-aineisto sisältääkin 67 alle 10 neliömetrin rakennusta. AI-lidar-aineistossa niitä on kymmenen ja AI-orto-aineistossa kahdeksan.

Tekoälyaineistot on tuotettu menetelmin, jotka kehitettiin Maanmittauslaitoksen ja paikkatietokeskuksen yhteisessä Advanced Technology for National Map Updating (ATMU2021/es) -projektissa. Projektissa luotiin menetelmät paitsi rakennusten, myös tiestön ja uomaverkoston automaattiseen tunnistamiseen laserkei-

laus- ja ilmakuva-aineistosta. Lisäksi kehitettiin menetelmä automaattiseen muu-
tostulkintaan koneoppimismenetelmillä rakennus- ja tiestökohteiden osalta. (Sa-
vola & Zhu 2023, 2.) Tekoälyaineistojen rakennusvektorit on puolestaan tuotettu
Maanmittauslaitoksen käynnissä olevassa Maastotietokannan tarkkuuden paran-
taminen tekoälyteknologian avulla (AI4TDB) -projektissa (Maanmittauslaitos
2023h).

Työssä on tarkoitus hyödyntää rakennus- ja huoneistorekisterin tietoja, jotka on
ladattu Digi- ja väestötietoviraston avoindata.fi-sivulta. Tiedot tuodaan QGIS-so-
vellukseen CSV-tiedostona ja ne sisältävät rakennusten osoitteet, postinumerot
ja WGS84-koordinaatit.

Opinnäytetyötä varten on saatu myös Pudasjärven kunnan pohjakartan raken-
nukset Kurenalan asemakaavan alueelta shape- ja dwg-formaatissa. Aineisto
kattaa vain osan S4333 karttalehden alueesta, joten se jätettiin tämän tutkimuk-
sen ulkopuolelle.

3.2 Menetelmät

Tutkimusaihetta lähestytään kvantitatiivisten eli määrällisten menetelmien kautta.
Aineistoille tehdään päällekkäisyysanalyysi, jonka jälkeen rakennukset luokitel-
laan eri luokkiin sen mukaan, kuinka paljon päällekkäistä pinta-alaa eri aineisto-
jen rakennuksilla on. Lopuksi menetelmän käytöstä tehdään ohje (Liite 1.), jonka
avulla tekoälyaineistosta voidaan poimia kohteet, jotka ovat avuksi maastotieto-
kannan rakennusten ajantasaistuksessa.

Rakennusaineistojen vertailuun käytetään QGIS-ohjelmaa. Siinä eri aineistoille
tehdään ristiin vertailu päällekkäisyysanalyysillä. Ristiin vertailut tehdään MTK:n
rakennusten ja AI-lidar-aineiston, MTK:n rakennusten ja AI-orto-aineiston sekä
AI-lidar- ja AI-orto-aineistojen kesken. Rakennukset luokitellaan neljään eri luok-
kaan päällekkäin osuvien rakennusten yhteisen pinta-alan mukaan taulukon 1
mukaisesti.

Taulukko 1. Rakennusten luokittelu päällekkäisen pinta-alan mukaan

Luokka	Kuvaus
80–100 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa vähintään 80 prosenttisesti päällekkäin.
50–80 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 79,99 prosenttisesti ja vähintään 50 prosenttisesti päällekkäin.
10–50 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 49,99 prosenttisesti ja vähintään 10 prosenttisesti päällekkäin.
0–10 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 9,99 prosenttisesti päällekkäin.

Kuviolla 1 havainnollistetaan aineistojen ristiin vertailua. Vasemmalla puolella rakennuksille on tehty päällekkäisyysanalyysi, jossa syötetaso on MTK:n rakennukset (oranssi) ja peitetasona AI-orto-aineiston rakennukset (violetti). AI-orto-rakennus peittää MTK:n rakennuksen pinta-alasta 44 prosenttia, joten MTK:n rakennus sijoittuu luokkaan 10–50 %. Kuvion 1 oikealla puolella syötetasona on AI-orto-rakennukset (vihreä) ja peitetasona MTK:n rakennukset (sininen). MTK:n rakennus peittää AI-orto-rakennuksen pinta-alasta 96 prosenttia, joten AI-orto-rakennus sijoittuu luokkaan 80–100 %.

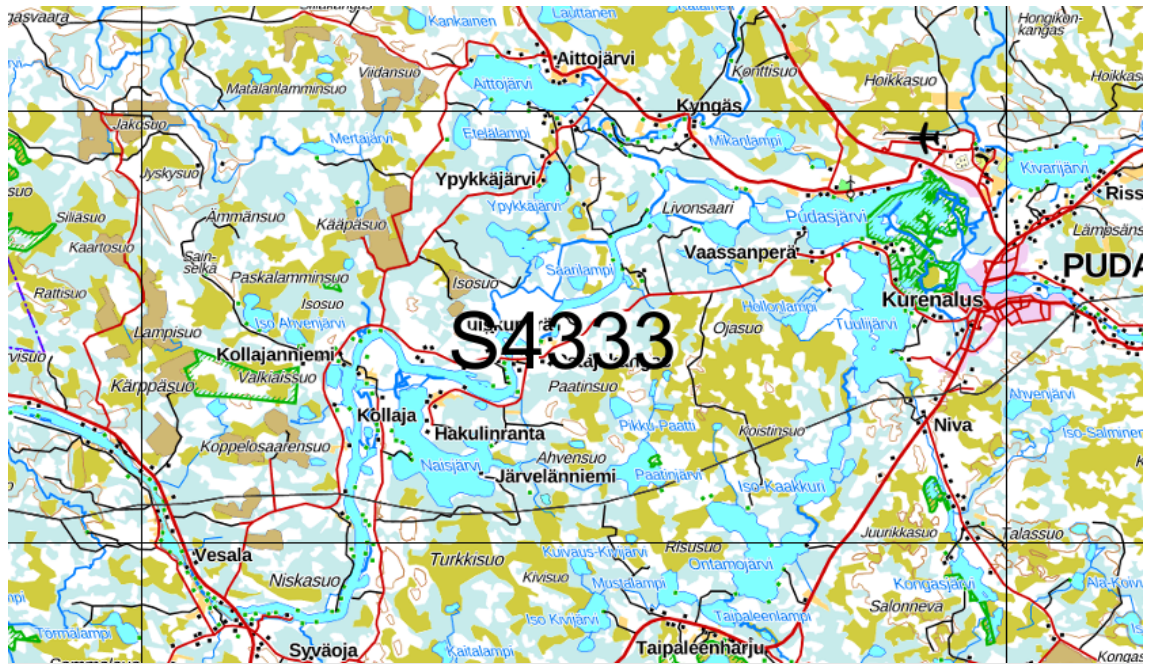


Kuvio 1. Esimerkki ristiin vertailusta. Vasemmalla puolella on päällekkäisyysanalyysi MTK:n rakennukset vs AI-orto-rakennukset ja oikealla puolella AI-orto-rakennukset vs MTK:n rakennukset.

Rakennusaineistojen päällekkäisyysanalyysit ja luokittelu on kuvattu tarkemmin liitteessä 1. Ohje on tarkoitettu helpottamaan tekoälyn avulla tuotettujen rakennusaineistojen käsittelyä niin, että niitä voidaan hyödyntää maastotietokannan rakennusten ylläpidossa. Ohjetta on myös helppo soveltaa, jos on tarpeellista esimerkiksi muuttaa rakennusten luokittelun rajoja.

3.3 Alue

Opinnäytetyössä tarkastellaan rakennusaineistoja 1:25 000 karttalehden S4333 (Kuvio 2) alueelta. Se on pinta-alaltaan 288 neliökilometriä. Karttalehti S4333 sijoittuu Pudasjärven tuotantoalueelle, joka on sekä ilmakuvaattu että laserkeilattu kesällä 2022. Alueella tehtiin ilmakuviin perustuva ajantasaistus vuosien 2022 ja 2023 aikana. (Maanmittauslaitos 2023k.) Opinnäytetyön lähtöoletuksena on, että maastotietokannan rakennukset ovat ajan tasalla. Karttalehti S4333 käsittää noin puolet Pudasjärven kunnan keskustasta sekä harvaan asuttua aluetta mukaan lukien vesistöjen rannoilla sijaitsevia vapaa-ajan rakennuksia.



Kuvio 2. Karttalehti S4333, jonka alueelta rakennusaineistoja käsitellään.

Alue valikoitui opinnäytetyöhön, koska Pudasjärven tuotantoalue oli ainoa, josta oli saatavilla laadultaan riittävän hyvää tekoälyn avulla tuotettuja rakennusaineistoja. Pudasjärven tuotantoalueen on keilannut TerraTec Oy ja ilmakuvaannut SKM Gisair Oy (Maanmittauslaitos 2023k).

4 TULOKSET

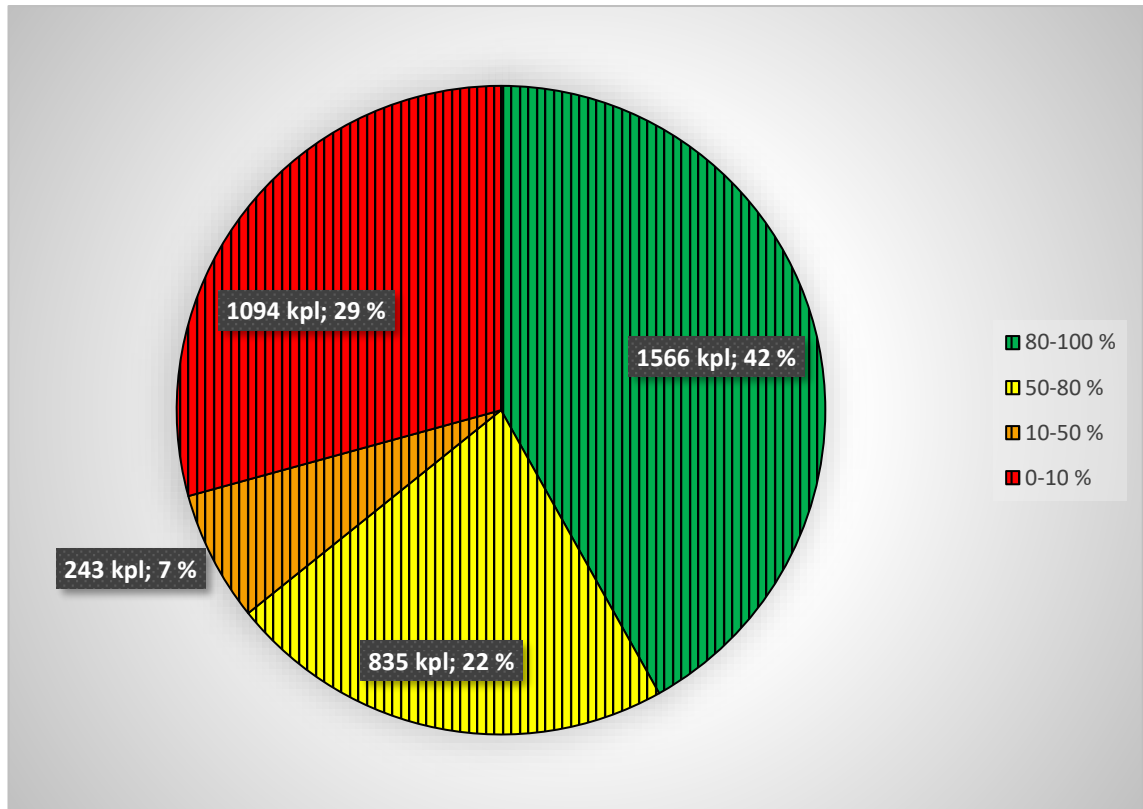
4.1 Aineistojen vertailu päällekkäisyysanalyysillä

Rakennusaineistoille tehdään ristiin vertailu QGIS-ohjelman päällekkäisyysanalyysillä. Syötetason rakennukset jaetaan neljään eri luokkaan (taulukko 1) sen mukaan, kuinka paljon rakennuksella on päällekkäistä pinta-alaa peitetason rakennuksen kanssa.

Luvuissa 4.1.1–4.1.5 on kuvattu päällekkäisyysanalyysin tulokset eri aineistojen kesken. Otsikoiden ensimmäinen rakennusaineisto on päällekkäisyysanalyysin syötetaso ja toinen aineisto on peitetaso.

4.1.1 MTK:n rakennukset vs AI-lidar-rakennukset

Kuvio 3 kuvaa päällekkäisyyssanalyysin, jossa syötetasona on MTK:n rakennukset ja peitetasona on AI-lidar-rakennukset, tuloksia. Siitä nähdään, että 42 % (1566 kpl) MTK:n rakennuksista löytävät vastinrakennuksen AI-lidar-aineistosta vähintään 80 prosenttisesti. Tarkasteltaessa luokkaa 0–10 %, huomataan, että jopa 29 % (1094 kpl) MTK:n rakennuksista ei käytännössä ole löytynyt AI-lidar-aineistosta. Vertailtavissa aineistoissa on 3738 MTK:n rakennusta ja 2886 AI-lidar-rakennusta. Väliluokkiin 50–80 % ja 10–50 % kuuluu yhteensä 29 % MTK:n rakennuksista.



