



Niko Anttiroiko, Floris Jan Groesz, Janne Ikkäheimo,
Aleksi Kelloniemi, Stian Rostad & Oula Seitsonen

Pesiöjärvi opettaa – tervahautoja tunnistavaa tekoälyä kehittämässä

Maanmittauslaitoksen uusi, vuodesta 2020 alkaen julkaisema tarkempi 5p-laserkeilausaineisto on avannut uusia mahdollisuuksia arkeologisten kohteiden paikantamiseen tekoälymenetelmiä hyväksikäyttäen. Artikkelissa esitellään Museoviraston ja Oulun yliopiston LIDARK-hankkeen kehittelemiä uusia lähestymistapoja ilmalaserkeilausaineistojen algoritmipohjaiseen käsittelyyn. Hankkeen tavoitteena on kehittää ja pilotoida arkeologisten kohteiden puoliautomaattista tunnistusta laserkeilausaineistosta. Tässä artikkelissa avataan yleisesti tekoölyyn ja koneoppimiseen perustuvien tunnistusalgoritmien toimintaa ja soveltuvuutta arkeologiassa, sekä käsitellään esimerkkitapauksena tervahautojen ja hiilimiilujen tunnistamista lähes 8000 km² kattavalla alueella Pohjois-Pohjanmaalla ja Kainuussa.

Pesiöjärvi lär ut – utvecklandet av en artificiell intelligens som känner igen tjärdalar

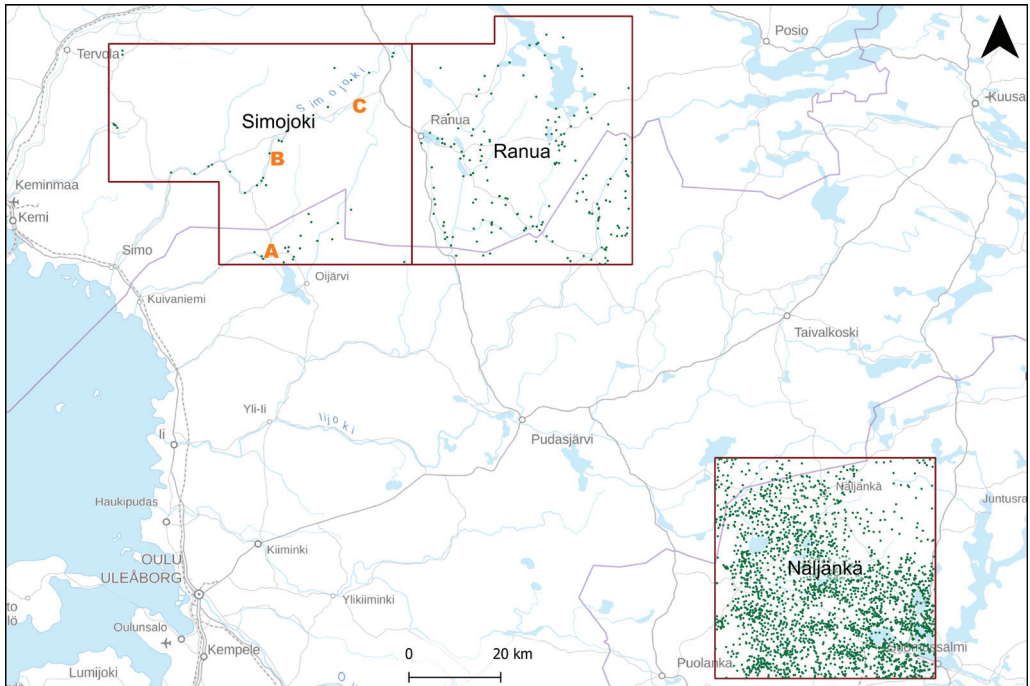
Lantmäteriverkets nya noggrannare laserskanningsdata har öppnat nya möjligheter för arkeologer att till exempel lokalisera objekten med artificiell intelligens. Artikeln presenterar nya metoder för algoritmbaserad bearbetning av data från flygburen laserskanning utvecklats av Museiverkets och Uleåborgs universitets LIDARK-projekt. Pilotprojektets syfte är att utveckla metoder för halvautomatisk identifiering av arkeologiska objekt från laserskanningsdata. Funktionen och tillämplighet av sådana igenkänningsalgoritmer inom arkeologi diskuteras här på allmänt sätt. Som exempel diskuteras igenkänning av tjärdalar och kolmilor på områden som sträcker sig nästan 8000 km² i Norra Osterbotten och Kajanaländ.

Johdanto

Muutamassa *Muinaistutkijassa* vuonna 2021 ilmestyneessä artikkelissa on esitelty Maanmittauslaitoksen syksystä 2020 alkaen julkaistua tarkempaa 5p-laserkeilausaineistoa ja käsitelty sen arkeologisia soveltamismahdollisuuksia (Ikäheimo 2021; Ikäheimo & Seitsonen 2021; Seitsonen 2021). Artikkeleissa on moneen otteeseen viitattu algoritmipohjaisiin tekoälymenetelmiin, joita on alkuvuodesta 2021 lähtien kehitetty Museoviraston koordinoiman *Arkeologisten kohteiden automaattinen tunnistaminen laserkeilausdatasta* -kehittämishanke LIDARK:in (ks. Hamari 2020) puitteissa. Tämä artikkeli tarjoaa ensimmäisen maistiaisen projektin tuloksista.

Tässä artikkelissa esitellään algoritmiin perustuvaa puoliautomoitua koh-

teentunnistusta sovellettuna tervahautojen ja hiilimiilujen tunnistamiseen kahdella Pohjois-Pohjanmaalle ja Kainuuseen sijoittuvalla tutkimusalueella (Kuva 1). Niistä ensimmäinen on 2304 km² laaja Näljängän laserkeilaustuotantoalue, joka sijoittuu pääosin Suomussalmen ja Puolangan, mutta vähäisemmässä määrin myös Pudasjärven, Taivalkosken ja Hyrynsalmen kuntien alueelle. Tätä aluetta on jo sivuttu aiemmin sen kaakkoisnurkan osalta Pesiöjärven tutkimusalueen tervahautoja käsittelevässä artikkelissa (Ikäheimo 2021). Näljängän alueella pilotoidun prosessin toimivuutta testattiin lisäksi Ranuan ja Simojoen laserkeilaustuotantoalueista muodostetulla tutkimusalueella, jonka yhteenlaskettu laajuus on yli 5500 km². Näistä ensiksi



Kuva 1. Tutkimusalueet, paikallistetut tervahaudat (vihreällä) ja tekstissä käsitellyt miilukentät A-C (ks. Kuva 8). Kuva: Janne Ikäheimo.

mainittu tuotantoalue kattaa Ranuan kunnan itäpuoliskon lisäksi Pudasjärven pohjois- ja Posion länsiosia, ja jälkimmäinen tuotantoalue Ranuan länsiosan lisäksi osia Keminmaan, Simon, Iin, Pudasjärven ja Tervolan kunnista.

Ennen varsinaisten tutkimustulosten ja käytettyjen menetelmien esittelyä, artikkelin tavoitteena on avata yleisellä tasolla tekoälyyn ja koneoppimiseen perustuvien tunnistusalgoritmien peruseriaatteita sekä käsitellä hieman tarkemmin niiden soveltamista ja käyttökelpoisuutta arkeologiassa. Tätä seuraa kuvaus kehittämämme tervahautojen tunnistamiseen soveltuvan algoritmin suunnittelu- ja testausprosessista. Sen jälkeen raportoidaan ja tulkitaan lyhyesti molemmilta tutkimusalueilta saadut tutkimustulokset. Mukaan on sisällytetty tulosten vertailu ilman tekoälyä suoritettuun pilottitutkimukseen Pesiöjärven eteläpuolella (Ikäheimo 2021), sillä tässä raportoitavat algoritmipohjaiset tulokset eroavat niistä varsin merkittävästi. Näiden erojen syiden purkaminen, vertailu ja analysointi avaavat samalla kouriintuntuvasti laserkeilausaineiston tekoälyavusteisen hyödyntämisen etuja arkeologiassa verrattuna aiemmin käytettyyn ”ihmislähtöiseen” lähestymistapaan. Edut eivät näet suinkaan rajoitu kohteiden paikantamis- ja tunnistamistyössä saavutettavaan ajan säästöön (Gallwey et al. 2019: 1–2; kts. alla).

Tekoäly muinaisjäännösten tunnistajana

Tekoälyyn pohjautuvien menetelmien soveltaminen eri tutkimusaloilla, arkeologia ja kulttuuriperinnöntutkimus mukaan lukien, on kasvanut valtavasti viime vuosina (esim. Caspari et al. 2014; Orengo et al. 2020; Yaworsky et al. 2020). Tähän ovat vaikuttaneet esimerkiksi käytettävissä olevien menetelmien ja soveltamiskohteiden kirjon laajeneminen, algoritmien nopeutuminen ja helpottu-

minen tietokoneiden laskentatehon kasvun myötä, sekä parantuneet lähdeaineistot, kuten esimerkiksi aiempaa tarkemmat ilmalaserkeilausaineistot (esim. Küçükdemirci et al. 2021; Trier et al. 2021). Tekoälyyn, koneoppimiseen ja syväoppimiseen perustuva kulttuuriperintökohteiden puoliautomoitoinen tunnistaminen laserkeilausaineistoista ja muista kaukokartoitusaineistoista on yksi arkeologisen ennustavan mallinnuksen osa-alueista (Wachtel et al. 2018; Orengo et al. 2020; Yaworsky et al. 2020). Ennustavien mallien tavoitteena on pyrkiä paikantamaan toistettavasti ja tilastollisesti merkittävästi ennestään tuntemattomia kohteita tai rakenteita jollain tietyllä alueella jonkin tietyn koulutusaineiston ja valittujen algoritmien avulla.

