



**TURUN
YLIOPISTO**

Historiallisten peruskarttojen syväoppiva karttaprosessointi

Pellot ja suot Suomessa 1949–1978

Iiro Seppä

Maantiede

Pro gradu -tutkielma

Laajuus: 30 op

Ohjaajat:

Niina Käyhkö

Jukka Heikkonen

29.04.2024

Turku

Pro gradu -tutkielma

Pääaine: Maantiede

Tekijä: Iiro Seppä

Otsikko: Historiallisten peruskarttojen syväoppiva karttaprosessointi: Pellot ja suot Suomessa 1949–1978

Ohjaajat: Niina Käyhkö, Jukka Heikkonen

Sivumäärä: 55 sivua + liitteet 1 sivu

Päivämäärä: 29.04.2024

Historialliset kartat sisältävät paljon arvokasta tietoa menneiden aikojen maankäytöstä. Tämän takia skannattujen karttojen muuttamista paikkatiedoksi, eli karttaprosessointia on tutkittu jo pitkään. Suomalaisia peruskarttoja on tosin karttaprosessoitu vain yhdessä aiemmassa tutkimuksessa, joka oli hyvin pienialainen. Tämän tutkielman tavoitteena oli laajentaa tämä alue peltojen ja ojittettujen sekä ojittamattomien soiden osalta kattamaan kaikki ensimmäisen valtakunnan peruskartoituksen aikana tuotetut ja myöhemmin skannatut peruskartat, eli lähes koko Suomen alue. Lisäksi tavoitteena oli arvioida tuotetun aineiston käyttökelpoisuutta.

Aluksi skannatut kartat georeferoitiin, ja georeferoinnin tarkkuus mitattiin. Tämän jälkeen digitoitiin käsin 359 km² koulutusaineistoa. Seuraavaksi syväoppiva konvoluutioneuroverkko, joka pohjautui U-net-arkkitehtuuriin, opetettiin tunnistamaan pellot sekä ojitetut ja ojittamattomat suot kaikista ennen vuotta 1979 valmistuneista peruskarttalehdistä, ja koulutuksen laatu arvioitiin. Lopuksi koulutetulla mallilla prosessoitiin 3090 peruskarttalehteä, jotka jälkikäsiteltiin ja kursittiin yhdeksi yhtenäiseksi aineistoksi.

Mallin kokonaistarkkuus oli 96,6 % ja Cohenin kappia oli 0,93. Georeferoitujen karttakohteiden sijaintivirheet olivat pääasiassa noin kymmenen metrin luokkaa. Tuotetun osituksen laatu vastaa aiemmassa tutkimuksessa peruskartoille tehtyä karttaprosessointia, mutta on tehty yli 300 kertaa suuremmalle alueelle.

Tuotettu aineisto tarjoaa mahdollisuuksia monenlaiseen peltojen ja soiden muutosten tutkimukseen 1950-luvulta eteenpäin.

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

Avainsanat: Karttaprosessointi, Maankäyttö, Historialliset kartat, Peruskartta, Konenäkö, Syväoppiminen

Master's thesis

Subject: Geography

Author: Iiro Seppä

Title: Deep learning map processing of historical Finnish Basic maps: Fields and mires in Finland 1949–1978

Supervisors: Niina Käyhkö, Jukka Heikkonen

Number of pages: 55 pages + 1 page appendix

Date: 29.04.2024

Historical maps include plenty of valuable information about land usage in times past. Therefore transforming scanned maps to digital geospatial data, also called map processing, has been researched for a long time. Despite this, there has been only one previous study, that has map processed historical Finnish Basic maps, and its study area was relatively limited. The objective of this Master's thesis was to expand the map processed area to cover almost the entire Finland, matching the extent of the first Basic mapping of Finland. Additional goal was to assess the usability of the produced dataset. The land use classes that were the focus of the study were fields, drained mires and undrained mires.