Kuvio 3. MTK:n rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun peitetasona on AI-lidar-rakennukset

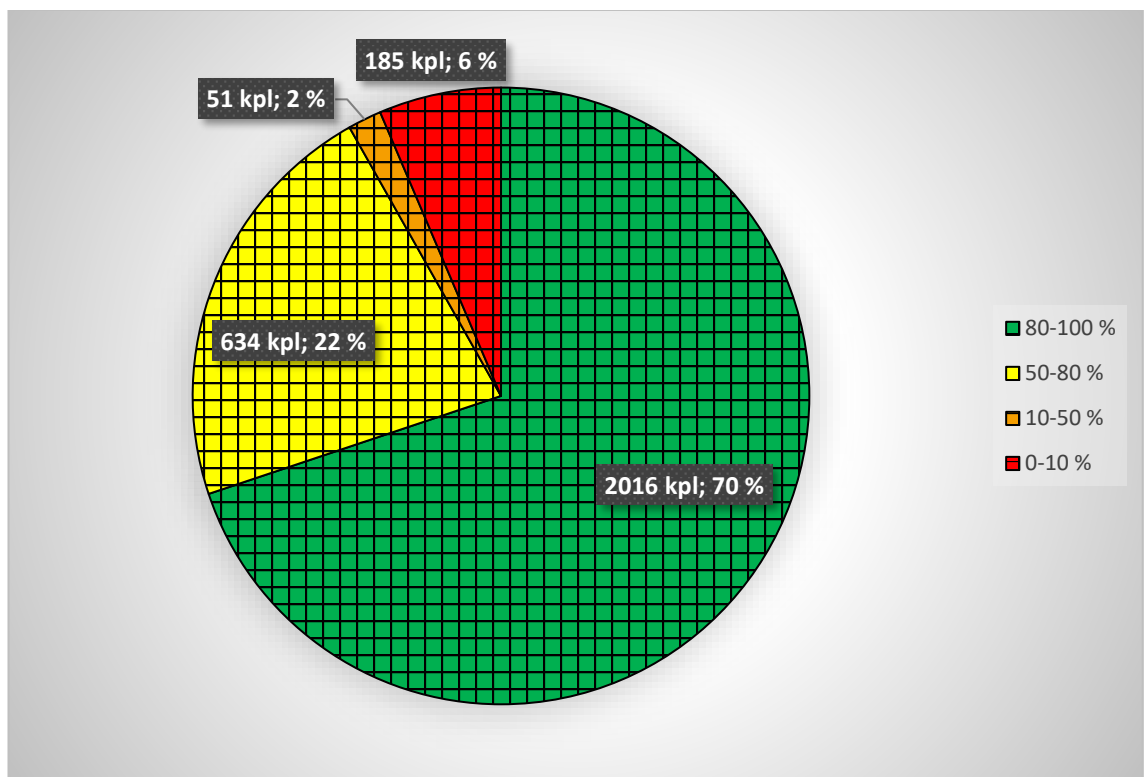
Kuviolla 4 havainnollistetaan AI-lidar-rakennuksia suhteessa MTK:n rakennuksiin. Kuvion 4 yläosassa on ortokuva. Alaosassa näkyy MTK:n rakennukset punaisella, keltaisella, oranssilla ja vihreällä taulukon 1 jaottelun mukaisesti ja AI-lidar-aineiston rakennukset ruskealla. Punaisia rakennuksia ei ole tunnistettu AI-lidar-aineistossa ja oranssista rakennuksesta on tunnistettu 10–50 %. Vihreän rakennuksen muoto on tunnistettu erilaiseksi kuin MTK:n rakennuksen muoto, vaikka pinta-ala kattaa MTK:n rakennuksen vähintään 80 prosenttisesti.



Kuvio 4. AI-lidar-rakennukset ruskealla ja MTK:n rakennukset vihreällä, keltaisella, oranssilla ja punaisella.

4.1.2 AI-lidar-rakennukset vs MTK:n rakennukset

Kun tarkastellaan tuloksia, jossa aineistojen vertailu tehdään toisin päin eli päällekkäisyysanalyysin syötetasona on AI-lidar-rakennukset ja peitetasona MTK:n rakennukset (kuvio 5) huomataan, että 70 % (2016 kpl) AI-lidar-aineiston rakennuksista saa vastaavuuden MTK:n rakennuksesta vähintään 80 prosenttisesti. Luokkaan 0–10 % kuuluu 6 % (185 kpl) AI-lidar-rakennuksista. Näille rakennuksille ei löydy vastinrakennusta MTK:sta. Väliluokkiin 50–80 % ja 10–50 % sijoittuu 24 % AI-lidar-aineiston rakennuksista.



Kuvio 5. AI-lidar-rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun peitetasona on MTK:n rakennukset

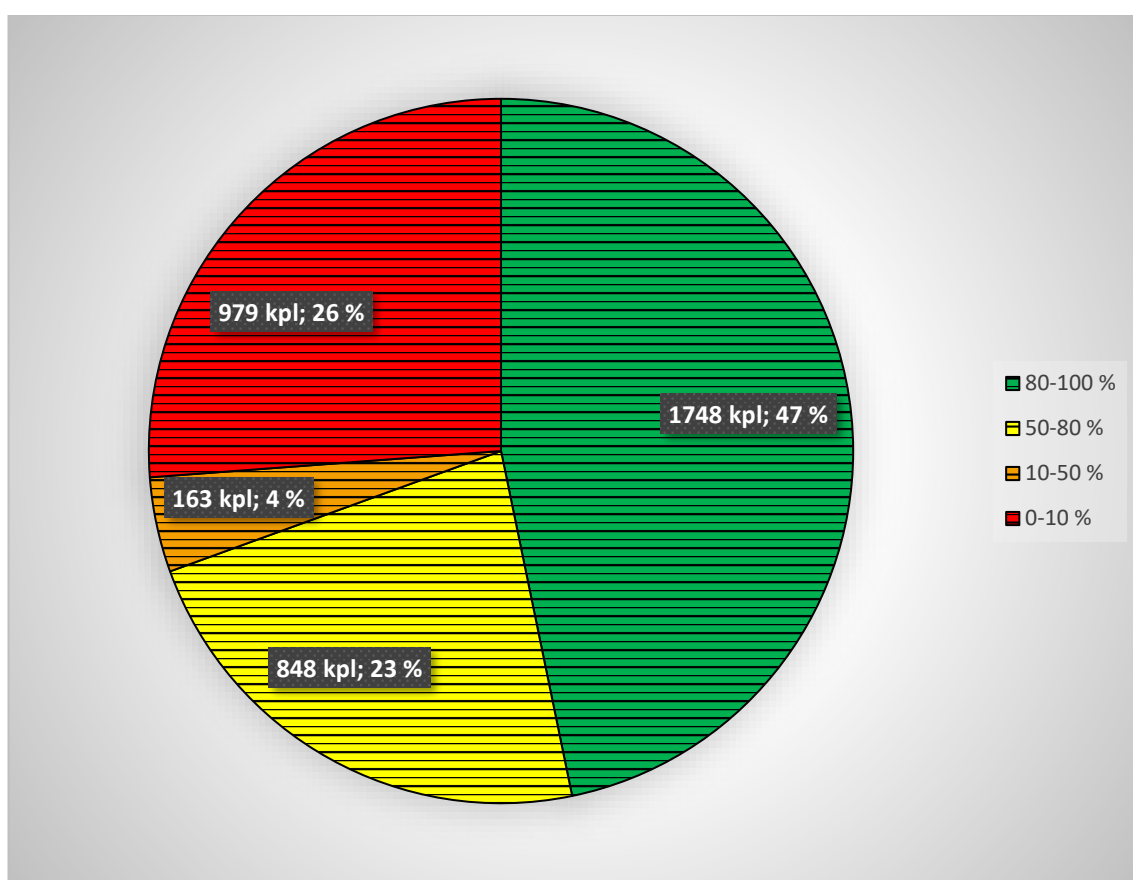
Kuvion 6 yläosassa näkyy ortokuva ja alaosassa MTK:n rakennukset sinisellä sekä AI-lidar-aineiston rakennuskohteet punaisella, oranssilla, keltaisella ja vihreällä. AI-lidar-aineistoon on tunnistettu yksi maastotietokannasta puuttuva todellinen rakennus (iso, punainen rakennus kuvion 6 alareunassa). Lisäksi AI-lidar-aineistossa rakennuksiksi on tunnistettu monta pientä kohdetta, jotka todellisuudessa eivät ole rakennuksia. Kuvion 6 alareunassa näkyy keltaisella AI-lidar-rakennus, joka on osittain tunnistanut MTK:n rakennuksen.



Kuvio 6. MTK:n rakennukset sinisellä ja AI-lidar-rakennukset vihreällä, keltaisella, oranssilla ja punaisella.

4.1.3 MTK:n rakennukset vs AI-orto-rakennukset

Seuraavaksi vertaillaan MTK:n ja AI-orto-aineistojen rakennuksia. Kuvio 7 osoittaa, että vihreään 80–100 %-luokkaan kuuluu 47 % (1748 kpl) MTK:n rakennuksista. Punaiseen 0–10 %-luokkaan kuuluu 26 % (979 kpl) MTK:n rakennuksista. Voidaankin todeta, että tekoälyn avulla tuotetusta aineistosta ei löydy noin neljäsosaa MTK:n rakennuksista. Luokkiin 50–80 % ja 10–50 % kuuluu yhteensä 27 % MTK:n rakennuksista, mikä on hyvin lähellä MTK:n rakennukset vs. AI-lidar-rakennukset-vertailun tulosta 29 %.



Kuvio 7. MTK:n rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun peitetasona on AI-orto-rakennukset

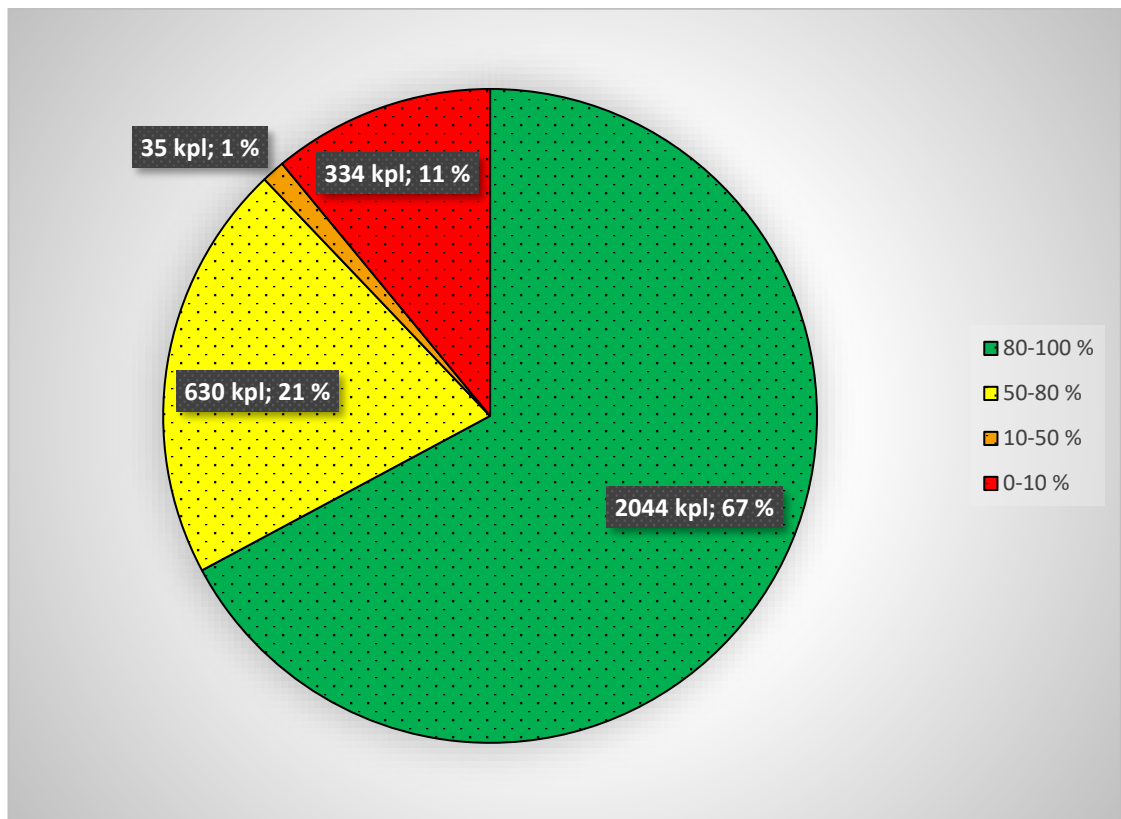
Kuviossa 8 on ortokuvan lisäksi AI-orto-aineiston rakennukset violetilla ja MTK:n rakennukset vihreällä, keltaisella, oranssilla ja punaisella. Siitä voidaan todeta, että AI-orto-aineistossa on vaihtelevasti tunnistettu kuvassa näkyvät kasvihuoneet.



Kuvio 8. AI-orto-rakennukset violetilla ja MTK:n rakennukset vihreällä, keltaisella, oranssilla ja punaisella.

4.1.4 AI-orto-rakennukset vs MTK:n rakennukset

Kuvio 9 näyttää päällekkäisyysanalyysin tulokset, kun syötetasona on AI-orto-aineiston rakennukset ja peitetasona MTK:n rakennukset. Luokkaan 80–100 % sijoittuu 67 % (2044 kpl) AI-orto-aineiston rakennuksista, mikä on lähellä AI-lidar-aineiston vastaavaa tulosta 70 % (kuvio 5). Luokassa 0–10 % on 11 % (334 kpl) AI-orto-aineiston rakennuksista, joille ei löydy vastinrakennusta MTK:sta. Väli-luokkiin 50–80 % ja 10–50 % sijoittuu 22 % AI-orto-aineiston rakennuksista.