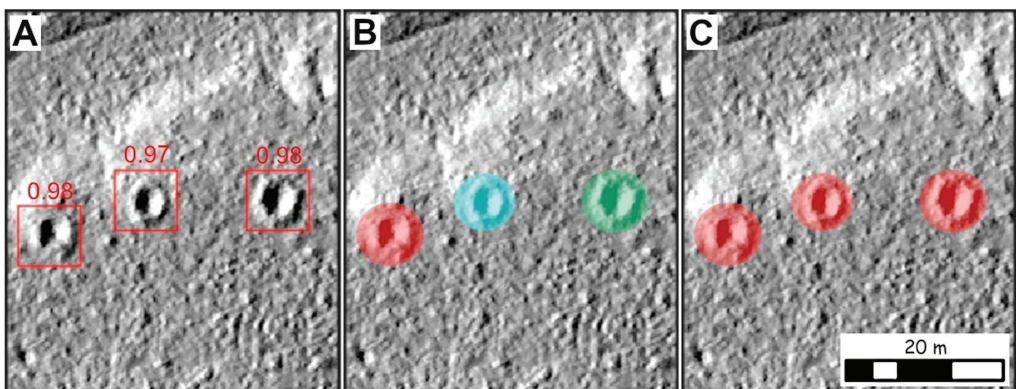
Viime aikoina varsinkin koneoppimiseen ja syväoppimiseen perustuvien lähestymistapojen hyödyntäminen arkeologiassa on lisääntynyt huomattavasti (esim. Yaworsky et al. 2020; Bonhage et al. 2021; Davis et al. 2021). Syväoppiminen on yksi koneoppimisen osa-alueista, joka pyrkii hahmottamaan datasta haluttuja abstraktioita hierarkkiseen arkkitehtuuriinsa pohjautuen (Guo et al. 2016; Verschoof-van der Vaart & Lambers 2019). Konvoluutioneuroverkko (Convolutional Neural Networks, CNN) CNN on monenlaisten aineistojen kuvantamisessa yleisesti käytetty kuvaluokittelualgoritmi, joka pyrkii toiminnassaan mukailemaan ihmisen näköaivokuoren käyttäytymistä (Verschoof-van der Vaart & Lambers 2019). Siihen perustuvat menetelmät ovat osoittautuneet erityisen tehokkaiksi lähtökohdiksi arkeologisten kohteiden paikantamisanalyysseissa (esim. Caspari & Crespo 2019; Lambers et al. 2019; Trier et al. 2021).

Arkeologit ovat hyödyntäneet jatkuvasti kehittyviä CNN-pohjaisia menetelmiä kohteiden tunnistuksessa vuodesta 2016 alkaen (Trier et al. 2016). Varsinkin aluemaisten kohteiden tunnistamiseen perustuva R-CNN (Region-Based Convolutional Neural Networks; Girshick 2015), sekä tähän pohjautuvat Faster R-CNN (Ren et al. 2017) ja Mask

R-CNN (He et al. 2017) -menetelmät ovat olleet suosittuja arkeologisissa yhteyksissä (ks. Kazimi et al. 2019; Verschoof-van der Vaart & Lambers 2019; Bonhage et al. 2021; Guyot et al. 2021a; 2021b; Trier et al. 2021). Näiden menetelmien pääasialliset erot voi tiivistää seuraavasti (Kuva 2): A) perinteinen CNN luokittelee koko analysoitavan kuva-alan ja määrittelee siihen liittyvän ilmiön luokan tai tyyppin; B) R-CNN vie analyysin pelkkää luokittelua pidemmälle ja kykenee määrittelemään luokiteltavalle ilmiölle myös sen tarkan paikan maisemassa ja rajaamaan sille “sijaintia rajaavan suorakaiteen” (Bounding Box, BBOX); ja C) Mask R-CNN on Faster R-CNN:iin perustuva menetelmä, joka pystyy ilmiön havaitsemisen, luokittelun ja sijainnin BBOX-määrittelyn lisäksi myös luonnehtimaan ilmiön morfologiaa rajaamalla sen kaksiulotteisen jalanjäljen ja erottelemalla (segmentoimalla) ilmiön taustastaan omaksi instanssiksi (Guyot et al. 2021a; 2021b).

Alun perin lääketieteelliseen kuvantamiseen kehitettyä U-net-konvoluutioneuroverkkoa (Ronneberger et al. 2015) on myös hyödynnetty viime aikoina arkeologiassa

(esim. Garcia-Molsosa et al. 2021; Ikäheimo & Seitsonen 2021). Mask R-CNN:n tavoin U-net kykenee luonnehtimaan ilmiöiden morfologiaa ja segmentoimaan ne taustastaan. Analyysituloksissa erona on, että U-Net tuottaa semanttisen segmentoinnin (*semantic segmentation*), joka luokittelee määriteltävät ilmiöt samaan luokkaan, muttei erottele niitä erillisiksi entiteeteiksi kyseisen luokan sisällä. Mask R-CNN tuottaa sen sijaan jokaiselle ilmiölle oman erillisen segmentoinnin (*instance segmentation*) (Kuva 2). U-Net-algoritmit ovat osoittautuneet arkeologiassa erityisen tehokkaiksi mallinnusmenetelmiksi pienten koulutusaineistojen kanssa toimittaessa (Küçükdemirci et al. 2021; Seitsonen & Ikäheimo 2021). LIDARK-hankkeessa kohteiden tunnistamiseen kehittelemämme lähestymistapa perustuikin juuri U-Net-konvoluutioneuroverkkoon. Tämä tarkoittaa esimerkiksi sitä, että kulttuuriperintökohteessa lähekkäin sijaitsevat jäännökset voivat segmentoitua yhdeksi yhteiseksi polygoniksi. Siksi tunnistusalgoritmin tuottamia tuloksia on tarpeen tarkastella ja jälkikäsitellä myös manuaalisesti.



Kuva 2. Arkeologisten kohteiden tunnistaminen CNN-pohjaisilla menetelmillä (esimerkkinä pyyntikuoppakohteet): A. ilmiöiden sijaintia rajaavat suorakaiteet ja havainnon todennäköisyysarvo (BBOX, R-CNN). B. Erillisten ilmiöiden instanssisegmentointi (Instance segmentation, Mask R-CNN) C. Ilmiöluokan semanttinen segmentointi (Semantic segmentation, U-net). Sisältää Maanmittauslaitoksen Laserkeilausaineisto 5 p -aineistoa vuodelta 2020. Kuva: Oula Seitsonen.

Algoritmin kehittäminen

Algoritmin kehittämisen lähtöaineistona käytettiin Maanmittauslaitokselta hankittua Näljängän Laserkeilausaineisto 5p -tuotantoaluetta kokonaisuudessaan (käyttölisenssi MML 15920/05 00 00/2021). Kyseinen laserkeilausaineisto on ladattavissa Maanmittauslaitokselta laz-tiedostoformaattissa 1x1 km:n kokoisina ja automaattiluokiteltuina blokkeina (Ikäheimo & Seitsonen 2021). Näljängän alueen aineiston on keilannut Terratec AS kesäkuun 2020 puolivälissä ja sen keskimääräinen paikannusvirhe on horisontaalisesti noin 4 cm ja vertikaalisesti noin 5 cm. Neliömuotoisen tuotantoalueen, jonka mitat ovat 48 x 48 km, kaakkoisnurkka sijaitsee hieman Ämmänsaaren kirkosta etelälounaaseen ja luoteisnurkka kuroutuu Puhosjärven tienoille (Kuva 1). Aiemmassa artikkelissa (Ikäheimo 2021) käsitelty Pesiöjärven tutkimusalue sijoittuu tuotantoalueen kaakkoisnurkkaan.

Automaattista tunnistusta varten laserkeilausaineistosta tuotettiin ensin LAStools-ohjelmalla digitaalinen pintamalli (DEM) 25 cm ruutukoolla, ja sen pohjalta laskettiin topografinen sijainti-indeksi (Topographic Position Index, TPI) käyttäen Zoran Čučkovićin kehittämää Numpy-kirjastoa hyödyntävää toteutusta (Čučković 2019). TPI muodostetaan vähentämällä tarkasteltavan solun korkeudesta sitä ympäröivien solujen korkeuksien keskiarvo halutulla säteellä, jolloin ympäristöään korkeammalla sijaitsevat solut saavat positiivisia ja matalammat negatiivisia arvoja (esim. Tagil & Jenness 2008; Seitsonen 2013; Argyriou et al. 2017). TPI:n avulla voidaan korostaa halutun kokoisia maastonmuotoja muuttamalla tarkasteltavan alueen sädettä. Tässä tutkimuksessa käytettiin 30 solun eli 12,5 metrin sädettä, jonka todettiin soveltuvan hyvin tervahautojen ja hiilihiilujen erottamiseen aineistosta.

Algoritmin opetusaineistona käytettiin osaa tutkimusalueen tervahautoista, joiden ympärille digitoitiin niitä rajaavat polygonit. Näiden polygonien perusteella algo-

ritmi etsii piirteiltään vastaavia kohteita TPI-aineistosta. Algoritmin opettamista varten TPI-aineistosta irrotettiin 512 x 512 pikselin kokoiset kuvakaappaukset kohdepolygonien sijaintia ilmaisevine maskeineen. Opetusaineiston koon kasvattaminen ja vaihtelun määrän lisääminen parantavat mallin opettamisesta saatavia tuloksia ja ehkäisevät ylisovittamista, eli tilannetta, jossa mallin yleistettävyyttä alkaa heikentyä mallin ikään kuin 'oppiessa ulkoa' opetusaineistossa esiintyvät ilmiöt.