At first, the scanned maps were georeferenced, and the accuracy of the georeferencing was measured. Following that, 359 km² of the maps were digitized manually to create a training dataset. Next, a deep learning convolutional neural network based on the U-net architecture was trained to extract the fields, drained mires and undrained mires from the scanned maps, and the performance was evaluated. Finally the trained model was used to process 3090 Basic map sheets, and the results were postprocessed and combined to one dataset.

The total accuracy of the model was 96,6 %, and the Cohen's kappa was 0,93. Location accuracy of the georeferenced map objects was usually around 10 m. The produced segmentation's quality matches or exceeds the previous study on map processing Finnish Basic maps, but covers an area that is 300 times larger.

The produced dataset provides opportunities for many types of land use change research that concerns fields and mires from 1950's onward.

Key words: Map processing, Land use, Historical maps, Basic map, Computer vision, Deep learning

Sisällysluettelo

1 Johdanto.....	1
2 Tutkimuksen teoreettinen viitekehys.....	4
2.1 Keskeisiä käsitteitä.....	4
2.2 Syväoppiva konenäkö.....	6
2.3 Luokkaosituksen laatumittarit.....	10
2.4 Haasteet ja epävarmuuden lähteet karttaprosessoinnissa.....	12
2.5 Karttaprosessin menetelmät.....	14
2.5.1 Georeferointi.....	14
2.5.2 Skannattujen karttojen luokkaositus.....	17
3 Aineistot ja menetelmät.....	21
3.1 Tutkielman metodologinen kokonaisuus.....	21
3.2 Skannatut peruskarttalehdet 1949–1978.....	23
3.3 Muut aineistot.....	28
3.4 Peruskarttojen esikäsittely.....	29
3.5 Opetusaineiston digitointi ja käsittely.....	30
3.6 Koneoppimismallin kouluttaminen.....	31
3.7 Karttalehtien luokkaositus ja jälkikäsittely.....	32
3.8 Luokittelun tarkkuuden testaus.....	33
4 Tulokset.....	35
4.1 Esikäsittely ja digitointi.....	35
4.2 Rakennusten ja risteysten sijaintivirhe.....	36
4.3 Karttakohteiden ositus.....	38
4.4 Osituksen laadunarviointi.....	42
5 Tulosten tarkastelu.....	49
5.1 Mallin ositustarkkuus.....	49
5.2 Tuotetun aineiston käyttökelpoisuus.....	50
5.3 Syväoppivan karttaprosessin haasteet ja mahdollisuudet.....	52

6 Aineistojen saatavuus.....	54
Kiitokset.....	55
Lähteet.....	56
Liitteet.....	67
Liite 1. Taulukot.....	67

1 Johdanto

Suomen maankäytössä tapahtui merkittäviä muutoksia 1900-luvun loppupuolella.

Koneellistuminen mahdollisti aiempaa tehokkaamman maa- ja metsätalouden sekä järeämmän maanmuokkauksen, kuten syvempien ojien tekemisen tehokkaasti. Maatalouden työvoiman tarve ja tilamäärät vähenivät, kun taas tilakoot kasvoivat (Voutilainen ym., 2012).