Kuvio 9. AI-orto-rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun peitetasona on MTK:n rakennukset

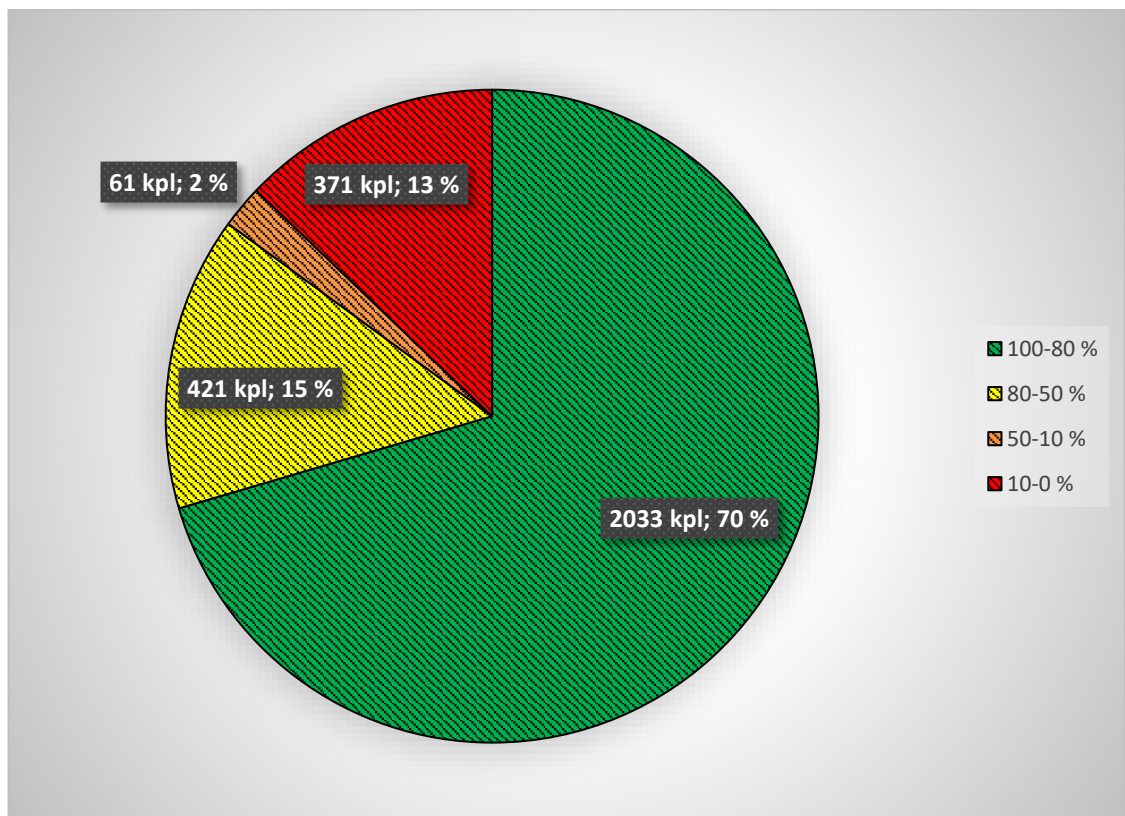
Kuvio 10 kuvapari on samasta kohtaa kuin kuviossa 6. Tässä kuviossa näkyy MTK:n rakennukset sinisellä ja AI-orto-aineiston rakennukset vihreällä, keltaisella ja punaisella. Myös AI-orto-aineistoon on tunnistettu maastotietokannasta puuttuva rakennus niin kuin AI-lidar-aineistossakin. Lisäksi ortokuvalla näkyviä puutavarakasoja on virheellisesti tunnistettu rakennuksiksi. Molemmissa tekoälyaineistoissa oli myös tunnistettu keltaisen rakennuksen lisäsiipi, joka on kuvion 10 alareunassa.



Kuvio 10. MTK:n rakennukset sinisellä ja AI-orto-aineiston rakennukset vihreällä, keltaisella ja punaisella.

4.1.5 Tekoälyaineistojen ristiin vertailu

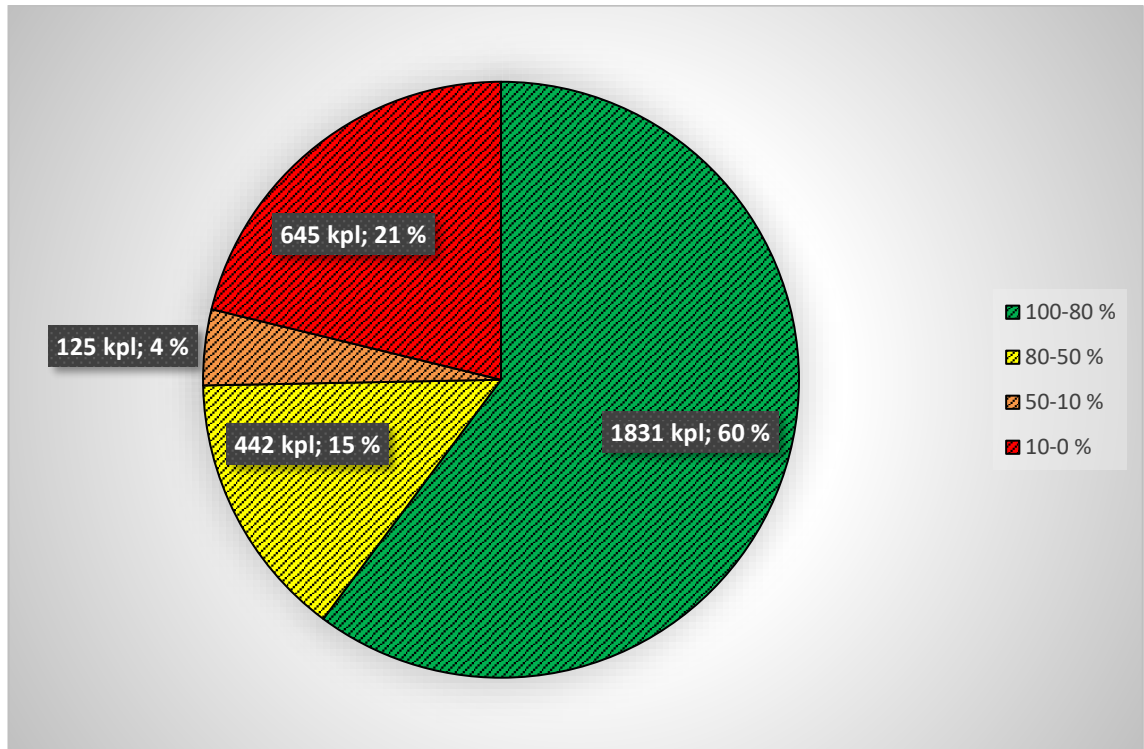
Kuvioissa 11 ja 12 on kuvattu AI-lidar-aineiston ja AI-orto-aineiston rakennusten luokittelu, kun aineistoille on tehty ristiin vertailu päällekkäisyysanalyysillä. AI-lidar-aineistossa rakennuksia on yhteensä 2886 kappaletta ja AI-orto-aineistossa 3043 kappaletta eli AI-orto-aineistossa on 157 rakennuskohdetta enemmän kuin AI-lidar-aineistossa.



Kuvio 11. AI-lidar-rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun peitetasona on AI-orto-rakennukset

AI-lidar-rakennukset vs AI-orto-rakennukset päällekkäisyysanalyysin jälkeen luokassa 0–10 % on 13 % (371 kpl) AI-lidar-rakennuksista (kuvio 11). AI-orto-rakennukset vs AI-lidar-rakennukset päällekkäisyysanalyysin jälkeen luokassa 0–10 % on 21 % (645 kpl) AI-orto-rakennuksista (kuvio 12). Tämä voi osittain selittyä sillä, että AI-orto-rakennuksia on 157 kappaletta enemmän kuin AI-lidar-rakennuksia. Muissa luokissa prosenttiosuudet ovat melko lähellä toisiaan: 80–100 %-luokassa 70 % ja 60 %, 50–80 %-luokassa 15 % ja 15 % sekä 10–50 %-luokassa 2

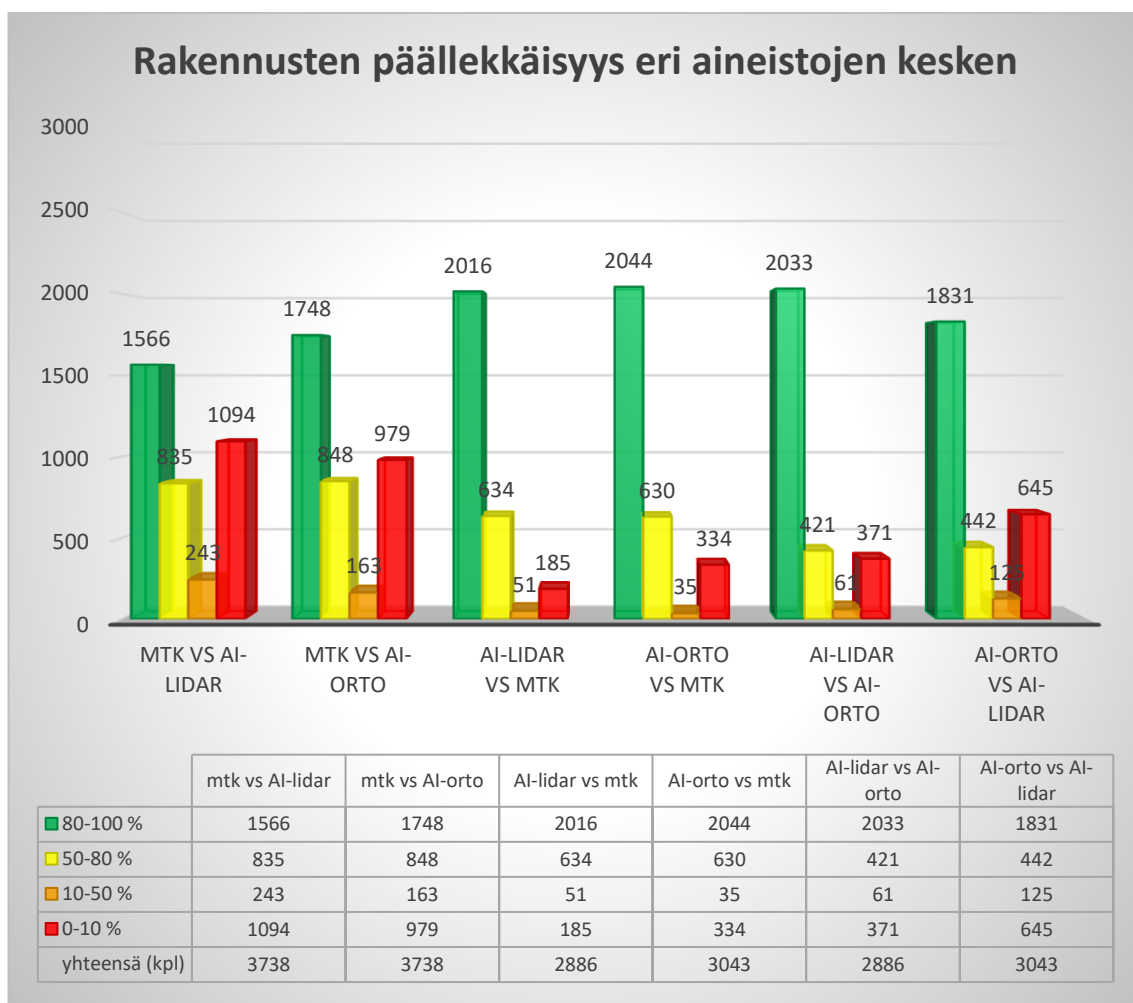
% ja 4 %, kun ensin on kerrottu AI-lidar-rakennukset vs AI-orto-rakennukset-ver-
tailun tulokset.



Kuvio 12. AI-orto rakennusten luokittelu päällekkäisyysanalyysin jälkeen, kun pei-
tetasona on AI-lidar-rakennukset

4.1.6 Rakennusten päällekkäisyys verrattuna eri aineistojen kesken

Kuviossa 13 on kuvattu kaikkien päällekkäisyysanalyysien tulokset niin, että ra-
kennusten määrät ovat luokittain omina pylväinä. Koska rakennusaineistoissa on
eri määrä rakennuksia, analyysien tulokset eivät ole suoraan verrattavissa toi-
siinsa. Rakennuskohteiden eri määrä aineistoissa vaikuttaa pylväiden eroihin.



Kuvio 13. Rakennusten päällekkäisyys eri aineistojen kesken

Luvussa 4.2. kerrotaan tarkemmin, kuinka luokan 0–10 % rakennukset eri aineistoissa ovat jakautuneet. Käytännössä MTK:n, AI-lidar- ja AI-orto-aineistojen kaikki rakennukset käytiin yksitellen läpi ja jaoteltiin eri luokkiin. Koska aineistojen kaikki rakennukset käytiin läpi, voidaan todeta, että AI-lidar ja AI-orto-aineistossa tunnistettiin kohteita eri tavalla, mikä näkyy myös AI-lidar vs AI-orto ja AI-orto vs AI-lidar rakennuskohteiden määrässä luokassa 0–10 % (kuvio 13.). Eri kohteiden tunnistamisen voi huomata myös vertaamalla kuviota 6 ja 10.

4.2 Rakennusten jakautuminen 0–10 %-luokassa

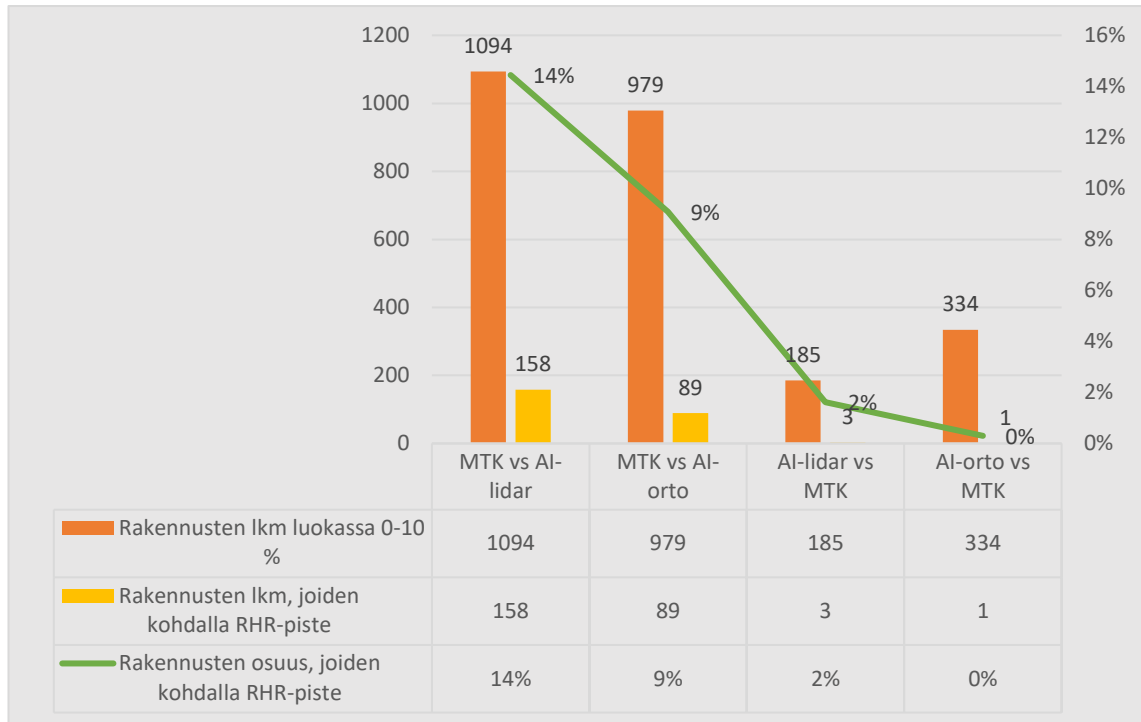
Seuraavissa luvuissa esitellään rakennusten jakautuminen päällekkäisyysanalyysin jälkeen luokassa 0–10 %, jossa peitetason rakennus peittää syötetason rakennuksen enintään kymmenen prosenttisesti. Jokainen rakennus luokassa 0–

10 % on käyty läpi ja luokiteltu taulukon 2 mukaisesti eri luokkiin. Jokaisen rakennuksen läpikäynti yksitellen osoitti, että käytännössä lähes kaikki kohteet olivat sellaisia, joille vastinpinta-alaa toisesta aineistoista ei ollut lainkaan.