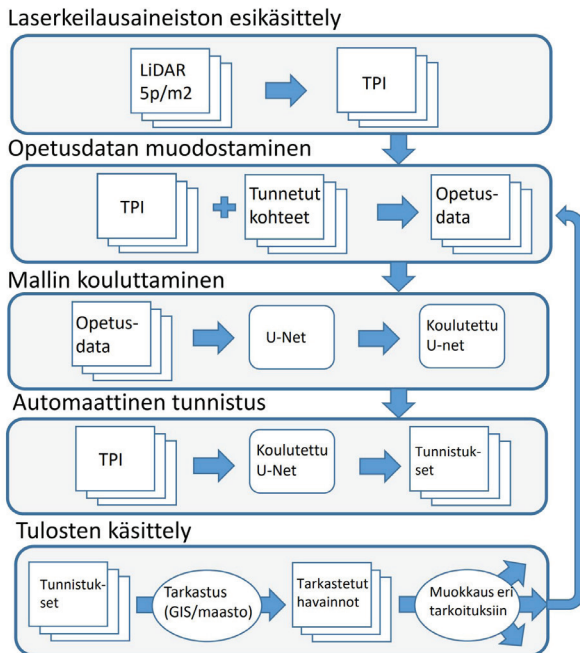
Opetusaineistoa augmentoitiin otamalla jokaisesta opetusdatan muodostamiseen käytetystä kohteesta kaksi sattumanvaraisesti keskitettyä kuvakaappausta, joista tehtiin lisäksi myös x- ja y-akselien suhteen peilattut sekä 90 ja 270 astetta käännetyt versiot. Yksittäiset kohteet saattavat kuitenkin esiintyä opetusaineistossa useampiakin kertoja, mikäli yhteen kuvakaappaukseen osuu useampia lähekkäin sijaitsevia kohteita, tai kohde sijaitsee kahden tai useamman karttalehden rajalla. Dataan lisättiin myös esimerkkejä muista tervahautaa muistuttavista maastonmuodoista, kuten maakellareista, jotta malli oppisi erottamaan ne oikeista tervahautoista. Mallin kouluttamista varten käytetty opetusaineisto jaettiin satunnaisotannalla opetus- (80%) ja validointiaineistoon (20%). Opetusaineistoa käytetään nimensä mukaisesti mallin kouluttamiseen, kun taas validointiaineiston avulla arvioidaan mallin suoriutumista aineistolla, jota ei ole käytetty mallin opettamiseen. Tämän takia sovelsimme Näljängän tutkimusalueella koulutettua mallia Ranua-Simojoki-alueelle saadaksemme vertailuaineistoa. Tämä antaa paremman kuvan kyseisen mallin yleistettävyydestä muihin aineistoihin.

Käyttämämme U-Net-mallinnusalgoritmit on kehitetty Python-ohjelmointiympäristössä hyödyntäen TensorFlow-koneoppimisalustaa (esim. Caspari & Crespo 2019; <https://www.tensorflow.org/>). U-net-mallin opettaminen tapahtuu niin sanotuissa epookkeissa. Epookilla tarkoitetaan yhtä kierrosta,

jonka aikana mallia koulutetaan käymällä opetusaineisto kertaalleen läpi, minkä jälkeen mallin suoriutumista arvioidaan validointiaineiston avulla. Koulutuksen aikana mallin suoriutuminen yleensä kohentuu tiettyyn pisteeseen asti, minkä jälkeen tulokset alkavat ylisovittamisesta johtuen jälleen heiketä. Mallin suoriutumisen mittariksi valittiin niin sanottu leikkausalue (IoU, intersect over union), joka kuvastaa sitä, miten tarkasti mallin tuottamat ennusteet vastaavat opetusdatan maskeja. IoU lasketaan jakamalla opetusdatan ja mallin tuottamien polygonimaskien leikkauksen ala niiden yhteispinta-alalla. Tällöin IOU saa arvon väliltä 0–1, joka on sitä suurempi mitä paremmin ennuste ja opetusdatan maski vastaavat toisiaan. Tässä tutkimuksessa koulutusta jatkettiin, kunnes mallin suoriutuminen ei kohentunut kymmenen edellisen eepookin aikana. Validointiaineiston perusteella parhaiten suoriutunut malli tallennettiin jatkokäyttöä varten.

Kohteita tunnistettaessa käytettiin 512 x 512 pikselin (128 metrin) liukuvaa ikkunaa, jonka avulla käytiin saumattomasti läpi koko tutkimusalue. Kohteita tunnistettaessa U-Net-malli tulostaa kuvan perusteella samankokoisen matriisin arvoja väliltä 0–1, jotka kuvaavat todennäköisyyttä, jolla kyseinen pikseli on osa etsittävää ja tunnistettavaa ilmiötä. Tunnistetuille kohteille muodostettiin rajaukset vektoroimalla polygoneiksi alueet, joilla pikselien saamat arvot ovat yli 0,5. Kohteiden ominaisuustiedoksi laskettiin rajauksen sisään jäävien pikselien keskiarvo, jota voidaan käyttää kuvaamaan tunnistuksen luotettavuutta (ks. alla). Lopuksi tunnistetut kohteet sidottiin alkuperäiseen koordinaatistoon ja yhdistettiin paikkatietojärjestelmissä geospaatialaisen tiedon tallentamiseen yleisesti käytetyksi shapefile-tiedostoksi (*.shp).

Tutkimuksen aikana mallia ja opetusaineistoa kehitettiin iteratiivisesti, eli toistamalla prosessia, jossa opetusaineistoa ja



Kuva 3. Algoritmin/mallin toimintaa havainnollistava kaavio. Kuva: Niko Anttiroiko.

mallia parannettiin edellisellä kierroksella saatujen tulosten perusteella (Kuva 3). Ensimmäisellä kierroksella opetusaineistona käytettiin noin 400 tervahaudasta ja sadasta muusta kohteesta muodostuvaa aineistoa, joka pohjautui pääosin aiempaan Pesiöjärven aluetta käsittelevään tutkimukseen (Ikäheimo 2021). Jokaisen kierroksen jälkeen mallin tunnistamat kohteet tarkastettiin visuaalisesti ja luokiteltiin havainnot todellisiin, epäselviin ja virheellisiin tunnistuksiin. Tarkistettujen tulosten perusteella opetusdataa täydennettiin lisäämällä aineistoon aiemmin tunnistetut tervahaudat sekä otos virheellisesti tunnistetuista kohteista. Tätä prosessia toistettiin kolmen iteraation verran, minkä jälkeen uusilla iteraatioilla ei katsottu enää olevan merkittävää vaikutusta tuloksiin. Malliin tehdyt muutokset iteraatioiden välillä pitävät sisällään lähinnä pieniä korjauksia koskien esimerkiksi karttalehtien rajalla sijaitsevien kohteiden tunnistamista sekä käytettävyyteen liittyviä parannuksia.

Tulokset ja havainnot

Seuraavassa esitellään käsitellyiltä tutkimusalueilta saatuja tuloksia, sekä analyysiin ja sen eri vaiheisiin liittyviä yleisiä havaintoja esimerkiksi ajankäyttöön liittyen. Painotamme tässä yhteydessä algoritmin käytettävyyteen ja toimivuuteen liittyviä seikkoja, emmekä niinkään eri muinaisjäänöstyyppihin liittyviä tulkinnallisia havaintoja, joita tullaan käsittelemään tulevissa artikkeleissamme. Ensimmäisessä käymme läpi Näljängän tuotantoalueen ja sen jälkeen Ranuan ja Simojoen tutkimusalueen.

Näljängän tutkimusalue

Näljängän alueen tervahautojen määrä on Museoviraston ylläpitämässä Muinaisjäänösrekisterissä 39 kpl, kun taas Maanmittauslaitoksen Maastotietokantaan samalta alueelta on kirjattu 757 tervahautakohdetta. Ikähei-

mon (2021) alustavan arvion mukaan, joka perustui Pesiöjärven tutkimusalueelta löydettyjen uusien kohteiden määrästä laskettuun kertoimeen 3,98x, Näljängän tuotantoalueelta löytyisi noin 2826 tervahautaa. Edellä esitetyn algoritmin ajamisen ja 5192 tunnistuspolygonia sisältävien tulosten käsittelyn jälkeen päädyimme tulokseen, jonka mukaan Näljängän tuotantoalueella voidaan arvioida olevan noin 2750 tervahautaa sekä nelisenkymmentä kohdetta, joiden statuksen määrittely tervahaudaksi vaatii vielä maastokäynnin.