Ojitusteknologian kehityksen seurauksena Suomen 1950-luvun alun 9 miljoonan hehtaarin suoalasta ojitettiin vajaat 5 miljoonaa hehtaaria, jotta tuolloin hyödyttömiksi koetut suot saataisiin puuntuotantokäyttöön (Korhonen ym., 2017). Ojitukset tehtiin pääasiassa 50-, 60- ja 70-luvuilla. Lisäksi yhteensä noin vajaa miljoona hehtaaria suota on muutettu pelloiksi, suuri osa ennen 1960-lukua (Myllys, 1996; Vasander, 2006). Puuntuotannon näkökulmasta ojituksen voidaan katsoa jossain määrin onnistuneen, koska noin kolme miljoonaa hehtaaria suota muuttui kasvillisuudeltaan kangasmetsäkasvillisuutta muistuttavaksi turvekankaaksi (Korhonen ym., 2017). Samalla suo-ojitukset ovat kuitenkin aiheuttaneet merkittävää haittaa ympäristölle, ja ne ovat merkittävin uhanalaistumisen syy monille suoluontotyypeille (Kaakinen ym., 2018). Kuivuneen suon tilalle kasvaneet metsät ovat myös lajistoltaan melko yksipuolisia, vaikkakin poikkeuksiakin tähän on esim. ruohoturvekankailla (Hotanen ym., 2015; Kaakinen ym., 2018). Etenkin rämesyntyiset turvekankaat ovat lajistoltaan suhteellisen köyhiä kivennäismaan metsiin verrattuna (Hotanen ym., 2006). Ojitetulla suolla on kuitenkin ympäristövaikutuksia myös kyseistä suota laajemmalla alueella. Ojitetut suot päästävät vesistöjä rehevöittävästä fosforia ja typpeä sekä tummentavaa humusta alavirran vesistöihin vielä vuosikymmeniä ojituksen jälkeen, turpeen kuivumisen ja hajoamisen takia (Finér ym., 2020). Vuonna 2020 turpeen hajoamisesta johtuvien kasvihuonekaasupäästöjen on arvioitiin olevan noin 17 Mt CO₂-ekvivalenttia, kun Suomen nettopäästöt olivat samana vuonna 30 Mt CO₂-ekvivalenttia (Lång ym., 2022). Näiden ympäristömuutosten täsmällisten vaikutusten ja dynamiikan ymmärtämiseksi tarvitaan luotettavaa ja tarkkaa tietoa sekä nykyisestä, että menneestä maankäytöstä (Barnosky ym., 2012; Zu Ermgassen ym., 2012).

Yksi tällainen historiallisen maankäyttötiedon lähde ovat historialliset kartat. Ne tarjoavat melko yksityiskohtaista, sijainniltaan tarkkaa, maastotarkastettua ja alueellisesti kattavaa informaatiota menneiden aikojen ihmistoiminnasta ja ympäristöstä (Vuorela ym., 2002).

Vanhat kartat siis sisältävät tietoa, jota ei ole helposti saatavilla muista maankäytön historiaa valottavista aineistoista, kuten ilmakuvista tai tilastoista. Ilmakuvat ovat yksityiskohtaisia ja alueellisesti kattavia, mutta maanpeiteluokkien erottaminen niistä vaatii joko työlästä käsityötä tai yleensä suhteellisen virhealtista automaattista prosessointia (Ratajczak ym.,