Taulukko 2. Luokan 0–10 % kohteiden luokittelu

Luokka	Kuvaus
On rakennus	Rakennus, jonka voi ortokuvalta havaita ja tunnistaa rakennukseksi.
Ei ole rakennus	Kohde, joka näkyy ortokuvalla, mutta ei ole rakennus.
Ei näy ortolla	Kohde, jota ei pysty ortokuvalla havaitsemaan esim. puuston vuoksi.
Eos	Kohde, joka näkyy ortokuvalla, mutta ei voi varmuudella tunnistaa rakennukseksi.

Luokan 0–10 % rakennuksia läpi käydessä seurattiin myös Digi- ja väestötietoviraston rakennus- ja huoneostotietopisteitä (RHR-piste). RHR-pisteen tarkoituksena oli osoittaa, että kohde on todennäköisemmin rakennus. RHR-pisteistä ei kuitenkaan ollut hyötyä, koska niitä osui rakennuskohteiden kohdalle vähän (kuvio 14). Tämän vuoksi RHR-pisteiden merkittävyys jätettiin huomioimatta. Luokan 0–10 % rakennusten, joiden kohdalle RHR-piste osui, lukumäärät ja osuudet on kuitenkin kerrottu seuraavissa luvuissa.

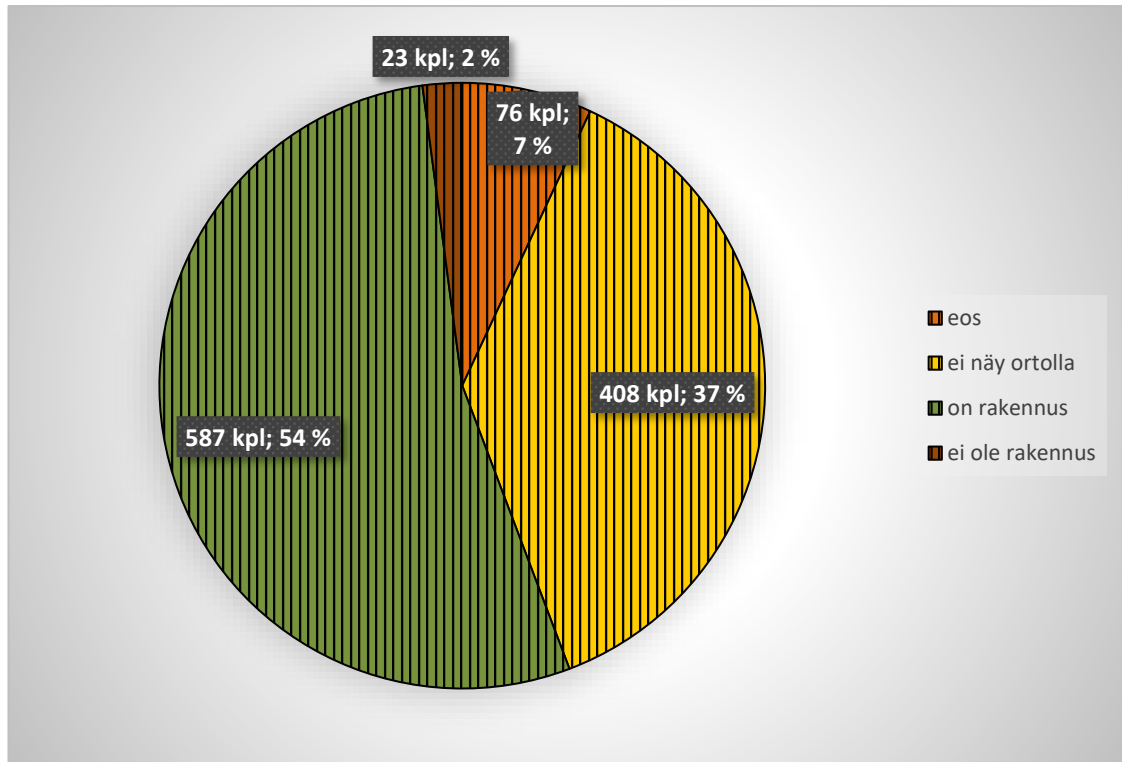


Kuvio 14. RHR-pisteen osuminen luokan 0–10 % rakennuksen kohdalle

4.2.1 MTK:n rakennukset vs AI-lidar-aineiston rakennukset

MTK:n rakennukset vs AI-lidar-rakennukset päällekkäisyysanalyysin tuloksena 0–10 %-luokassa on 1094 MTK:n rakennusta. Käytännössä AI-lidar-aineistossa ei ollut tunnistettu tämän luokan MTK:n rakennuksia. Rakennukset käytiin yksitellen läpi ja jaettiin taulukon 2 mukaisesti neljään eri luokkaan sen mukaan, mitä ortokuvalta pystyi silmäämäärisesti tulkitsemaan.

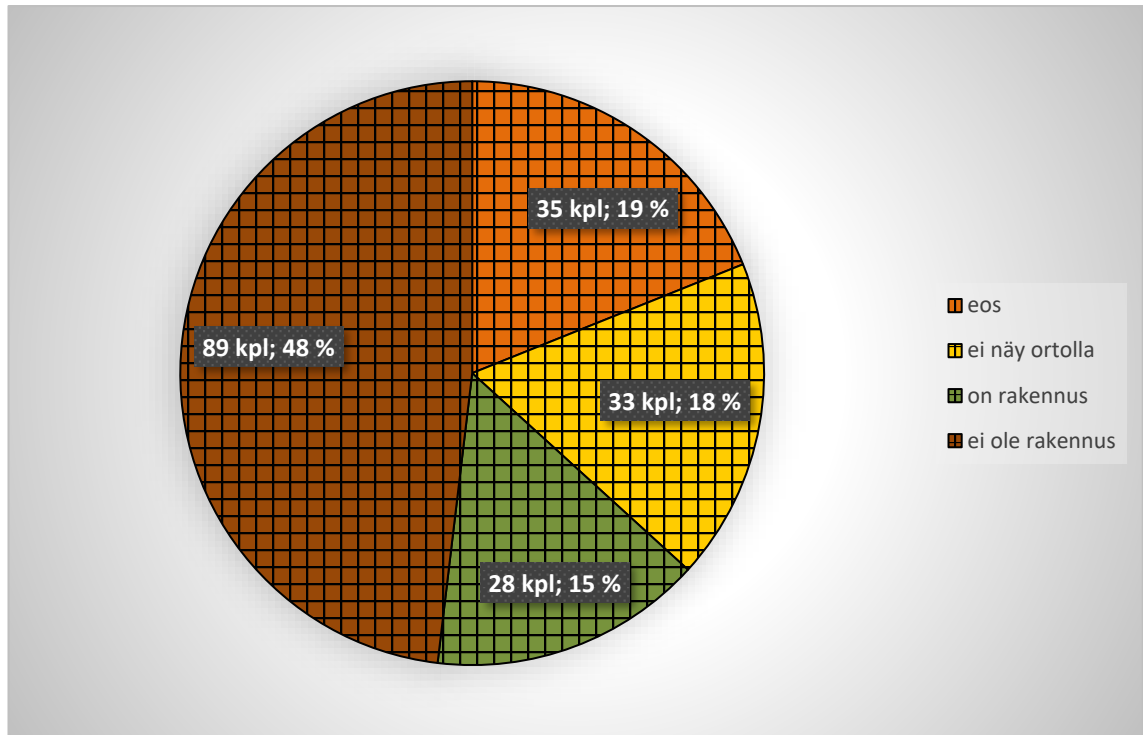
Kuvio 15 osoittaa, että yli puolet, 54 % (587 kpl) 0–10 %-luokan MTK:n rakennuksista todella oli rakennuksia, joita tekoäly ei ollut tunnistanut rakennuksiksi. Rakennuksia, joiden kohdalle osuu RHR-piste, oli kuitenkin vain 158 kappaletta, mikä on 14 % kaikista luokan rakennuksista (kuvio 14). 2 % (23 kpl) rakennuksista ei ollut ortokuvan mukaan rakennuksia. Näissä kohteissa oli mukana purettuja rakennuksia ja kohteita, joissa ortokuvalla näkyi perustukset, mutta ei vielä rakennusta. 37 % (408 kpl) rakennuksista ei pystynyt ortokuvalla havaitsemaan ja 7 % (76 kpl) kohteista ei pystynyt varmuudella määrittämään rakennuksiksi.



Kuvio 15. Päällekkäisyysanalyysi MTK:n rakennukset vs AI-lidar-rakennukset: 0–10 %-luokan rakennusten jakautuminen

4.2.2 AI-lidar-rakennukset vs MTK:n rakennukset

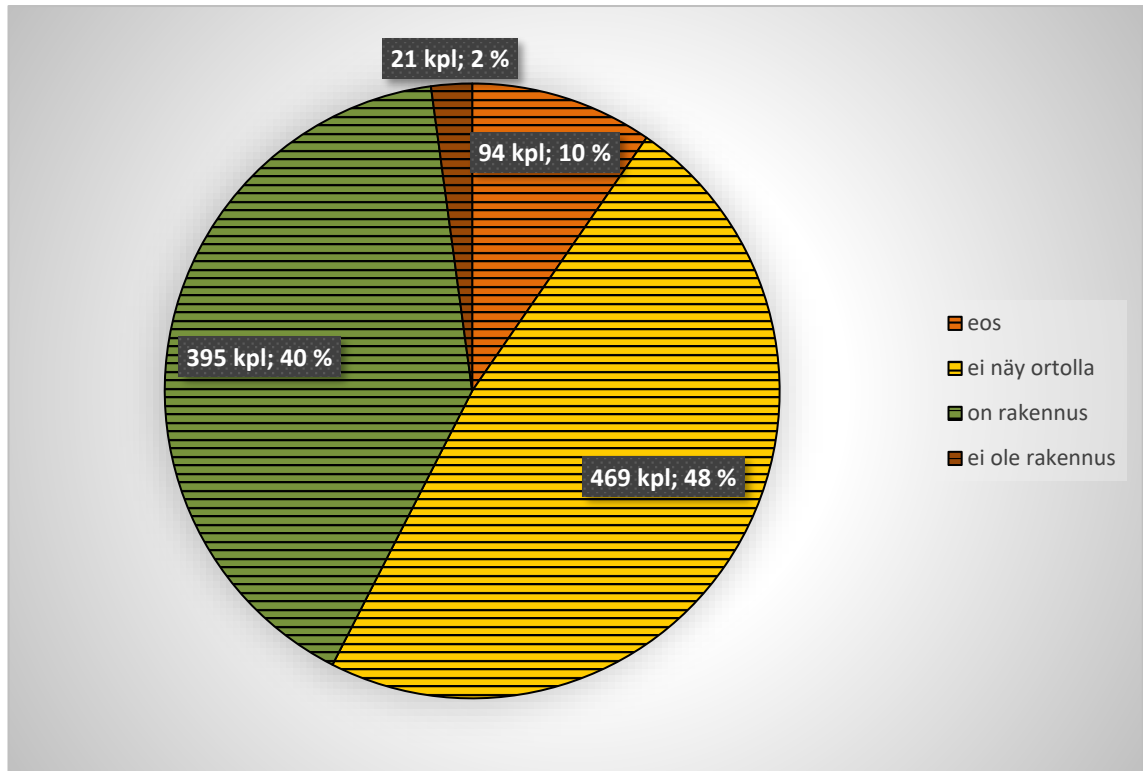
Kuvio 16 osoittaa päällekkäisyysanalyysin AI-lidar-rakennukset vs MTK:n rakennukset luokan 0–10 % rakennusten jakautumisen. Yhteensä 185 rakennuksesta 48 % oli sellaisia kohteita, jotka eivät olleet rakennuksia. Tekoäly oli luokitellut rakennuksiksi muun muassa puupinoja, puutavarakasoja, kontteja ja rekan. 15 % (28 kpl) oli todellisia rakennuksia, jotka puuttuvat maastotietokannasta. RHR-piste osui kolmen rakennuksen kohdalle (kuvio 14). 19 % (35 kpl) kohteista oli ortokuvalta mahdotonta luokitella rakennukseksi. Kohteista 18 % (33 kpl) ei pystynyt havaitsemaan ortokuvalta.