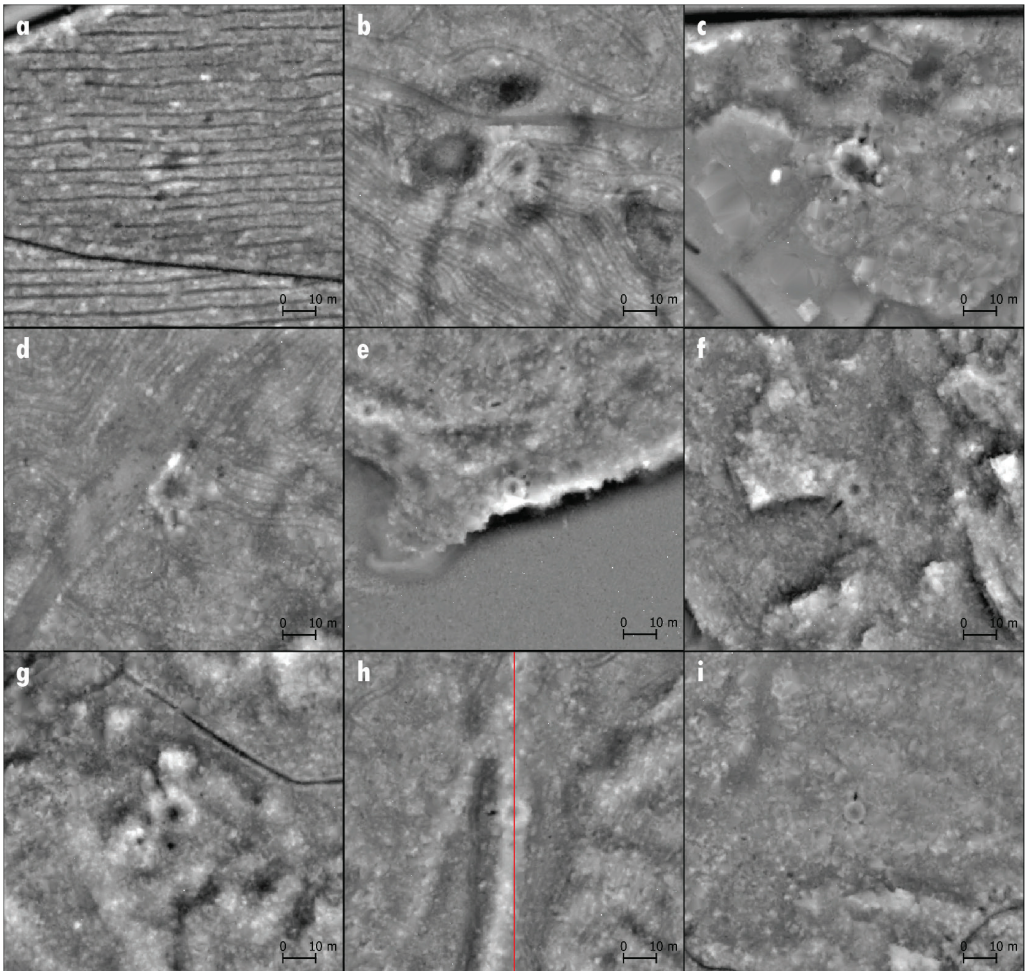
Lisäksi Näljängän tuotantoalueen kaakkoiskulmaan sijoittuvalta Pesiöjärven tutkimusalueelta paikannettiin algoritmin avulla peräti 78 sellaista tervahautaa, jotka olivat jääneet visuaaliseen tarkasteluun perustuvassa pilottitutkimuksessa (Ikäheimo 2021) ihmissilmältä huomaamatta. Tervahautojen lisäksi algoritmi paikansi alueelta yhteensä 177 hiilimiilua, joita tunnettiin ennestään 11 kappaletta. Tarkemmassa tarkastelussa myös osa aiemmin tervahaudoiksi rekisteröidyistä kohteista osoittautui hiilimiiluiksi. Näiden lisäksi algoritmi havaitsi yhdeksän suorakulmaista jäännöstä, esimerkiksi tukkikämppien ja rakennusten pohjia.

Ihmisen ja algoritmin suorittaman laserkeilausaineistoperusteisen paikantamisen erojen ymmärtämiseksi on hyödyllistä tarkastella lähemmin syitä, jotka johtivat näin huomattavaan kohteiden määrän lisääntymiseen hiljattain silmämääräisesti, mutta silti yksityiskohtaisesti läpikäydylle alueelle (Ikäheimo 2021). Tämän vuoksi jokaista pilottitutkimuksessa havaitsematta jäänyttä kohdetta (N=78) tarkasteltiin algoritmia varten tuotetuista visualisoinneista painottaen kohteen näkyvyyttä sekä näkyvyyttä selittäviä tekijöitä. Sellaisiin voidaan lukea erityisesti maankäytön eri muodot, kuten asuinrakentaminen, liikenteen infrastruktuuri, sekä metsänparannukseen tähtäävä maanmuokkaus.

Tarkastelu osoitti, että huomaamatta jääneistä tervahautoista kaikkiaan 33 oli sellaisia, joiden visuaaliseen havaitsemiseen ei ollut käytännössä mitään edellytyksiä. Näistä

lähes kaksi kolmasosaa sijaitsee maapohjaltaan läpikotaisesti äestetyllä alueella (Kuva 4a). Myös tervahaudan osittainen äestäminen ja joissain tapauksissa kohteen kiertäminen äestämisen yhteydessä vaikeuttavat kohteen havaitsemista (Kuva 4b). Silti aineistoon sisältyy 45 kohdetta, jotka olisi lähtökohtaisesti pitänyt havaita myös visuaalisessa tarkasteluun

perustuvassa tutkimuksessa. Tämä alleviivaa entisestään algoritmiin pohjautuvan lähestymistavan merkitystä alueellisesti yhtenäisen ja vertailukelpoisen tuloksen saamiseksi. Ihmistarkastelijan tarkkaavaisuuteen vaikuttavat aina monenlaiset osa- ja häiriötekijät, joista tietokone ja algoritmi eivät häiriinny, vaan suorittavat tehtävänsä ennako-oletusten ja



Kuva 4. Valikoima visualisoinneissa heikosti erottuvia tervahautakohteita: a) läpikotaisin äestetty; b) päältä ajettu; c) asutuksen tuntumassa; d) metsätien leikkaama ja päältä ajettu; e) rantatörmässä; f) kalliopaljastuman ympäröimä; g) soistuneen kankaan reunassa; h) kahden laserkeilausblokin rajalla; i) pienikokoinen kohde piirteettömässä maastossa.¹ Kuva: Janne Ikäheimo.

-määritysten mukaisesti samaan tapaan koko tarkastellun alueen laajuisesti.

Visuaalisessa tarkastelussa havaitsematta jääneiden kohteiden sijainnin ja olemuksen analyysissä nousi esille yleisimmän syyn, eli metsänparannustarkoituksissa suoritettua maapohjan muokkauksen myös muita kohteiden havaittavuuteen vaikuttavia tekijöitä. Pääosa näistäkin on ihmislähtöisiä. Ilmeisimpiä tekijöitä tässä suhteessa ovat asuinrakentaminen (Kuva 4c) sekä pellonraivaus, jotka tuottavat paikalliseen topografiiaan, ja näin ollen myös visualisointeihin, suurikontrastisia elementtejä, jotka vaikeuttavat pieni- ja piirteisten arkeologisten ilmiöiden havaitsemista. Sama huomio pätee yleisellä tasolla myös tieverkko. Jos tie ei suoranaisesti leikkaa tervahautaa, visuaalisena kiinnostopisteenä toimiva tiestö vetää helposti huomion pois lähiympäristöstään (Kuva 4d). Ihmisen epäsuora vaikutus tervahautakohteiden tuhoutumisen edistäjänä näkyy myös muun muassa vesivoimalaitosten sähköntuotannon tarpeista kumpuavasta järvien vedenkorkeuden säätelystä. Sen seurauksena etenkin Kiantajärven rannoilta tapaa rantatörmien eroosion myötä eriasteisesti tuhoutuneita tervahautakohteita (Kuva 4e).

Törmässä tai jyrkässä rinteessä sijaitsevia tervahautoja voi olla muutenkin vaikea havaita. Tämä riippuu ensisijaisesti tarkasteluun käytetystä visualisoinnista. Alkuperäisessä tutkimuksessa (Ikäheimo 2021) tervahautakohteiden tunnistamisessa hyödynnettiin pääasiassa 8-bittistä analyttistä vinovalovarjostusta sekä taivasnäkyä (Sky View Factor, SVF). Tässä artikkelissa raportoitava algoritmi hyödyntää puolestaan laskennallisesti kevyempää, maaston suhteellista korkeutta kuvaavaa TPI-visualisointia. TPI näyttää siis toimivan paremmin kohteissa, joiden lähiympäristössä esiintyy jyrkkiä korkeusvaihteluja.

Tervahaudan sijaintiympäristö vaikuttaa muutenkin sen havaittavuuteen. Tämä käy selväksi etenkin silloin, kun ympäristön pinnanmuodostus tuottaa visualisoinnissa

kohteeseen rinnastuvia samankaltaisia elementtejä. Tutkimusalueella tällaisiksi osoittautuivat mm. kalliopaljastumat (Kuva 4f), jäätikkösyntyiset moreeniharjanteet, sekä kankaiden soistuneet reunat (Kuva 4g). Maanmittauslaitoksen tuottaman laserkeilausaineiston tekninen jakelutapa 1 x 1 kilometrin blokeissa, jotka on koostettu yhdeksän neliökilometrin kokonaisuuksiksi (ks. Ikäheimo & Seitsonen 2021), vaikuttaa sekin kohteiden havaittavuuteen. Mikäli kohde sattuu sijaitsemaan juuri kahden laserkeilausblokin saumakohdassa (Kuva 4h), jää se pahimmassa tapauksessa havaitsematta niin algoritmilta kuin visualisointeja tarkasteltavalta tutkijaltakin.