2019). Kartoittaja on tulkinut kartalla esitettävän maankäytön maastossa, joten skannatut kartat sisältävät myös informaatiota, jota ilmakuviissa ei ole saatavilla. Ilmakuvat eivät esimerkiksi sisällä tietoa puunlatvojen alla olevasta maanpeitteestä (Bhattacharjee ym., 2021). Tilastot taas eivät sisällä sijaintitietoa, joka on tarpeen, jos maankäyttötietoja haluttaisiin yhdistää paikkatietomenetelmin muihin aineistoihin. Monet historialliset karttasarjat ovat nykyään myös avoimesti saatavilla skannattuina valokuvina (esim. ‘Vanhat painetut kartat’, 2023; ‘Mapy archiwalne Polski i Europy Środkowej’, 2024; ‘Historische Karten’, 2024; ‘Historiska kartor’, 2024; ‘Historical Topographic Maps - Preserving the Past’, 2024). Ne eivät kuitenkaan ole sellaisenaan käytettävissä paikkatietoanalyysissä, vaan ne vaativat karttaprosessointia eli georeferointia ja maankäyttöluokkien tunnistamista skannatusta kuvasta. Laaja-alaisuutensa takia karttojen digitointi käsin on hyvin työvoimaintensiivistä (Ostafin ym., 2017). Tämän takia karttaprosessointi on ollut aktiivisen tutkimuksen kohteena vuosikymmeniä (Boyle, 1980; Wu ym., 1994; Reiher ym., 1996; Bin & Cheong, 1998; Wise, 2002; Khotanzad & Zink, 2003; Dhar & Chanda, 2006; Leyk ym., 2006; Leyk & Boesch, 2009; Mello ym., 2012; Chiang ym., 2014; Iosifescu ym., 2016; Liu, Miao, Tian, ym., 2016; Liu, Miao, Xu, ym., 2016; Ostafin ym., 2017). Pääasiassa nämä ei-syväoppivien menetelmin tehdyt tutkimukset ovat keskittyneet värien erotteluun, mutta löysin myös yhden tutkimuksen, joka oli onnistunut muotoihin perustuvassa symboleiden tulkinnassa ilman koneoppimista 95 %:n tarkkuudella (Leyk ym., 2006). Viimeisten neljän vuoden aikana syväoppivat kuvantunnistusmenetelmät ovat tehneet läpimurron myös karttaprosessoinnissa, sekä georeferoinnissa, että luokkaosituksessa. Nämä menetelmät ovat mahdollistaneet etenkin monimutkaisempien ja muiden symboleiden kanssa samanväristen symboleiden tunnistamisen aiempaa tarkemmin ja paljon helpommin (Uhl ym., 2020; Ståhl & Weimann, 2022; Wu ym., 2022; Mäyrä ym., 2023). Syväoppivien menetelmien hyödyntämistä georeferoinnin automatisoinnissa on tutkittu vähemmän, mutta tulokset ovat olleet riittävän tarkkoja (Dong ym., 2018; Burt ym., 2020). Myös suomalaisia historiallisia peruskartoja on ehditty myös jo karttaprosessoimaan ja hyödyntämään maankäytön muutosten tutkimukseen, mutta vain muutaman karttalehden alueelta (Mäyrä ym., 2023). Valtaosa Suomen historiallisista peruskartoista on siis vielä täysin karttaprosessoimatta.