Kuvio 16. Päällekkäisyysanalyysi AI-lidar-rakennukset vs MTK:n rakennukset: 0–10 %-luokan rakennusten jakautuminen

4.2.3 MTK:n rakennukset vs AI-orto-rakennukset

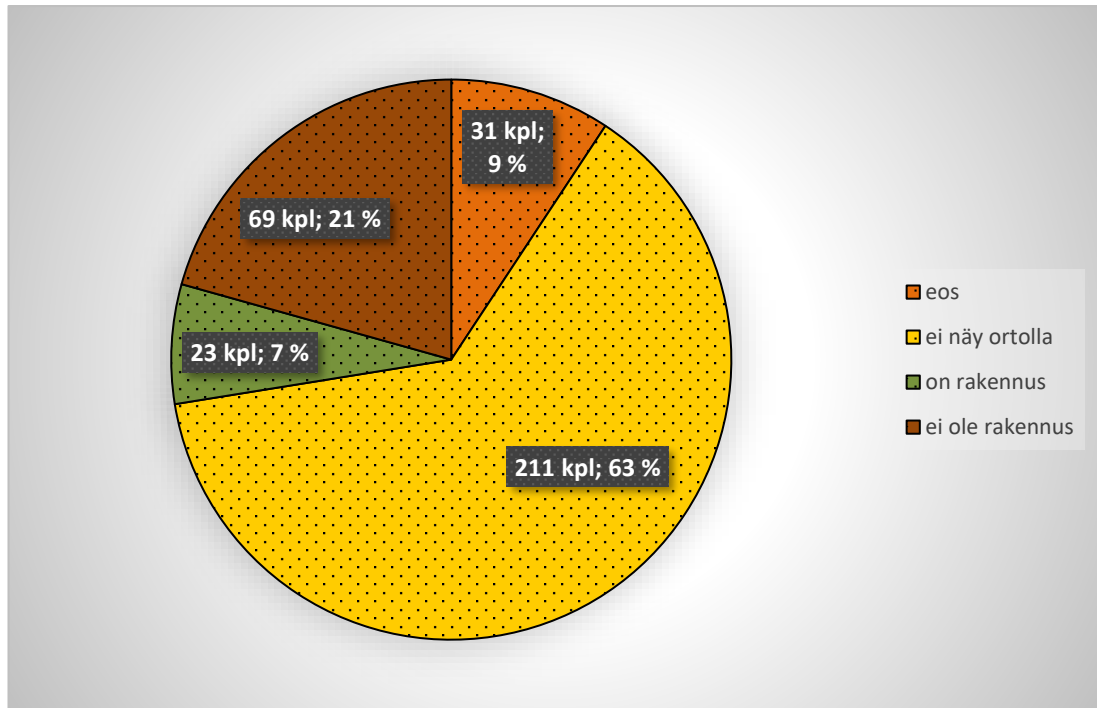
Kun päällekkäisyysanalyysin MTK:n rakennukset vs AI-orto-rakennukset luokan 0–10 % rakennukset luokitellaan, huomataan kuviosta 17, että 40 % (395 kpl) MTK:n rakennuksista ei ole tunnistettu. 9 % eli 98 rakennuksen kohdalla oli RHR-piste (kuvio 14). Yhteensä rakennuksia tässä luokassa on 979 kappaletta. 2 % (21 kpl) kohteista ei ole rakennuksia, vaan purettuja rakennuksia sekä rakennusten perustuksia. 48 % (469 kpl) kohteista ei näy ortokuvalla ja kohteista 10 % (94 kpl) on sellaisia, joita ortokuvalta on varmasti rakennukseksi vaikea tulkita.



Kuvio 17. Päällekkäisyysanalyysi MTK:n rakennukset vs AI-orto-rakennukset: 0–10 %-luokan rakennusten jakautuminen

4.2.4 AI-orto-rakennukset vs MTK:n rakennukset

Päällekkäisyysanalyysi AI-orto-rakennukset vs MTK:n rakennukset luokan 0–10 % rakennukset jakautuivat kuvion 18 mukaisesti. Yhteensä rakennuksia on 334 kappaletta, joista 63 % (211 kpl) ei näy ortokuvalla. AI-orto-aineistosta löytyy 7 % (23 kpl) rakennuksia, joita ei ole maastotietokannassa ja vain yhden kohdalla oli RHR-piste (kuvio 14). 9 % (31 kpl) on rakennuksia, joita ortokuvalla ei voi varmasti rakennukseksi tulkita. Kohteista 21 % (69 kpl) ei ollut rakennuksia. Näissä kohteissa oli esimerkiksi puupinoja, puutavarakasoja sekä kontteja.



Kuvio 18. Päällekkäisyysanalyysi AI-orto-rakennukset vs MTK:n rakennukset: 10–0 %-luokan rakennusten jakautuminen

4.3 Tekoälyaineistojen geometriat

AI-lidar- ja AI-orto-rakennusten geometrioita tarkasteltiin silmämääräisesti. Kuvio 19 osoittaa, että tekoälyaineistojen geometriat saattavat poiketa huomattavasti maastotietokannan rakennusten geometrioista. Jotkut rakennukset ovat tekoälyaineistossa tunnistettu useammaksi rakennukseksi.



Kuvio 19. MTK:n rakennukset sinisellä, AI-orto-rakennukset violetilla ja AI-lidar-rakennukset ruskealla.

Toisaalta jotkut kohteet ovat tekoälyaineistoissa tunnistettu yhtä hyvin tai jopa tarkemmin kuin miten rakennus oli tallennettu maastotietokantaan (kuvio 20).



Kuvio 20. MTK:n rakennus sinisellä, AI-orto-rakennus violetilla ja AI-lidar-rakennus ruskealla.

4.4 Tulosten yhteenveto

Tähän lukuun on kerätty päällekkäisyysanalyysien tuloksista tehtyjä huomioita. MTK:n rakennuksista alle puolet (42 %) on rakennuksia, joille löytyi vastinrakennus AI-lidar-aineistosta vähintään 80 prosenttisesti (kuvio 3). Yli neljäsosa (29 %) MTK:n rakennuksista jäi tunnistamatta AI-lidar-aineistossa (kuvio 3). Näistä kohteista yli puolet (54 %) osoittautuivat kuitenkin todellisiksi, ortokuvalla näkyviksi rakennuksiksi (kuvio 15). AI-lidar-rakennuksista 6 % on rakennuksia, joita ei ollut maastotietokannassa (kuvio 5). Tästä osasta kuitenkin vain 15 % oli todellisuudessa rakennuksia (kuvio 16).

Kun MTK:n ja AI-orto-aineiston rakennuksia verrattiin keskenään, seuraavia huomioita nousi esiin. AI-orto-rakennukset vastasivat vähintään 80 prosenttisesti MTK:n rakennuksia 47 %:n osalta (kuvio 7). Osuus on lähes sama kuin AI-lidar-

rakennusten kohdalla 42 % (kuvio 3). Noin neljäsosa (26 %) MTK:n rakennuksista jäi tunnistamatta AI-orto-aineistossa (kuvio 7). Tämäkin on lähellä AI-lidar-aineiston vastaavaa lukua 29 % (kuvio 3). MTK:n tunnistamattomista rakennuksista kuitenkin 40 % oli rakennuksia ja lisäksi lähes puolet (48 %) oli sellaisia kohteita, joita ei pystynyt ortokuvulta havaitsemaan puuston peitteisyyden vuoksi (kuvio 17). AI-orto-aineistossa oli noin viidennes (21 %) kohteita, jotka puuttuivat MTK:n rakennuksista (kuvio 12). Tästä vain 7 % oli rakennuksia, jotka näkyvät ortokuvalla, mutta puuttuivat maastotietokannasta (kuvio 18). Kohteista 63 % oli tulkittu rakennuksiksi, vaikka rakennusta ei pystynyt silmämääräisesti ortokuvulta havaitsemaan (kuvio 18).

Tekoälyaineistojen geometrioita verrattiin silmämääräisesti MTK:n rakennusten geometrioihin ja ortokuvaan. Geometrioiden laatu vaihteli paljon tekoälyaineistoissa. Joissakin kohteissa oli siisti ja tarkka geometria ja niillä olisi suoraan voinut korvata maastotietokannassa olevat rakennukset. Toisaalta molemmissa tekoälyaineistoissa oli hyvin paljon sellaisia geometrioita, jotka eivät muistuttaneet rakennusta lainkaan. Joidenkin rakennusten geometriat olisivat muuttuneet paremmin hyödynnettäviksi, jos niissä olisi olleet suorat kulmat. Rakennusten sijaintitarkkuutta ei käytetyllä menetelmällä pystynyt seuraamaan. Maastotietokannan rakennus saattoi olla tekoälyaineistossa tulkittu kolmeksi eri rakennukseksi. Tekoälyaineistojen geometriat tulisikin ensin saada paremmiksi, minkä jälkeen rakennusten sijaintitarkkuutta olisi järkevä lähteä tutkimaan.

5 JOHTOPÄÄTÖKSET

Tämän tutkimuksen tulokset osoittavat, että tekoälyn avulla tuotettuja rakennusaineistoja voidaan hyödyntää maastotietokannan rakennusten ylläpidossa. Tekoälyaineistoista voidaan päällekkäisyysanalyysiä käyttämällä löytää maastotietokannasta puuttuvia, mutta maastossa olevia rakennuksia sekä rakennuksia, jotka maastotietokannasta pitäisi poistaa, koska niitä ei enää maastossa ole.

Päällekkäisyysanalyysillä, jossa syötetasona on AI-aineisto ja peitetasona MTK:n rakennukset löydetään tekoälyaineistosta luokan 0–10 % kohteet eli kohteet, joille ei löydy vastinrakennusta MTK:n rakennuksista. Nämä kohteet voitaisiin käydä läpi yksitellen ja tarkistaa, onko kyseessä uusi rakennus, joka pitäisi lisätä maastotietokantaan. AI-lidar-aineistoissa luokan 0–10 % rakennusten osuus oli 6 % (185 kpl / 2886 kpl) ja AI-orto-aineistossa 11 % (334 kpl / 3043 kpl), joten menetelmä voisi olla mahdollinen.

Kun päällekkäisyysanalyysi tehdään toisin päin eli syötetasona on MTK:n rakennukset ja peitetasona on AI-aineiston rakennukset, löydetään luokan 0–10 % kohteet eli MTK:n rakennukset, joille ei löydy vastinrakennusta AI-aineistosta. Näiden kohteiden osuudet ovat merkittävästi suuremmat kuin päinvastaisessa vertailussa. AI-lidar-aineiston kohdalla 29 % (1094 kpl / 3738 kpl) ja AI-orto-aineiston kohdalla 26 % (979 kpl / 3738 kpl). Nämä kohteet yksitellen läpikäymällä voitaisiin tarkastella, onko rakennus tosiaan poistunut maastosta ja voidaanko se poistaa maastotietokannasta. Koska rakennusten osuudet ovat melko suuret, pitäisi pohtia, saadaanko työhön käytetystä ajasta tarpeeksi suuri hyöty.

Työn tuottavuus paranee, jos tulevaisuudessa on saatavilla yksi, yhdistetty tekoälyaineisto, jota voidaan hyödyntää maastotietokannan rakennusten ylläpidossa. Työtä helpottaisi myös, jos tekoälyaineiston rakennuskohteiden geometriat vastaisivat laadultaan maastotietokannan laatuvaatimuksia, jotta rakennukset voitaisiin suoraan viedä maastotietokantaan eikä rakennuksia tarvitsisi erikseen digitoida.

6 POHDINTA

Tämän opinnäytetyön tavoitteena oli selvittää, voiko tekoälyn avulla tuotettuja rakennusaineistoja hyödyntää maastotietokannan rakennusten ylläpidossa. Tekoälyä ja koneoppimismenetelmiä hyödynnetään nykypäivänä paljon monilla eri aloilla ja tulevaisuudessa niiden käyttö lisääntyy. Tutkimuksen seurauksena kehitettiin menetelmä, jolla voidaan löytää tekoälyaineistosta ne rakennuskohteet, jotka mahdollisesti voidaan maastotietokannasta poistaa tai lisätä sinne. Opinnäytetyön tietoperusta pohjautuu vahvasti Maanmittauslaitoksen ATMU2021/es-projektin tuloksiin. Menetelmät, joilla tässä opinnäytetyössä käytetyt tekoälyaineistot on tuotettu, on kehitetty ATMU2021/es- ja AI4TDB-projekteissa.

ATMU2021/es-projektin loppuraportissa kerrotaan, että rakennusten automaattisen tunnistamisen osalta projektissa tuotettu menetelmä tunnistaa rakennukset jopa 97,9 % tarkkuudella. Tämä on Maanmittauslaitoksen asiantuntijoiden tekemä arvio. (Savola & Zhu 2023, 6.) Tässä opinnäytetyössä saadut tulokset eivät ole yhtä hyviä, kun MTK:n rakennuksista oli AI-lidar-aineistossa vähintään 80 prosenttisesti tunnistettu vain 42 % ja AI-orto-aineistossakin alle puolet (47 %). Suora vertailu ATMU2021/es-projektin tuloksiin on kuitenkin hankalaa, koska projektissa käytettyjä menetelmiä ja tuloksia ei ole avattu yksiselitteisesti. Yritin selvittää projektiin osallistuneilta henkilöiltä, millä menetelmillä tulokset on saatu. En kuitenkaan saanut sellaista materiaalia käyttööni, jota olisin voinut hyödyntää opinnäytetyön tulosten vertailuun. Voin vain todeta, että mielestäni tämän opinnäytetyön tulokset eivät tue projektissa saatuja tuloksia rakennusten tunnistamisessa.

Tuotannon näkökulmasta tekoälyn avulla tuotettuja rakennusaineistoja tulisi saada vielä paremmiksi, jotta niistä olisi enemmän hyötyä maastotietokannan rakennusten ylläpidossa. Rakennukset tulisi tunnistaa luotettavammin ja rakennusten geometrioiden pitäisi vastata paremmin todellisuutta. Näitä tekoälyaineistoja voidaan hyödyntää, mutta mielestäni niitä kannattaa vielä kehittää, jotta hyödystä saadaan suurempi.

Opinnäytetyössä saatiin vastattua siinä asetettuihin tutkimuskysymyksiin. Tekoälyn avulla tuotetut rakennusaineistot ovat tarpeeksi laadukkaita, jotta niitä voidaan hyödyntää maastotietokannan rakennusten ylläpidossa. Tekoälyaineistojen avulla maastotietokannasta löytyi sellaisia rakennuksia, jotka tulisi poistaa ja tekoälyaineistoista löytyi rakennuskohteita, jotka tulisi lisätä maastotietokantaan. Lisäksi tekoälyaineistoista on mahdollista poimia kohteita, jotka olisi hyvä käsitellä maastotietokannan rakennusten ajantasaistuksen yhteydessä. Tämä nopeuttaa työskentelyä ja vaikuttaa maastotietokannan rakennusten laatuun positiivisesti.

Opinnäytetyössä kehitettiin myös menetelmä, jonka avulla kartoittaja voi hyödyntää tekoälyaineistoja maastotietokannan rakennusten ylläpidossa. Valittu tutkimusmenetelmä ja aineistot olivat toimivia. Lisäarvoa opinnäytetyöhön olisi tuonut, jos Pudasjärven rakennusaineisto olisi ollut kattavampi ja aineistosta olisi myös tehty päällekkäisyysanalyysi. Lisäksi RHR-pisteiden aineistoa ei pystynyt täysin hyödyntämään maastotietokannasta poistettavien ja sinne lisättävien kohteiden tukena.