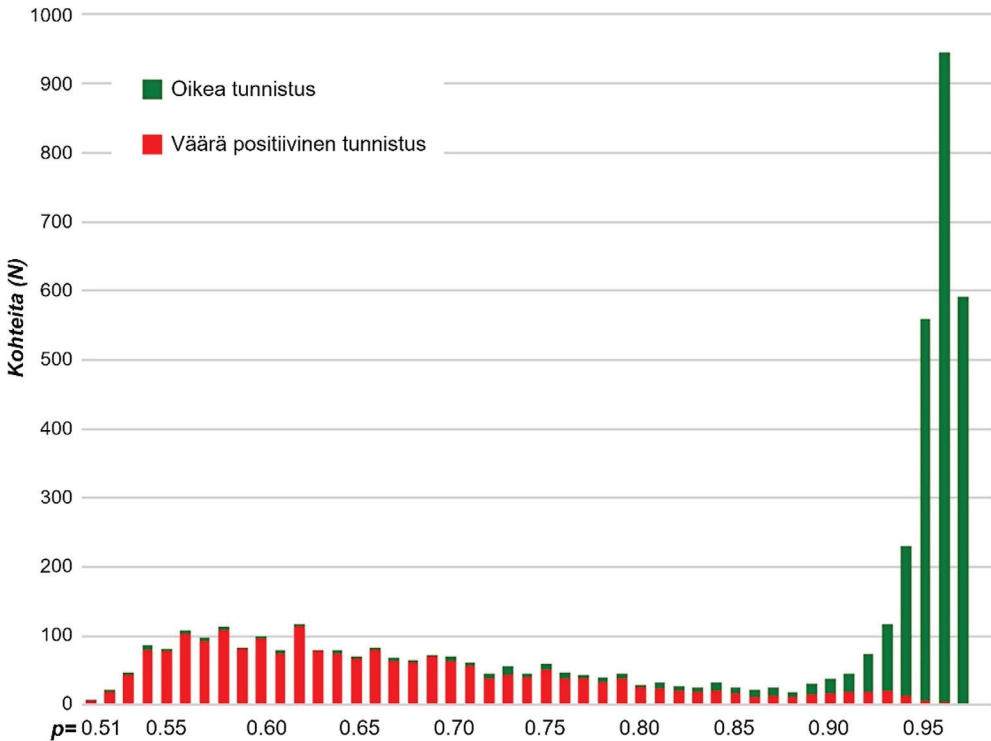
Viimeisin ja samalla ehkä ilmeisin havaitsemiseen vaikuttava tekijä on luonnollisesti tervahaudan koko. Voidaan perustellusti olettaa, että mitä pienempi kohde on kyseessä, sitä todennäköisemmin se jää tutkijalta havaitsematta visuaalisessa tarkastelussa (Kuva 4i). Varmistimme tämän hypoteesin vertaamalla kohteiden kokonaisjakaumaa kokoluokittain sekä visualisoinneista tunnistettuihin kohteisiin, että algoritmin tuottamiin lisähavaintoihin (Taulukko 1). Tulokset osoittavat, että algoritmi tunnistaa erityisesti uusia pieniä ja keskikokoisia tervahautoja, kun taas ihmisen suorittamassa tarkastelussa päästään verrattain kattavaan tulokseen kokoluokissa keskisuuresta valtavaan.

Kun jokaiselle mallin ennustamalle kohteelle on laskettu p-arvo, korkeamman arvon saaneet kohteet ovat lähtökohtaisesti todennäköisemmin tervahautoja ja matalamman p-arvon kohteet vääriä tunnistuksia (Kuva 5). Tunnistusalgoritmin tuloksia arvioidessa onkin syytä huomata, että jos tarkoituksena on huomioida kaikki mallin ehdottamat kohteet, olivatpa ne sitten tervahautoja tai toisella tavoin koulutetun ennustavan mallin hakemia muita muinaisjäännoiksi, on myös väärin positiivisten tunnistusten määrä suurempi. Jos raja-arvo lasketaan tässä mallissa perustasolle $p=0.5$, lopputulokseen sisältyvät kaikki tunnistukset, niin oikeat kuin väärät-

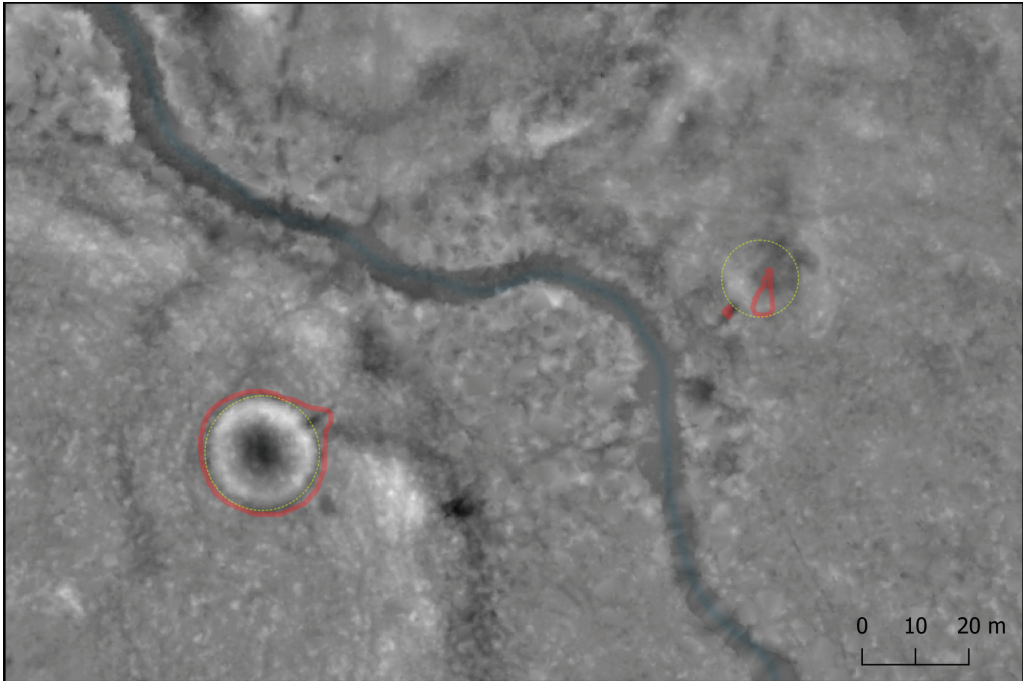
Taulukko 1. Tervahautojen suhteellinen (%) osuus Pesiöjärven tutkimusalueella kokoluokittain.

	Halkaisija	Tot	Man	Alg
Pieni	5–7,5 m	3,76	2,99	7,69
Pienukoinen	7,5–10 m	20,04	17,96	30,77
Keskikokoinen	10–12,5 m	30,69	30,42	32,05
Keskisuuri	12,5–15 m	22,96	24,94	12,82
Suurehko	15–17,5 m	12,73	13,22	10,26
Suuri	17,5–20 m	6,68	6,98	5,13
Hyvin suuri	20–22,5 m	2,30	2,49	1,28
Valtava	22,5–25 m	0,84	1,00	-

Halkaisija= tervahaudan läpimitta vallin ulkoreunoilta mitattuna; Tot= tutkimusalueelta tunnistetut tervahaudat (N=479); Man= tutkijan visualisoinneista tunnistamat kohteet (N=401, ks. Ikäheimo 2021); Alg= algoritmin tunnistamat uudet kohteet (N=78).



Kuva 5. Näljängän alueelta algoritmin tunnistamien kohteiden lukumäärä ja todennäköisyysarvot. Kuva: Floris Jan Groesz ja Niko Anttiroiko.



Kuva 6. Kaksi algoritmin tunnistamaa tervahautaa Suomussalmen Tervajärven kaakkoisreunalta: täydellisesti tunnistetun tervahaudan ($p= 7211573$; $i= 570364$) p -arvo on 0.97, vaillinaisesti ja kahtena osana tunnistetun tervahaudan ($p= 7211606$; $i= 570457$) saadessa huomattavasti matalampia arvoja – kehä 0.69 ja halssi 0.52. Tervahautojen ulkokehät on merkitty ohuella keltaisella katkoviivalla. Kuva: Janne Ikäheimo.

kin (Kuva 6). Raja-arvoa nostettaessa osa oikeista tunnistuksistaikin karsiutuu kuitenkin pois samalla kun väorien tunnistusten määrä vähenee oleellisesti. Raja-arvon alapuolelle jää kuitenkin aina myös oikeita tunnistuksia, jotka muodostavat ongelman aineiston automaattista mekaanista suodattamista ajatellen. Eräänlaisena optimiarvona voi pitää arvoa $p= 0.9$.

Mitä tunnistusalgoritmin sisäiseen toimivuuteen tulee, suoriutui Näljängän aineistolla koulutettu malli kolmen iteraation jälkeen seuraavasti, kun polygonien suodatuksen arvona käytetään todennäköisyyttä $p > 0.9$, jolloin kynnyksrajaa alapuolelle jäävät kohteet poistetaan aineistosta. Mallin yleinen tunnistustarkkuus on 93 % ja väorien positiivisten tunnistusten määrä on suhteellisen vähäinen

eli noin 3 % aineistosta. Sen sijaan mallilta tunnistamatta jääneiden, mutta aiemmin tunnettujen tervahautojen osuus oli 9 %.

Mitä enemmän aineistoa malli tuottaa, sitä pidemmän ajan sen tarkastaminen manuaalisesti vie. Aineiston suodattamisessa on siis loppupeleissä kyse hyväksyttävistä ”kustannuksista” tunnistus/työpanos -suhdetta optimoitaessa. Mallin käyttäjän täytyy siis tiedostaa ja hyväksyä tämä työpanokseen liittyvä periaatteellinen valinta. Lisäksi on muistettava, että aina sovellettaessa koulutettua mallia uuteen tutkimusalueeseen tai muinaisjäätännöstyyppeihin, voivat tulokset ja siten myös raja-arvot poiketa esimerkiksi maaperän, topografian ja kohdetyypin ominaisuuksien vuoksi oleellisesti aiemmasta.