Tämän opinnäytteen tavoitteena on muuttaa historialliset, kaikkien vuosina 1949–1978 valmistettujen skannattujen peruskarttalehtien pellot ja ojitetut sekä ojittamattomat suot mahdollisimman korkealaatuiseksi paikkatietoaineistoksi ja arvioida sen käyttökelpoisuutta maankäytön muutosten tutkimukseen. Tämä mahdollistaa tulevaisuudessa aiempaa paremman tiedon tuottamisen soiden ja peltojen muutoksista Suomessa 1900-luvun jälkipuolella.

Tutkielma pyrkii vastaamaan seuraaviin kysymyksiin:

1. Kuinka tarkasti konvoluutioneuroverkot pystyvät tunnistamaan alue- ja viivamuotoisia karttakohteita vanhoista skannatuista kartoista?
2. Kuinka käyttökelpoinen tuotettu aineisto on maankäytön muutosten tutkimukseen?
3. Millaisia haasteita ja mahdollisuuksia syväoppivaan karttojen ositukseen liittyy?

2 Tutkimuksen teoreettinen viitekehys

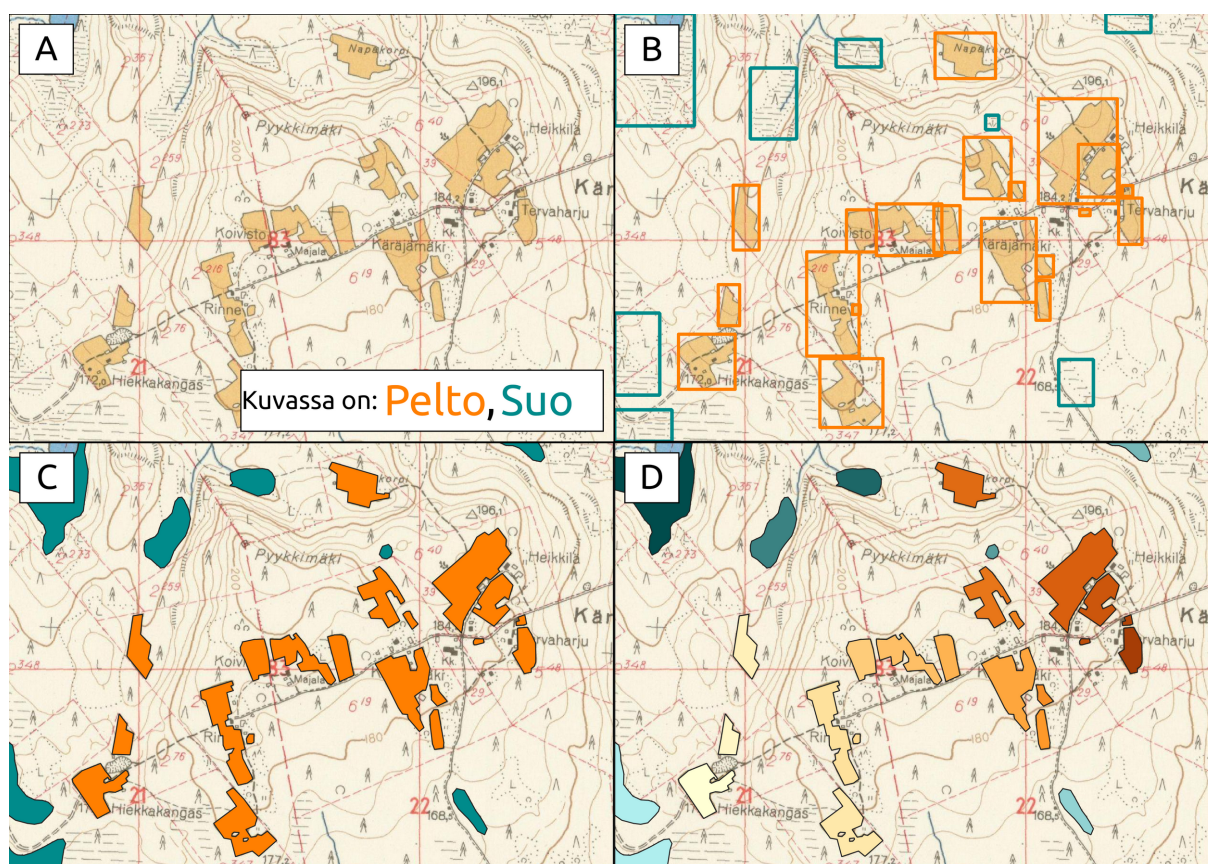
2.1 Keskeisiä käsitteitä

Kartta on maantieteellinen, yksinkertaistettu visualisointi asioiden ominaisuuksista ja sijainnista. **Maanpeite** ja **maankäyttö** ovat yleisiä kartoilla kuvattavia asioita. Maanpeite kuvaa maanpinnan biologisia ja fyysisiä ominaisuuksia (Wang ym., 2023). Esimerkkejä maanpeiteluokista ovat esim. puustot, ruohikot, vesistöt ja hietikot. Vaikka maanpeite kuvaakin fyysisen maailman ominaisuuksia, se ei ole siltikään täysin objektiivista, koska maailman jaottelu kartoitettaviin luokkiin johtaa väistämättä yksinkertaistuksiin, joihin sisältyy myös arvovalintoja. Maankäytöllä taas tarkoitetaan maanpeitteen luokittelua ihmisen ja yhteiskunnan toiminnan näkökulmasta (Wang ym., 2023). Esimerkiksi maanpeitteeltään puustoinen alue voi olla maankäytöltään niin puisto, talousmetsä, suojelualue kuin maatalousalue. **Topografisella kartalla** tarkoitetaan etenkin historiallisessa kontekstissa käytännössä maastokarttaa, vaikka kirjaimellisesti nimi tarkoittaa korkeuden kuvaamista. Niiden pääpaino on useimmiten maankäytössä, mutta niissä esiintyy myös kohteita, jotka ovat lähempänä maanpeitteen kuin maankäytön kuvausta ('Peruskartta 1:20 000.', 1948; 'Historiska kartor', 2024; 'Historische Karten', 2024). Maankäytön ja maanpeitteen raja on kuitenkin usein varsin häilyvä.