Tulevaisuutta on vaikea ennustaa, mutta tekoälyn käyttö lisääntyy ja tekoälyaineistojen laatu tulee hyvin todennäköisesti kehityksen myötä paranemaan, koska tekoälyn opetusaineiston määrä lisääntyy ja laatu paranee sekä menetelmiä kehitetään koko ajan. Sen seurauksena tekoälyaineistojen käyttö, niin rakennusten osalta kuin myös muiden maastotietokannan kohteiden osalta lisääntyy. Tekoälyaineistojen hyödyntäminen nopeuttaa maastotietokannan ajantasaistusta ja siihen tullaan tarvitsemaan vähemmän resursseja. Tekoälyaineistojen avulla maastotietokannan ylläpitoa turvataan, koska lähivuosina PATI-kartoittajia eläköityy. Uutta työvoimaa voi olla hankala saada rekrytoitua, koska väestö ikääntyy ja kilpailu työntekijöistä on kovaa. Tekoälyaineistojen hyödyntäminen maastotietokannan kohteiden ylläpidossa tarjoaa varmasti jatkossakin mielenkiintoisia opinnäytetöiden aiheita.

LÄHTEET

Adobe 2023. DWG-tiedostot. Viitattu 19.11.2023
<https://www.adobe.com/fi/creativecloud/file-types/image/vector/dwg-file.html>.

Avoindata.fi 2023. Turun ilmakekuva 2018. Viitattu 19.11.2023
https://www.avoindata.fi/data/en_GB/dataset/turun-ilmakekuva-2018.

Digi- ja väestötietovirasto 2023. Kiinteistö-, rakennus- ja paikkatiedot. Viitattu 21.11.2023 <https://dvv.fi/kiinteisto-rakennus-ja-paikkatiedot>.

Maanmittauslaitos 2023a. Korkeusmalli 2 m. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/asiantuntevalle-kayttajalle/tuotekuvaukset/korkeusmalli-2-m>.

– 2023b. Laserkeilaus ja ilmakekuvaus. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/laserkeilaus-ja-ilmakekuvaus>.

– 2023c. Maanmittauslaitoksen ilmakekuva. Viitattu 14.11.2023
<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/ilmakekuva>.

– 2023d. Maanmittauslaitoksen Maastotietokohteet-käsikirja. Ei julkinen.

– 2023e. Maanmittauslaitoksen ortokuva. Viitattu 21.11.2023
<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/ortokuva>.

– 2023f. Maastotietokanta. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/asiantuntevalle-kayttajalle/tuotekuvaukset/maastotietokanta-0>.

– 2023g. Peruspaikkatietopalveluiden ylläpito. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/peruspaikkatietojen-yllapito>.

– 2023h. Power Point-aineisto AI4TDB-projektista. Ei julkinen.

– 2023i. Tietoa Maanmittauslaitoksesta. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/organisaatio>.

– 2023j. Tietohallinto. Viitattu 13.11.2023
<https://mmlfi.sharepoint.com/sites/MITPA>. Ei julkinen.

– 2023k. Tuotannon tilannekartta. Viitattu 14.11.2023
<https://tilannekartta.maanmittauslaitos.fi/ilmakekuva>.

– 2023l. Tuotanto. Viitattu 13.11.2023
<https://mmlfi.sharepoint.com/sites/Tuotanto>. Ei julkinen.

– 2023m. Tutkimus. Viitattu 13.11.2023
<https://www.maanmittauslaitos.fi/tutkimus>.

– 2023n. Yleishallinto. Viitattu 13.11.2023
<https://mmlfi.sharepoint.com/sites/Yleishallinto>. Ei julkinen.

Savola, S. & Zhu, L. 2023. Advanced Technology for National Map Updating (ATMU2021/es). Loppuraportti. Ei julkinen.

Wikipedia 2023a. CSV. Viitattu 17.11.2023 <https://fi.wikipedia.org/wiki/CSV>.

– 2023b. Shapefile. Viitattu 19.11.2023 <https://fi.wikipedia.org/wiki/Shapefile>.

LIITTEET

- Liite 1. Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-
ohje

Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla

Veera Hautamäki

13.11.2023

Tässä ohjeessa havainnollistetaan päällekkäisyysanalyysin tekemistä rakennusaineistoille ja rakennusten luokittelua QGIS-ohjelmalla.

Tässä esimerkissä rakennukset luokitellaan alla olevan taulukon mukaisesti.

Taulukko 1. Rakennusten luokittelu päällekkäisen pinta-alan mukaan.

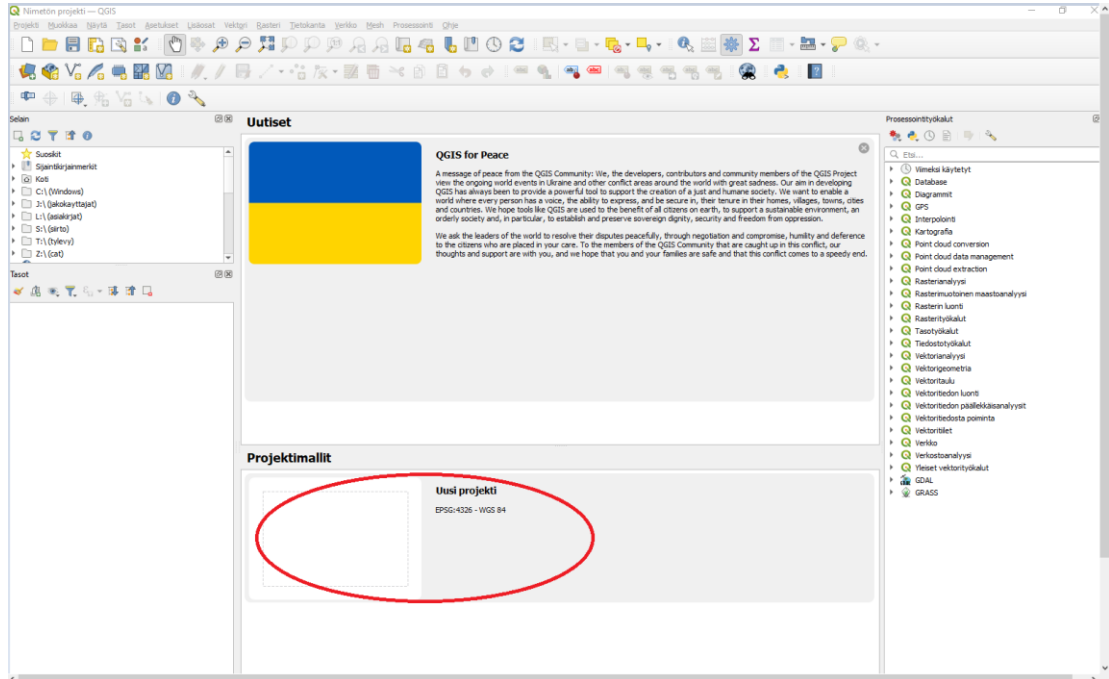
Luokka	Kuvaus
80–100 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa vähintään 80 prosenttisesti päällekkäin.
50–80 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 79,99 prosenttisesti ja vähintään 50 prosenttisesti päällekkäin.
10–50 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 49,99 prosenttisesti ja vähintään 10 prosenttisesti päällekkäin.
0–10 %	Sisältää syötetason rakennukset, jotka ovat peitetason rakennusten kanssa enintään 9,99 prosenttisesti päällekkäin.

Jos teet luokittelun eri prosenttien mukaan, vaihda ohjeissa oleviin kaavoihin haluamasi luvut.

Liite 1 3(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

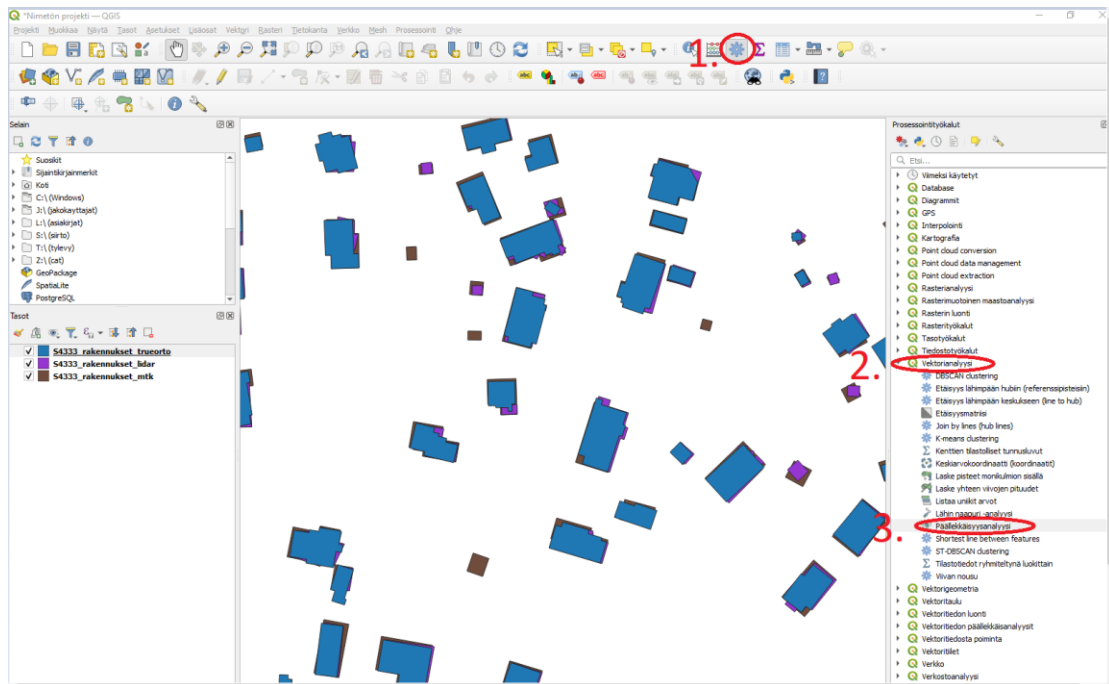
Päällekkäisyysanalyysi ja rakennusten luokittelu

Avaa QGIS-ohjelma ja valitse Uusi projekti



Raahaa tarvittavat rakennusaineistot QGIS:iin.

Avaa prosessointityökalut (1.) ja valitse kohdasta Vektorianalyysi (2.) ja Päällekkäisyysanalyysi (3.).



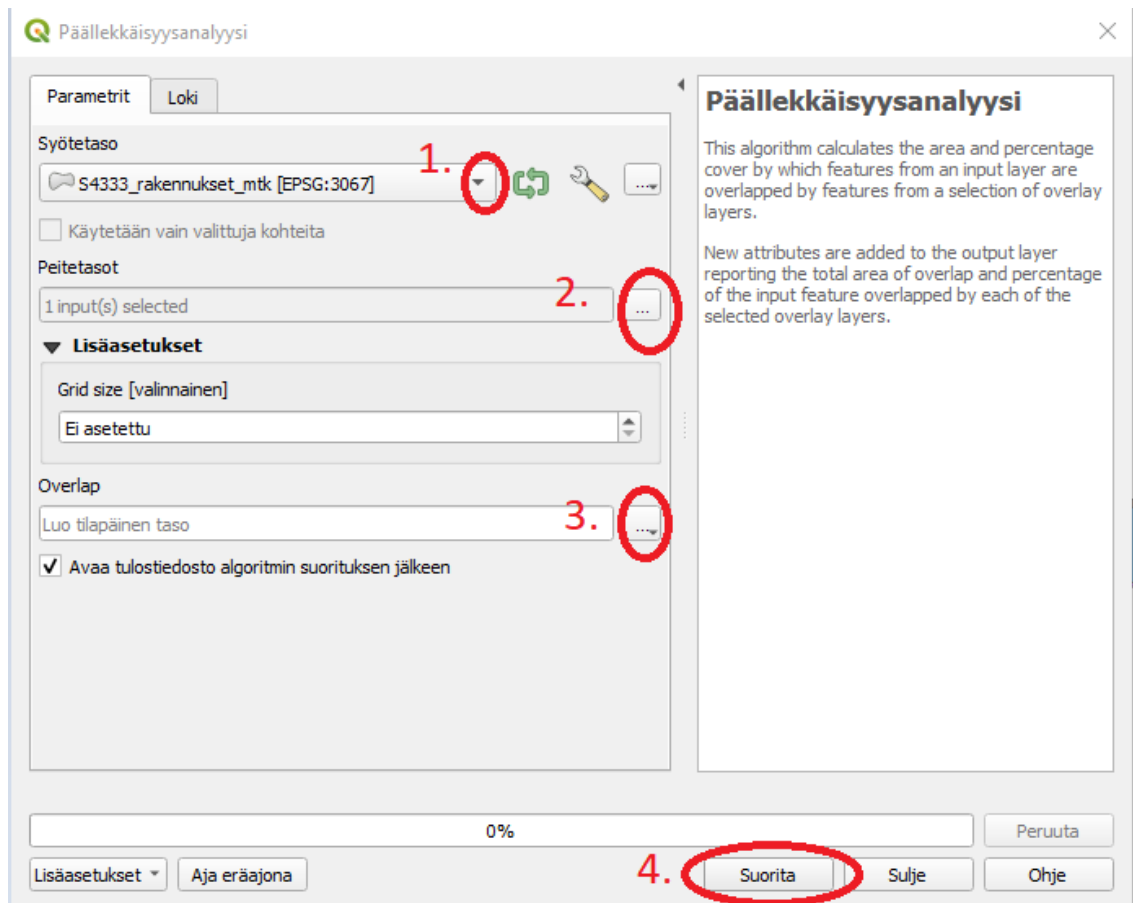
Liite 1 4(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Valitse haluamasi syötetaso nuolen takaa (1.) ja peitetaso kolmen pisteen takaa (2.). Tässä esimerkissä syötetasona on maastotietokannan rakennukset ja peitetasona lidar-aineiston rakennukset.

Voit jättää Overlap-kohdan tyhjäksi, jolloin ohjelma luo uuden Overlap-tason. Muista kuitenkin nimetä ja tallentaa taso haluamaasi paikkaan analyysin ajamisen jälkeen.

Voit myös tallentaa luotavan Overlap-tason kolmen pisteen (3.) takaa valitsemalla esim. Tallenna tiedostoon... Valitse tiedostolle haluamasi sijainti, nimi ja formaatti.