Ranua-Simojoki

Testasimme kehittämämme tervahautoja tunnistavan mallin toimivuutta tuotantoalueista Ranua ja Simojoki muodostetulla tutkimusalueella. Tavoitteena oli tarkastella esimerkiksi erilaisen topografian, maankäytön ja maaperän vaikutuksia tuloksiin, sekä etenkin tunnistusten tarkastamiseen kuluva työpainosta ja -aikaa. Tunnistusalgoritmia ajettiin tässä tarkoituksessa Lenovo Thinkstation P340 Core i7 3.8 GHz -työasemalla, joka on varustettu 64GB:n keskusmuistilla, NVIDIA Quadro RTX 4000 8GB -näytönohjaimella ja Windows 10 -käyttöjärjestelmällä. Tuotantoalueita sekä tutkimustilannetta ja -tuloksia koskevat perustiedot on koottu taulukkoon (Taulukko 2). Python-ohjelmointikielellä kirjoitetut skriptit ajettiin Anaconda 3 -alustalle ladattujen avoimen lähdekoodin kirjastojen kautta selainpohjaisessa Jupyter Notebook -kehittämisympäristössä.

Ensiksi käsitellyltä Ranuan tuotantoalueelta algoritmi rajasi noin 4500 kohdetta, joiden tarkastamiseen manuaalisesti kului työtahdilla n. 700 kohdetta/h yksi henkilötyöpäivä. Käytännössä tarkastaminen suoritettiin QGIS 3.10 -ohjelmassa, jonne muodostettiin virtuaalirasteri TPI 30 -visualisoinneista. Sen päälle projisoidut polygonit käytiin läpi yksittellen ominaisuustietotaulukon kautta ”Pan to feature” -komentoa käyttäen. Polygonin tarkastelun yhteydessä se luokiteltiin johonkin seuraavista kategorioista: 0= väärä havainto, 1= tervahauta, 2= hiilimiilu ja 3= muu kohde. Viimeksi mainittu kategoria laajeni työskentelyn edetessä sisältämään joukon erilaisia ja epämääräisiä kuoppajäännöksiä, asumuspainanteita, rakennuksenpohjia ja rakennusryhmiä, joiden ajoituksen ja luonteen tarkempi selvittely vaatii omat jatkotutkimuksensa.

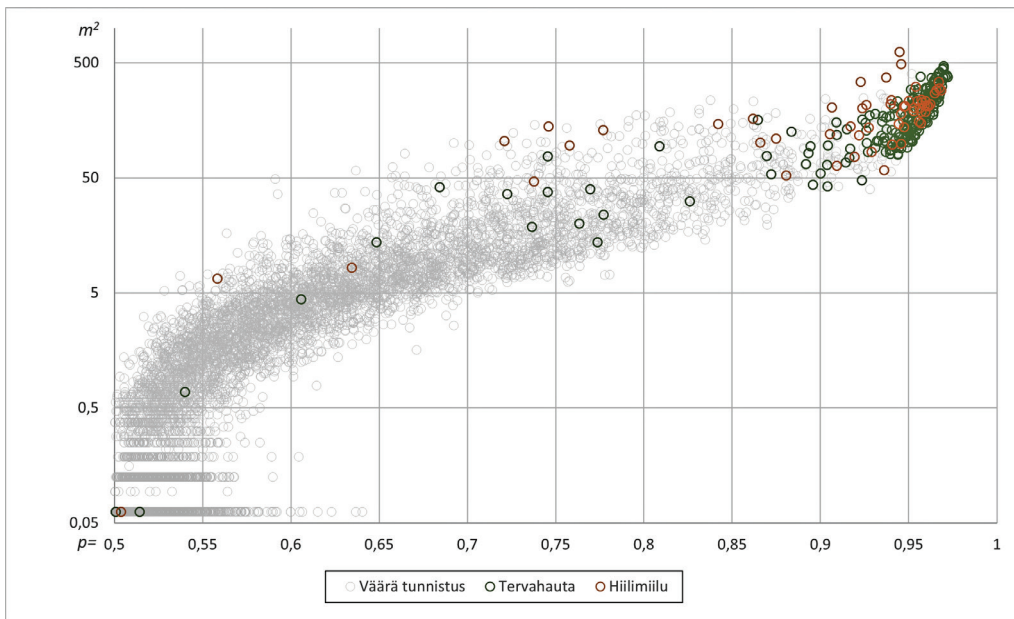
Taulukko 2. Tuotantoalueet Ranua & Simojoki: tunnuslukuja.

	Ranua	Simojoki
Keilausvuosi	2020	2021
Laajuus (km ²)	2848	2736
Tiedostokoko (GB)	159	110
Tunnetut tervahaudat		
Maanmittauslaitos	46	3
Museovirasto	13	1
LiDARK-prosessin kesto (h)		
DTM-tiedostot	3	3
TPI-tiedostot	7	7,5
tunnistaminen	22	39
LiDARK-tunnistukset		
polygoneja	4524	5188
tervahautoja	174	56
hiilimiiluja	3	122
muita kohteita	181	94

Simojoen tuotantoalueen yli 5000 kohdepolygonin kohdalla tarkastamisprosessia paikkatieto-ohjelmassa kehitettiin kirjaamalla väärät havainnot attribuuttitaulun sarakkeeseen vasta tarkastusprosessin lopuksi ja käyttämällä polygonien ominaisuustietojen tarkasteluun taulunäytön sijasta lomakenäyttöä. Aineiston tarkastaminen nopeutui tahtiin 2000 kohdetta per tunti, kun päälle kytkettiin ohjelman ”Automatically pan to current feature” ominaisuus ja polygonista toiseen siirryttiin nuolinäppäintä käyttäen. Tällöin ohjelma vilkuttaa ”Highlight current feature on map”-ominaisuus päälle kytkettynä polygonin sijainnin hetkellisesti visualisointipinnalla, vaikka polygonit sisältävä taso olisi näkyttömässä. Varoituksena mainittakoon, että tämä työtapa rasittaa tuntuvasti silmiä ja kasvattaa epileptisen kohtauksen riskiä, eli ominaisuuden käyttöä ei voi suositella kaikille.

Tunnistusalgoritmin tuottamat tulokset tutkimusalueella ovat kiinnostavia mo-

nessa mielessä. Tuloksista vähiten yllättävä on kohteiden lukumäärän moninkertaistuminen (Taulukko 2). Tervahautoja tunnistettiin aiempaan verrattuna noin nelinkertainen määrä. Tunnistettujen kohteiden jakauma painottuu suurikokoisiin ja todennäköisiin kohteisiin, kuten nämä kaksi muuttujaa ristiintaulukoiva kuvaaja osoittaa (Kuva 7). Pääosa tunnistuksista on alaltaan yli 50 m² ja todennäköisyysarvoltaan $p > 0.9$. Aineistoon tosin kuuluu myös kaksi pikseliparin avulla paikannettua ja arvon $p < 0.55$ omaavaa tervahautakohdetta. Algoritmilta jäi tunnistamatta kaksi ennestään tunnettua, osittain tuhoutunutta tervahautakohdetta, jotka molemmat sijaitsevat veden kuluttamassa ja romahtaneessa rantapenkassa. Lisäksi visuaalisessa tarkastelussa paikannettiin yksi tervahauta, jota algoritmi ei havainnut lainkaan. Koska algoritmin käyttämiä TPI_30 -visualisointeja ei käyty läpi tässä tarkoituksessa systemaattisesti, onkin mahdollista ja jopa todennäköis-



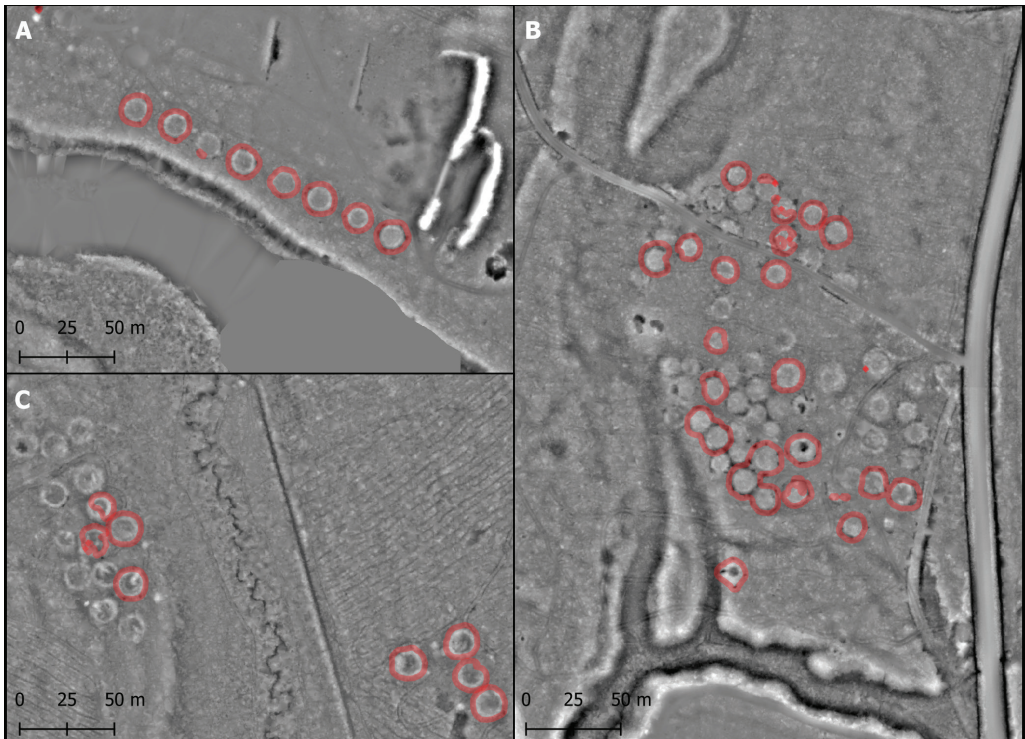
Kuva 7. Ranua-Simojoki-tutkimusalueen kohteet niiden pinta-alan ja todennäköisyysarvon mukaan visualisoituna. Kuva: Janne Ikäheimo.

tä, että tuotantoalueella sijaitsee tässä esitettyä enemmän tervahautoja. Paikantamatta jääneiden kohteiden määrä kokonaisuutta ajatellen lienee kuitenkin marginaalinen.