Karttaprosessoinnilla (eng. Map processing) tarkoitetaan tietojenkäsittelyllisiä menetelmiä, joilla valokuvatuisissa (usein skannatuissa) fyysisissä kartoissa kuvatut karttasymbolit saadaan tunnistettua ja muutettua paikkatiedoksi puoli- tai täysiautomaattisesti (Chiang ym., 2014; Liu ym., 2019). Karttaprosessointi voidaan jakaa karkeasti kahteen osa-alueeseen, **georeferointiin** (eng. Georeferencing) ja **digitointiin** (eng. Digitizing). Georeferoinnilla (eng. Georeferencing) tarkoitetaan kuvan asettamista oikeaan maantieteelliseen sijaintiin koordinaattijärjestelmässä (Bahgat & Runfola, 2021). Digitointi puolestaan tarkoittaa piste-viiva- tai aluekohteiden erottamista kuvasta.

Konenäkö (eng. Computer vision) tarkoittaa tietojenkäsittelymenetelmiä, joilla kuva-aineistoista saadaan eriytettyä käyttökelpoista informaatiota automaattisesti. Kuvien sisältöä voi jaotella luokkiin monella tavalla. Neljä yleisesti käytettyä tapaa ovat **luokittelu**, **kohteentunnistus**, **luokkaositus** ja **tapaositus** (eng. Classification, object detection, class segmentation & instance segmentation) (Tian ym., 2021). Luokittelussa kuvassa olevat luokat eivät saa tarkempaa sijaintia (Kuva 1A). Kohteentunnistuksesta puhutaan jos luokittelija

tuottaa vain luokkaennusteen ja kohdetta rajaavan suorakaiteen (Kuva 1B, (Alom ym., 2019; Alzubaidi ym., 2021; Tian ym., 2021)). Luokkaosituksella tarkoitetaan kuvan jakamista pikselikohtaisesti alueisiin siten, että yhdellä alueella on yhdentyypistä luokkaa (Alom ym., 2019; Alzubaidi ym., 2021). Skannattujen karttojen tapauksessa tämä tarkoittaa esimerkiksi sitä että kaikki pellot on merkattu samaan joukkoon (Kuva 1C). Luokkaosituksen vastinparina on tapausositus, joka tarkoittaa että jokaisen luokan jokainen erillinen tapaus rajataan omaksi joukokseen pikselintarkasti (Tian ym., 2021). Peltojen kohdalla tämä tarkoittaisi jokaisen toisiinsa koskemattoman pellon alueen rajaamista erikseen (Kuva 1D).