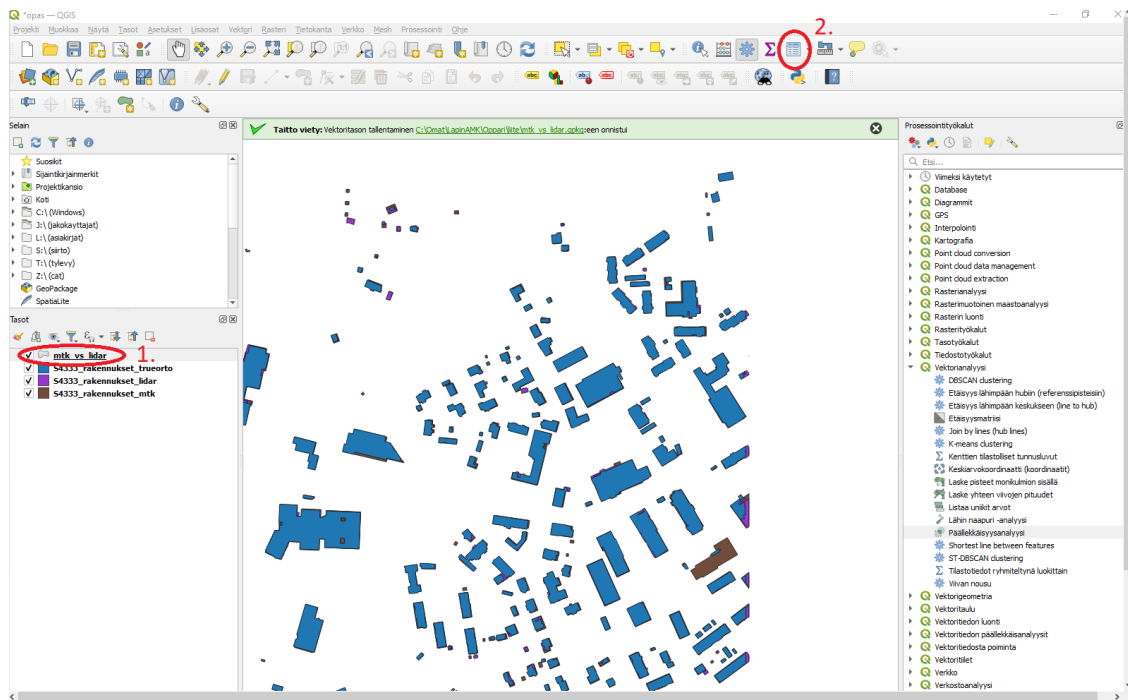
Lopuksi paina Suorita (4.), minkä jälkeen voit sulkea ikkunan.



Liite 1 5(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

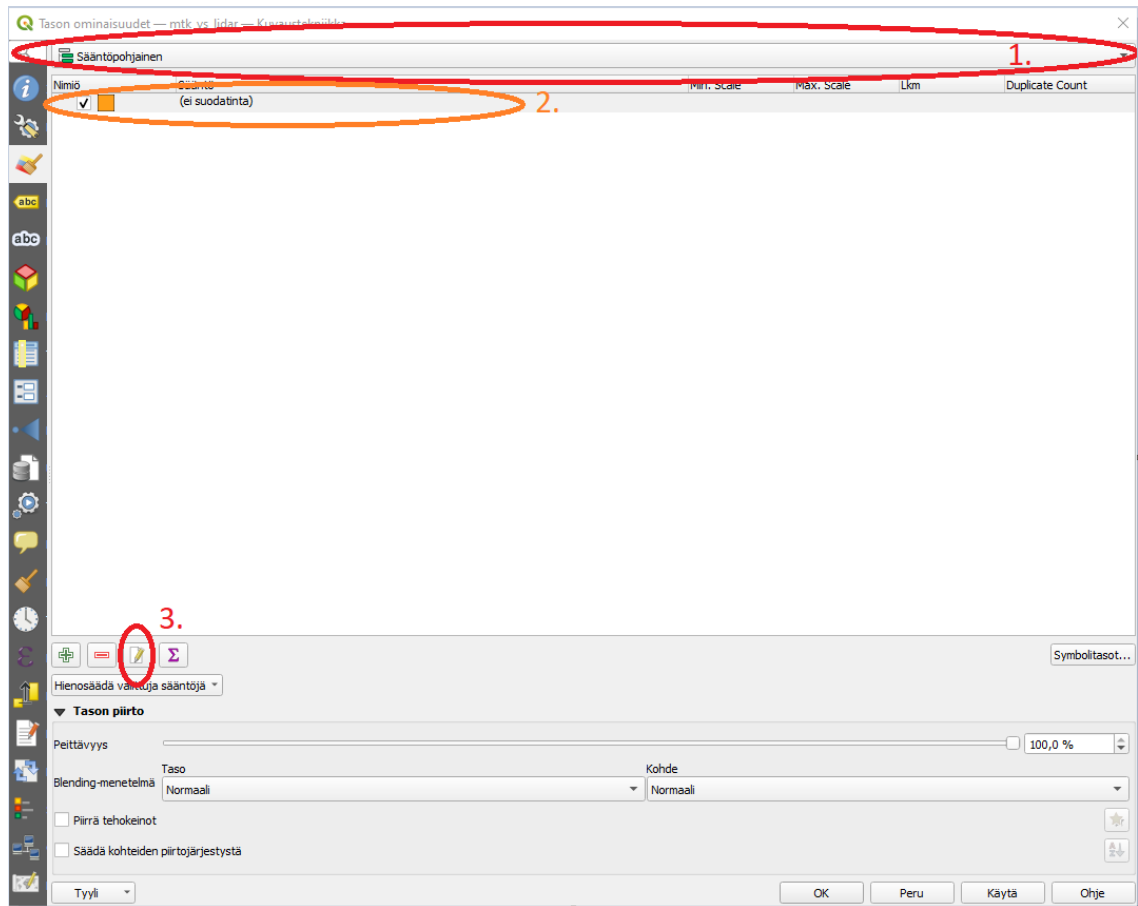
Avaa juuri muodostetun tason attribuuttitaulu valitsemalla taso aktiiviseksi (1.) ja klikkaamalla Avaa attribuuttitaulu-kuvaketta (2.).

Kaksoisklikkaa päällekkäisyysanalyysissa syntynyttä tasoa (1.) Tasot-ikkunassa, jolloin avautuu Tason ominaisuudet-ikkuna.



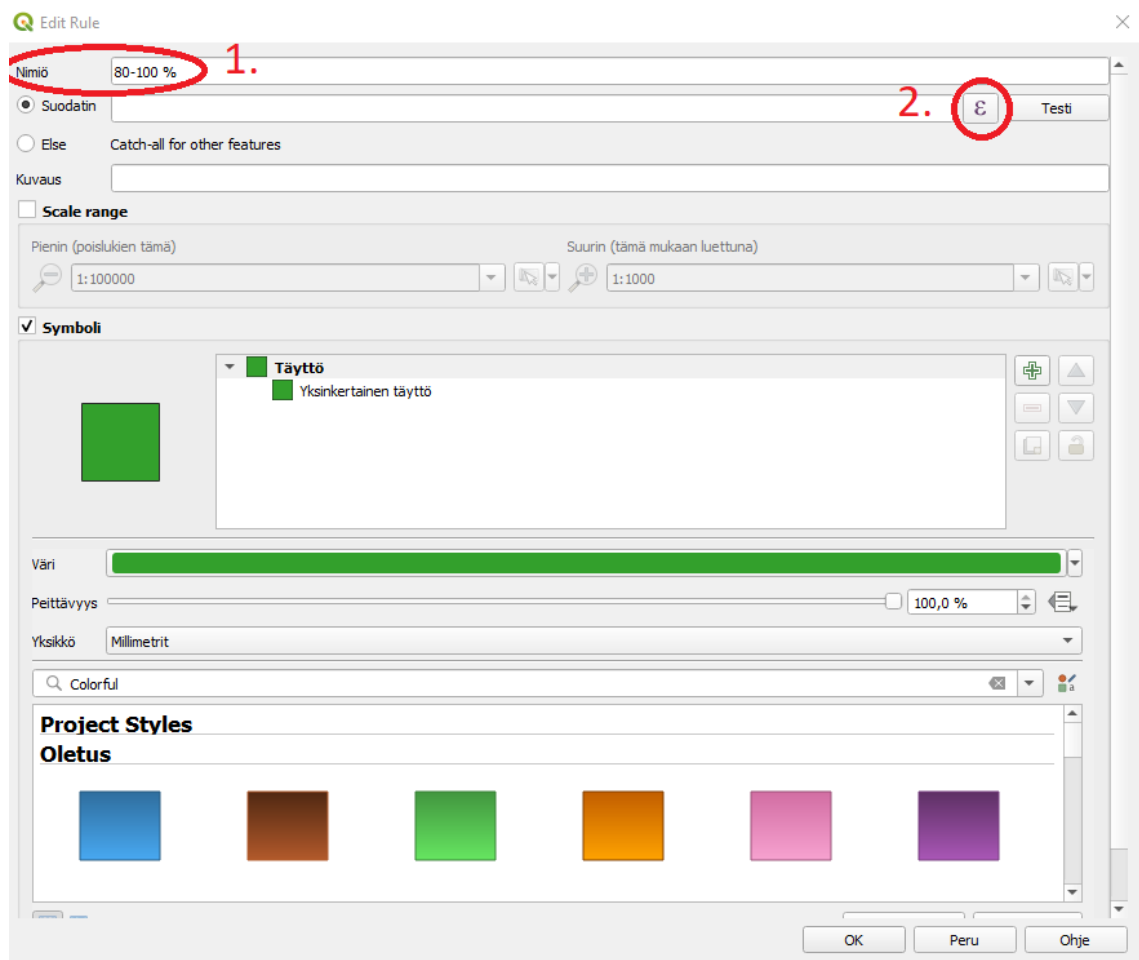
Liite 1 6(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Valitse ylimmäisestä valikosta nuolen takaa Sääntöpohjainen (1.) ja klikkaa alla oleva rivi aktiiviseksi (2.). Sen jälkeen klikkaa Muokkaa nykyistä sääntöä-kuva-
ketta (3.).

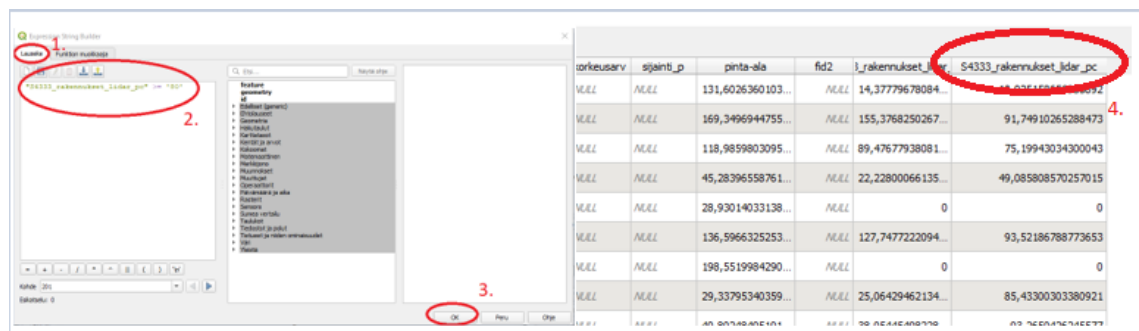


Liite 1 7(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Kirjoita säännölle nimi Nimiö-kohtaan (1.) ja klikkaa  kuvaketta (2.).

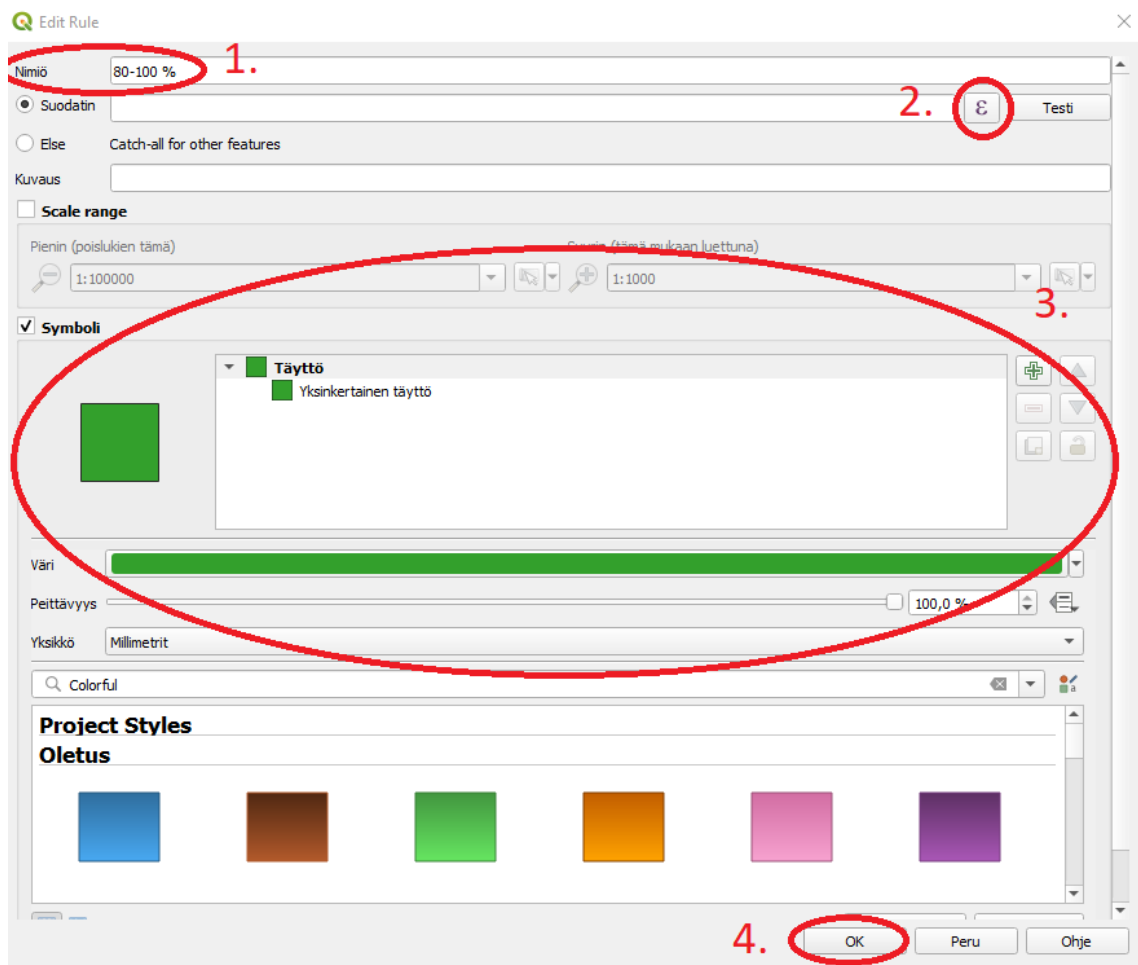


Kirjoita Lauseke-ikkunaan (1.) seuraava lauseke: "**S4333_rakennukset_lidar_pc**" >= '80'. Muuta vihreällä oleva tason nimi samaksi, mikä on attribuuttitaulussa viimeisen sarakkeen nimi (4.). Sarakkeen nimen alkuosa **S4333_rakennukset_lidar** on päällekkäisyysanalyysin peitetason nimi. Paina lopuksi OK (3.).