Tervahautojen lisäksi tutkimusalueelta löytyi aiemmin täysin tuntemattomia hiilimiiluja yhteensä 125 kappaletta. Niiden levintä painottuu Simojoen läheisyyteen, mistä yksittäisten miilujen ja miiluparien lisäksi paikannettiin useampi laaja hiilimiilukenttä. Lähes puolet kohteista, eli yhteensä 59 miilua, kuuluukin Iso-Näverryksen massiiviseen miilukenttään (Kuva 8b), joka sijaitsee Näverryksen kylän pohjoispuolella hieman yli 60 kilometriä Simojokea ylävirtaan. Raiskion parinkymmenen miilun kenttä sijaitsee siitä vielä 15 kilometriä jokivartta ylöspäin (Kuva 8c), kun taas pienempi ja olemukseltaan säännöl-

lisempi kahdeksan miilun keskittymä löytyi Kuivajoen Harjulan kylästä läheltä Oijärveä (Kuva 8a). Tätä Miiluranta-nimellä tunnettua kohdetta on paikallisten informanttien mukaan käytetty vielä 1950-luvulla. Onkin selvää, että puuhiilen tuotanto miilukentillä on ollut laajaa ja pitkäkestoista. Puuhiilen massatuotannon alku liittyyneen johonkin Pohjois-Pohjanmaan alueen varhaiseen teollisuuslaitokseen, esimerkiksi Iin Olhavassa 1700-luvun lopulla toimintaansa käynnistelleeseen Nybyn lasiruukkiin. Hiilentuotannon tarve on kuitenkin loppunut vasta toisen maailmansodan jälkeen, kun sodan aikana käytetyt häikäpönttöautot korvautuivat bensiinikäyttöisillä ajoneuvoilla.

Miilukentät sopivat hyvin myös kehittämämme algoritmin, ja yleisemmällä



Kuva 8. Algoritmin tunnistamia miilukenttiä tutkimusalueella Ranua-Simojoki (sijainneista, ks. Kuva 1). Kuva: Janne Ikäheimo.

tasolla myös semanttiseen segmentaatioon perustuvan U-Netin ongelmakohtien esittelyyn. Näistä ilmeisin on, että malli ylipäänsä tunnistaa automatisoidusti myös hiilimiiluja, vaikka se on koulutettu pelkästään tervahautoilla. Toisaalta sama ilmiö on havaittavissa myös arkeologien aiemmin rekisteröimiä tervahautoja tarkasteltaessa, sillä osa niistäkin on osoittautunut hiilimiiluiksi. Ehkä juuri koulutusaineistosta johtuen algoritmi ei kuitenkaan tunnista läheskään kaikkia miiluja, vaan ainoastaan osan niistä (Kuva 8). Tätä olisi syytä tarkastella erikseen pelkästään hiilimiiluilla koulutetun ennustavan mallin avulla. Aineistoon sisältyy myös sekä miilujen että tervahautojen kohdalla useita osittaisia tunnistuksia, joissa malli on havainnut esim. vain osan vallia tai pelkän halssiojan. Vähäisimmät niistä saattavat olla vain muutaman pikselin kokoisia ja mikäli useampi sellainen osuu samaan kohteeseen, näiden moninkertaisten tunnistusten siivoaminen aineistosta tuottaa melkoisesti lisätyötä. Toisaalta, koska semanttinen segmentointi ei erota vierekkäisiä alakohteita toisistaan, yksittäinen polygoni voi sulkea sisäänsä useamman alakohteen (esim. Kuva 8c), mikä puolestaan vaikuttaa kyseiselle tunnistukselle laskettuun pinta-ala ja todennäköisyyden sekä automaattisesti tunnistettujen kohteiden lukumäärään. Toisaalta tunnistettujen alakohteiden määrä yksittäisessä kohteessa antaa jonkinlaista osviittaa miilujen kokonaismäärästä, jota voidaan sitten täydentää visuaalisen tarkastelun avulla.

Onkin siis toistettava, että aineistolle suoritettu mekaaninen suodattaminen esimerkiksi todennäköisyyden tai polygonin koon perusteella johtaa työmäärän pienemisen ohella lähes aina myös oikeiden kohteiden menettämiseen. Esimerkiksi ”poista ala < 10m²” -suodattimen käyttäminen tutkimusalueella karsisi noin 80% läpikäytävästä polygoniaineistosta, mutta aiheuttaisi samalla neljän tervahaudan, kolmen hiilimiilun ja 258 muun potentiaalisen kohteen tappion. Toisaalta tämä säästäisi noin neljä tuntia aineiston tarkastamiseen käytettävää työaikaa.

Mikäli jossain tutkimuksessa tai jonkin aineiston käsittelyssä päädytään käyttämään polygoniaineiston mekaanista suodatusta, siitä sekä käytetyistä asetuksista on syytä aina tehdä merkintä tutkimus- tai selvityshankkeen dokumentaatioon ja raportteihin. Editoimattomien polygoniennusteiden pitkäaikainen tallentaminen on myös yksi varteenotettava vaihtoehto, sillä levytilasyöpöistä rasterivisualisoinneista poiketen niiden tiedostokoko ei tuota minkäänlaista ongelmaa. Esimerkiksi edellä käsitellyn tutkimusalueen vektoritiedostot ovat yhteiskooltaan vain 3 MB. Tämän toteuttaminen vaatii omat järjestelynsä tietoaineiston pitkäaikais säilytyksen ja arkistoinnin järjestämiseksi. Digitaalisten aineistojen arkistointiin ja tallennustilaan sekä käytettyjen menetelmien dokumentointiin liittyvät kysymykset ovat myös merkittäviä tekijöitä, jotka tulee ottaa huomioon koneoppimiseen perustuvien algoritmien lähestymistapojen laajempaa käyttöä suunniteltaessa.

Lopuksi

Tekoälyyn ja koneoppimiseen perustuva lähestymistapa aiemmin tuntemattomien arkeologisten kohteiden puoliautomaattisesti paikantamiseen tulee pilottitutkimustemme perusteella olemaan lähitulevaisuudessa merkittävä tekijä Suomen muinaismuistojen hallintoa ja tutkimusta hahmoteltaessa. Tervahautojen, sekä niiden oheistuotteena hiilimiilujen, paikantamiseen keskittyneet ensimmäiset syvemmälle menneet menetelmälliset kokeilut ovat nostaneet esiin monenlaisia edellä läpikäytyjä tekijöitä, jotka on syytä ottaa huomioon menetelmien jatkokehityksessä. Erityisen merkittävää on tiedostaa käsityön varsin suuri osuus algoritmien mallien koulutuksessa sekä automatisoidusti saatujen havaintojen tarkastelussa ja varmentamisessa. Tämä on tarpeen huomioida hankkeiden ja analyysien työmäärää sekä -panosta hahmoteltaessa. Manuaalista työpanosta voi

yrittää vähentää useilla tunneilla suodattamalla aineistoja mekaanisesti esimerkiksi ennustusten todennäköisyysarvojen tai ennustettujen kohteiden pinta-alan perusteella väärin positiivisten tunnistusten määrän minimoimiseksi. Tällöin on kuitenkin huomioitava ja myönnettävä, että mekanisoidun suodatuksen myötä osa todellisista kohteista menetetään. Aineiston mekaaninen suodatus esimerkiksi p-arvon perusteella tässä tarkastelluilla tutkimusalueilla ei muuttaisi merkittävästi kokonaiskuvaa, mutta aiheuttaisi joidenkin yksittäisten kohteiden menettämisen. Jos mekaaniseen suodatukseen joissain yhteyksissä päädytään, onkin tehtävä arkeologin näkökulmasta vaikeita päätöksiä siitä, kuinka paljon on hyväksyttävää ”heittää lapsia pesuveden mukana”. Muun muassa tämän takia on erityisen tärkeää, että aina algoritmisia menetelmiä käytettäessä kaikki tehdyt työvaiheet ja valinnat dokumentoidaan erityisen tarkasti, ja mahdollisuuksien mukaan myös työtiedostot tallennetaan tulevia tutkijoita varten. Tämä vaatii osaltaan erityisjärjestelyjä digitaalisten aineistojen pitkäaikaissäilytyksen ja arkistoinnin osalta (ks. esim. <https://arcticworldarchive.org/>).

Vaikka tässä esimerkkitapauksessa ennustava malli peittoosi selvästi ihmisen, ei algoritmikaan ole koskaan täydellinen. Se toteuttaa aina ainoastaan sille määritellyn tehtävän, eikä ota huomioon mitään sen ulkopuolelle jäävää. Kuitenkin tekoälyavusteinen laserkeilausaineiston käsittely on aina täsmällisempää, tuloksellisempaa, ja ehkä myös eettisesti kestävämpää, kuin yksittäisen tutkijan, tutkijaryhmän tai arkeologian harrastajan heuroinen yritys seuloa muinaisjäännökset esiin tutkimusalueeltaan joko manuaalisesti laserkeilausaineistojen visualisoinneista tai systemaattisiksi miellettyin kenttätyömenetelmin. Tämän pilottihankkeen puitteissa tunnistettujen kohteiden tarkastaminen maastossa olisi vaatinut huomattavia resursseja, eikä sitä ollut mahdollista toteuttaa tutkimusalueiden suuresta pinta-alasta ja kohdemäärästä johtuen. Tekoälyavusteisen kohteentunnistuksen

tehokkuudesta johtuen havaittujen kohteiden määrä tulee todennäköisesti jatkossakin ylittämään kyvyn tarkastaa havaintoja maastossa. Tämä korostaa tarvetta luoda alalle yhteiset käytännöt ja pelisäännöt sille, miten kaukokartoitusaineistoihin pohjautuvia havaintoja voidaan hyödyntää eri tarkoituksiin, huomioiden 5p-laserkeilausaineistojen myötä oleellisesti parantuneet edellytykset tunnistaa ja rajata kohteita luotettavasti.

Kiitokset

Kirjoittajat haluavat kiittää artikkelin sisältöä parantaneista kommentaista *Muinaistutkijan* kahta anonyymia vertaisarvioijaa. Artikkelin on laadittu Museoviraston koordinoimassa ja Maa- ja metsätalousministeriön rahoittamassa "Arkeologisten kohteiden tunnistaminen laserkeilausdatasta (LIDARK)" -hankkeessa (päätös VN/22710/2020-MMM-3).

Bibliografia

Arkistolähteet

Hamari, P. 2020. Arkeologisten kohteiden auto-maattinen tunnistaminen laserkeilaudatas-ta – LIDARK. Hankesuunnitelma. Helsinki: Museovirasto.

Verkkoaineistot

Čučković, Z. 2019. Topographic position index: a numpy based implementation. <<https://landscapearchaeology.org/2019/tpi/>> (Luettu 10.11.2021).

TensorFlow <<https://www.tensorflow.org/>> (Luettu 1.4.2021).

Tutkimuskirjallisuus

- Argyriou, A. V., Teeuw, R. M. & Sarris, A. 2017. GIS-based landform classification of Bronze Age archaeological sites on Crete Island. *PLoS ONE* 12(2): e0170727.
- Bonhage, A., Eltaher, M., Raab, T., Breuß, M., Raab, A. & Schneider, A. 2021. A modified mask region- based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the north German lowland. *Archaeological Prospection* 28: 177–186.
- Caspari, G., Balz, T., Gang, L., Wang, X. & Liao, M. 2014. Application of Hough forests for the detection of grave mounds in high resolution satellite imagery. *2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*: 906–909. DOI: <https://doi.org/10.1109/IGARSS.2014.6946572>
- Caspari, G. & Crespo, P. 2019. Convolutional neural networks for archaeological site detection – Finding “princely” tombs. *Journal of Archaeological Science* 110: 104998. DOI: <https://doi.org/10.1016/j.jas.2019.104998>
- Davis, D. S., Caspari, G., Sanger, M. C. & Lipo, C. P. 2021. Using deep learning to detect rare archaeological features: A case from coastal South Carolina, USA. *First International Meeting for Applied Geoscience & Energy*: 3265–3269.
- Gallwey, J., Eyre, M., Tonkins, M. & Coggan, J. 2019. Bringing lunar LiDAR back down to Earth: Mapping our industrial heritage through deep transfer learning. *Remote Sensing* 11(17): 1994.
- Garcia-Molsosa, A., Orenge, H. A., Lawrence, D., Philip, G., Hopper, K. & Petrie, C. A. 2021. Potential of Deep Learning Segmentation for the Extraction of Archaeological Features from Historical Map Series. *Archaeological Prospection* 28: 1–13.
- Girshick, R. 2015. Fast r-cnn. *Proceedings of the IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1109/ICCV.2015.169>
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. & Lew, M. S. 2016. Deep learning for visual understanding: A review. *Neurocomputing* 187: 27–48. DOI: <https://doi.org/10.1016/j.neucom.2015.09.116>
- Guyot, A., Lennon, M. & Hubert-Moy, L. 2021a. Objective comparison of relief visualization techniques with deep CNN for archaeology. *Journal of Archaeological Science: Reports* 38: 103027.
- Guyot, A., Lennon, M., Lorho, T. & Hubert-Moy, L. 2021b. Combined Detection and Segmentation of Archeological Structures from LiDAR Data Using a Deep Learning Approach. *Journal of Computer Applications in Archaeology* 4(1): 1–19. DOI: <https://doi.org/10.5334/jcaa.64>
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. 2017. Mask R-CNN, arXiv:1703.06870. In: *Proceedings of the IEEE International Conference on Computer Vision, 22–29 October 2017*: 2961–2969. Venice: IEEE. DOI: <https://doi.org/10.1109/ICCV.2017.322>
- Ikäheimo, J. 2021. Tervahautojen ilmalaserkeilausavusteinen työpöytäinventointi Suomussalmella. *Muinaistutkija* 3/2021: 2–15.
- Ikäheimo, J. & Seitsonen, O. 2021. Maanmittauslaitoksen uusi ja tarkempi laserkeilausaineisto (Laserkeilausaineisto 5p) kuoppajäännösten tutkimuksessa. *Muinaistutkija* 2/2021: 2–18.

- Kazimi, B., Thiemann, F. & Sester, M. 2019. Object Instance Segmentation in Digital Terrain Models. M. Vento & G. Percannella (toim.) *Computer Analysis of Images and Patterns. 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part II*: 488–495. Cham: Springer.
- Küçükdemirci, M., Landeschi, G., Dell'Unto, N. & Ohlsson, M. 2021. Mapping Archeological Signs From Airborne Lidar Data Using Deep Neural Networks: Primary Results. *ArcheoSciences* 45(1): 291–293. DOI: <https://doi.org/10.4000/archeosciences.10179>
- Lambers, K., Verschoof-van der Vaart, W. & Bourgeois, Q. 2019. Integrating Remote Sensing, Machine Learning, and Citizen Science in Dutch Archaeological Prospection. *Remote Sensing* 11(7): 794. DOI: <https://doi.org/10.3390/rs11070794>
- Orengo, H. A., Conesa, F. C., Garcia-Molsosa, A., Lobo, A., Green, A. S., Madella, M. & Petrie, C. A. 2020. Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data. *Proceedings of the National Academy of Sciences* 117(31): 18240–18250.
- Ren, S., He, K., Girshick, R. & Sun, J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6): 1137–1149.
- Ronneberger, O., Fischer, P. & Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. N. Navab, J. Hornegger, W. Wells & A. Frangi (toim.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Lecture Notes in Computer Science 9351. Cham: Springer.
- Seitsonen, O. 2013. LiDAR–kaukokartoitusaineistojen visualisointi ja analysointi: paikkatietoalgoritmeja arkeologeille. *Muinaistutkija* 1/2013: 2–16.
- Seitsonen, O. 2021. Kuoppia, kuoppia, enemmän kuoppia: Arkeologin keittokirja (pyynti) kuoppien tarkasteluun laserkeilausaineiston pohjalta. *Muinaistutkija* 3/2021: 16–37.
- Seitsonen, O. & Ikäheimo, J. 2021. Detecting archaeological features with airborne laser scanning in the Alpine tundra of Sápmi, Northern Finland. *Remote Sensing* 13(8): 1599. DOI: 10.3390/rs13081599.
- Tagil, S. & Jenness, J. 2008. GIS-based automated landform classification and topographic, landcover and geologic attributes of landforms around the Yazoren Polje, Turkey. *Journal of Applied Sciences* 8: 910–921.
- Trier, Ø. D., Salberg, A.-B. & Pilø, L. H. 2016. Semi-Automatic Mapping of Charcoal Kilns from Airborne Laser Scanning Data Using Deep Learning. *CAA2016 : Oceans of Data*: 221–32. Oslo: Archeopress.
- Trier, Ø. D., Rekstena, J. H. & Loseth, K. 2021. Automated Mapping of Cultural Heritage in Norway from Airborne Lidar Data Using Faster R-CNN. *International Journal of Applied Earth Observation and Geoinformation* 95: 102241.
- Verschoof-van der Vaart, W. B. & Lambers, K. 2019. Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications in Archaeology* 2(1): 31–40.
- Wachtel, I., Zidon, R., Garti, S. & Shelach-Lavi, G. 2018. Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China. *Journal of Archaeological Science* 92: 28–36.
- Yaworsky, P. M., Vernon, K. B., Spangler, J. D., Brewer, S. C. & Coddling, B. F. 2020. Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. *PLoS ONE* 15(10): e0239424.

Loppuviite

- 1 Kuvan 4 kohteiden sijaintikoordinaatit: a) p= 7201 517; i= 586 954; b) p= 7197 657, i= 587 426; c) p= 7197 073, i= 589 290; d) p= 7196 520, i= 586 708; e) p= 7204 217, i= 586 175; f) p= 7203 746, i= 589 249; g) p= 7205 643, i= 583 518; h) p= 7201 217, i= 575 000; i) p= 7200 434, i= 589 861.

Niko Anttiroiko on Museoviraston kulttuuriympäristöpalvelujen erikoisasantuntija.
niko.anttiroiko@museovirasto.fi

Floris Jan Groesz on Fieldin business manager.

Janne Ikäheimo on Oulun yliopiston arkeologian oppiaineen yliopistonlehtori.

Aleksi Kelloniemi on Oulun yliopiston arkeologian oppiaineesta valmistunut filosofian maisteri.

Stian Rostad on Fieldin projekti-insinööri.

Oula Seitsonen on Oulun yliopiston arkeologian oppiaineen tutkijatohtori.