Liite 1 8(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Voit muokata symbolia haluamallasi tavalla (3.). Paina lopuksi OK (4.)

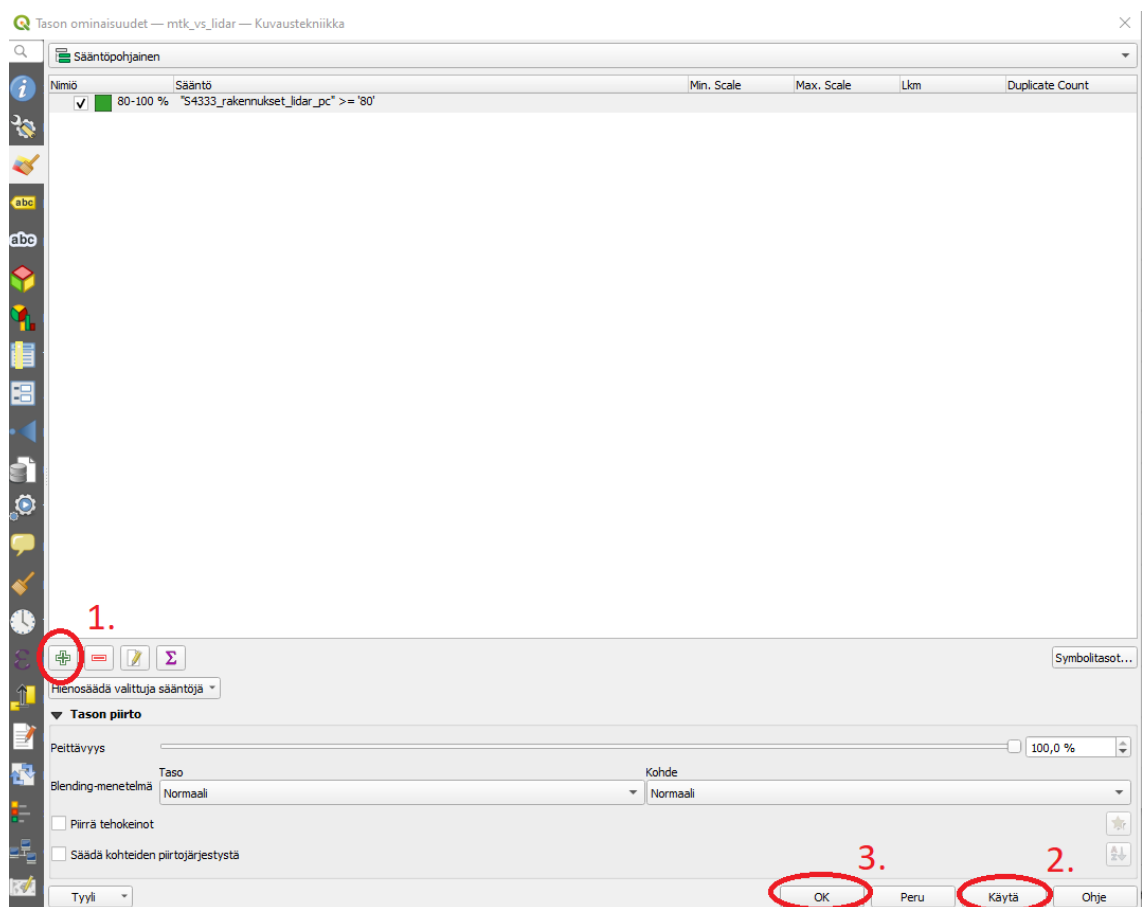


Liite 1 9(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Lisää loput säännöt plus-kuvakkeesta (1.) seuraavien mallien mukaan:

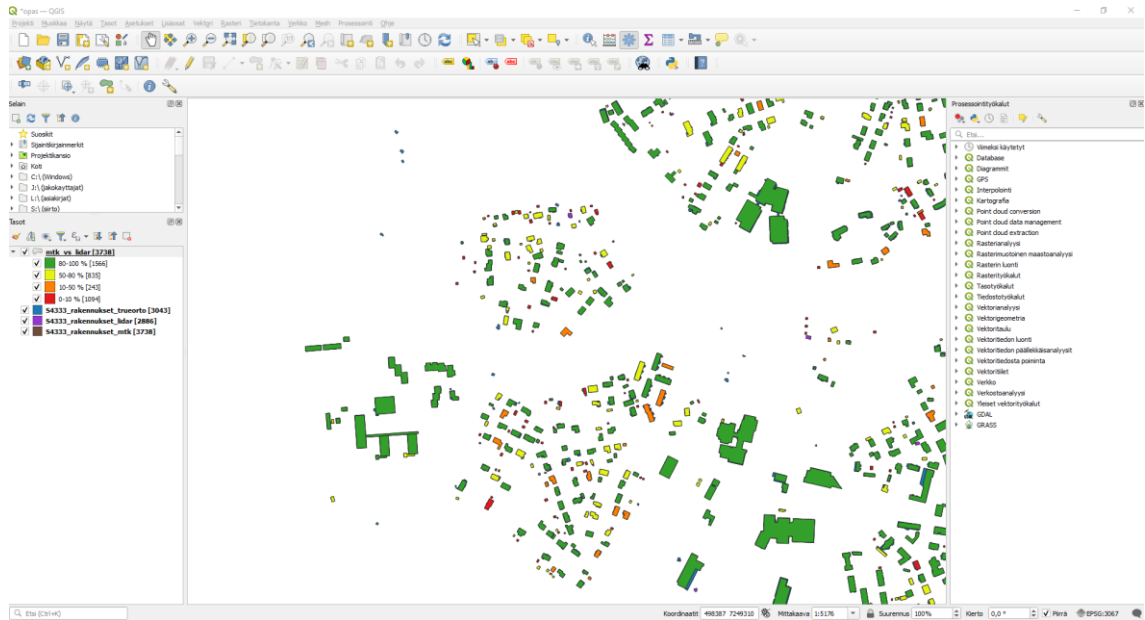
- "S4333_rakennukset_lidar_pc" < '80' AND "S4333_rakennukset_lidar_pc" >= '50'
- "S4333_rakennukset_lidar_pc" < '50' AND "S4333_rakennukset_lidar_pc" >= '10'
- "S4333_rakennukset_lidar_pc" < '10'

Paina lopuksi Käytä (2.) ja OK (3.).



Liite 1 10(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Nyt on päällekkäisyysanalyysissä luodun tason rakennukset luokiteltu neljään eri luokkaan sen mukaan, kuinka paljon lidar-aineiston rakennus peittää maastotietokannan rakennusta. Rakennusten lukumäärän saat näkyviin, kun klikkaat hiiren kakkospainikkeella tason päällä ja valitset Näytä kohteiden lukumäärä. Muista tallentaa projekti kohdasta Projekti -> Tallenna nimellä.



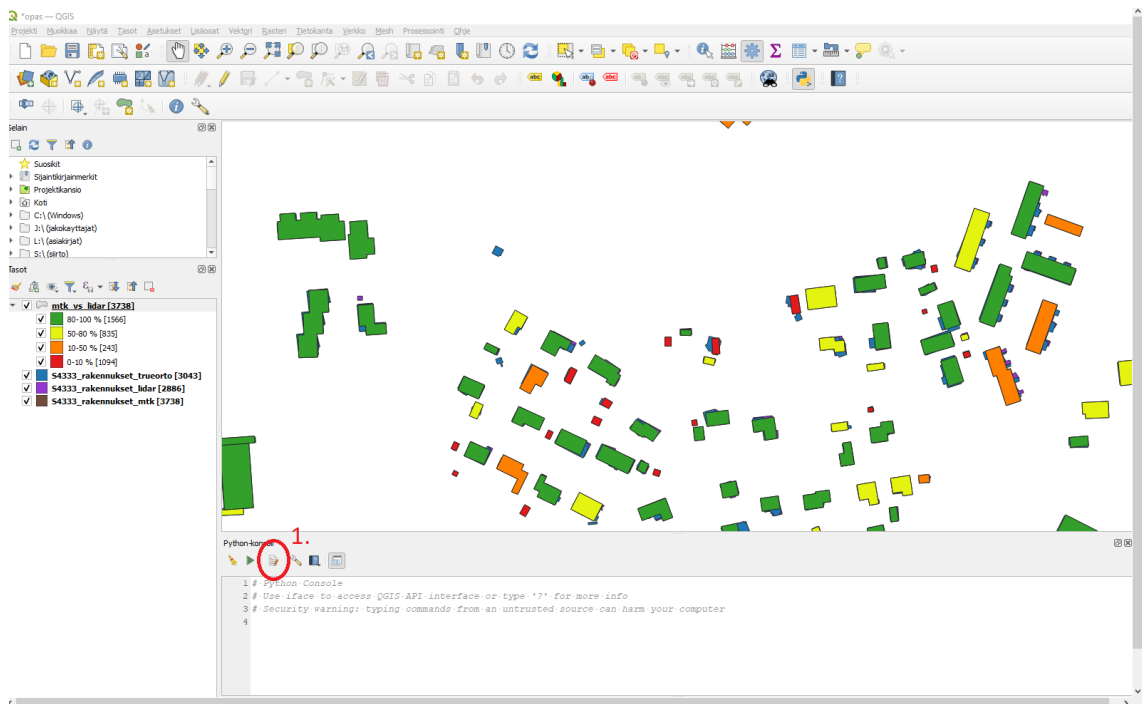
Liite 1 11(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Aineiston jokaisen rakennuksen läpi käyminen yksitellen

Kun haluat käydä aineiston jokaisen rakennuksen läpi yksitellen, voit käyttää apuna skriptiä (tehnyt Jere Raninen). Voit kopioida skriptin täältä xxx omaan tiedostoosi.

Valitse QGIS-ohjelmassa Lisäosat -> Python-konsoli

Avaa editori Näytä editori-kuvakkeesta (1.)

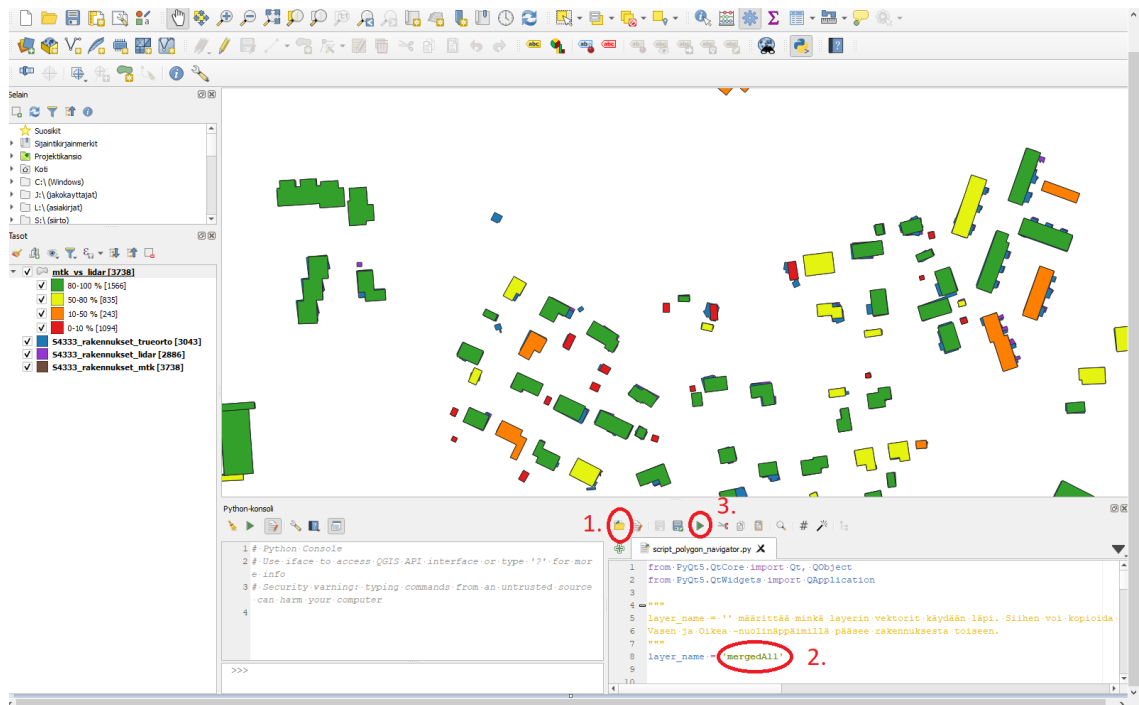


Liite 1 12(13). Pääallekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Avaa skripti omasta tiedostostasi (1.)

Muuta layer_name-kohtaan 'hipsujen sisään' sen tason (syötetaso) nimi, jonka rakennukset haluat käydä läpi (2.)

Paina lopuksi Suorita skripti (3.)



Nyt voit liikkua nuolinäppäimillä (oikea ja vasen) rakennuksesta toiseen.

Liite 1 13(13). Päällekkäisyysanalyysi ja rakennusten luokittelu QGIS-ohjelmalla-ohje

Näet Python-konsolin vasemmasta ikkunasta, minkä rakennuksen kohdalla olet menossa.