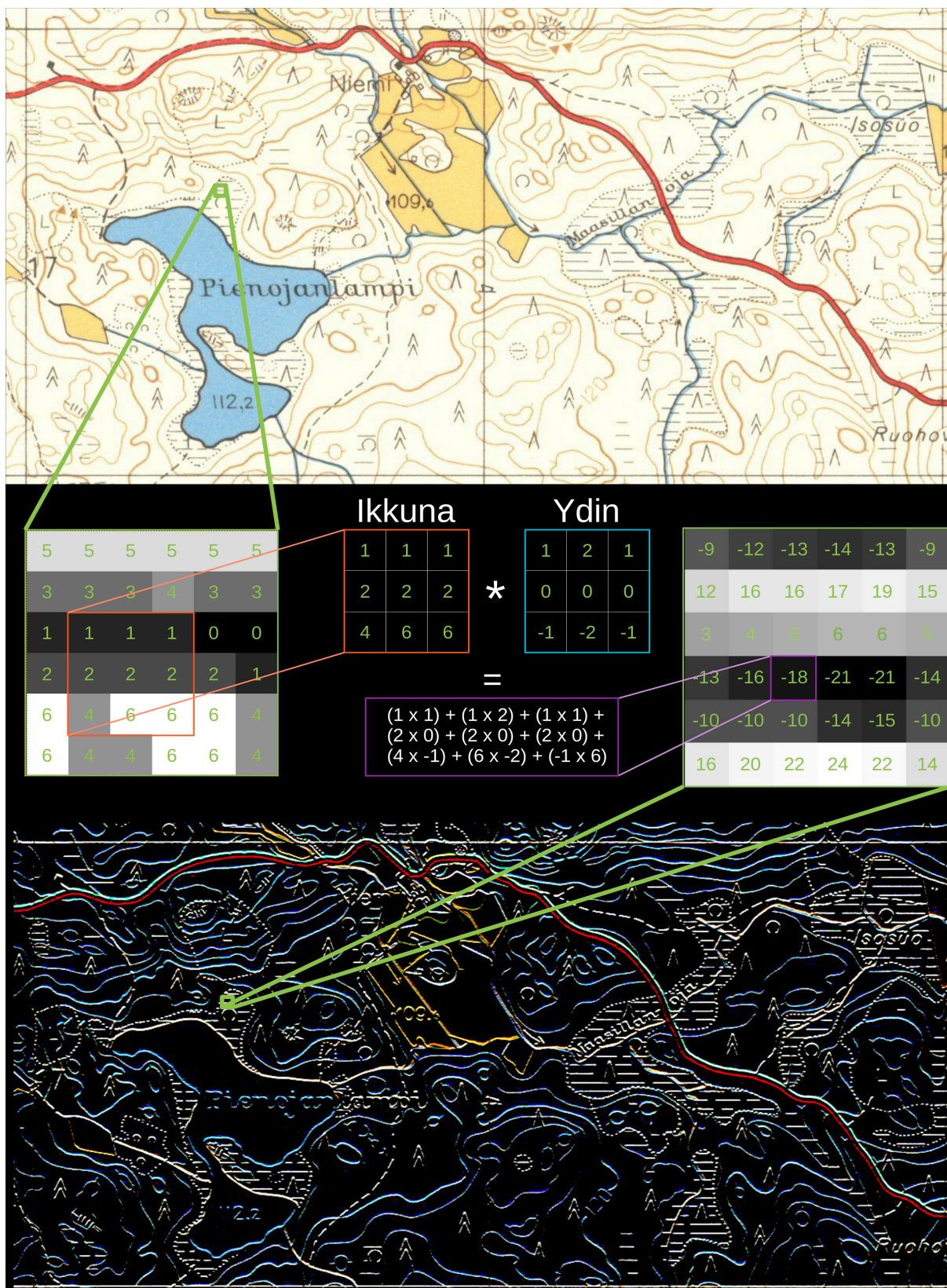
Kuva 1: Esimerkki erityyppisistä kuvanluokittelutavoista skannatulla kartalla. A) Luokittelija kertoo kuvan luokan, muttei tuota tarkempaa sijaintitietoa. B) Kohteentunnistuksessa malli ennustaa sekä luokan, että kohdetta rajaavan ruudun. C) Luokkaosituksessa kaikki pellot tai suot kuuluvat samaan joukkoon pikselikohtaisesti. D) Tapausosituksessa jokainen erillinen pelto tai suo muodostaa oman joukonsa pikselikohtaisesti. Taustakartta: (Peruskarttalehti 2333 10+331101 Vuorilahti, 1966)

Syväoppiminen (eng. Deep learning) on tekoälyn ja koneoppimisen osa-alue, joka pohjautuu keinoalgoritmeihin ja neuroverkkoihin (Alom ym., 2019; Alzubaidi ym., 2021). Nykyistä muistuttavat keinoalgoritmit keksittiin jo 50-luvulla, mutta niiden hyödyllisyys pysyi pitkään suhteellisen rajallisena muihin menetelmiin verrattuna (Schmidhuber, 2015). Se mikä

erottaa syväoppivat neuroverkot vanhoista neuroverkoista on niiden arkkitehtuurinen monimutkaisuus, suuruus ja syvyys, minkä algoritmiikan kehitys ja näytönohjainlaskenta ovat mahdollistaneet (Schmidhuber, 2015; Alom ym., 2019). Syväoppimisen suurimpana etuna on, että neuroverkko oppii itsenäisesti opetusaineistosta sen tehtävän kannalta oleelliset ominaisuudet (Alom ym., 2019; Alzubaidi ym., 2021). Toisaalta tämän takia syväoppiminen reagoi herkästi vinoumiin tai virheisiin käytetyssä aineistossa, ja malli saattaa tämän seurauksena oppia epätoivottuja asioita (Alom ym., 2019; Alzubaidi ym., 2021; Li ym., 2023). Tätä kutsutaan tekoälyjärjestelmän **huonoksi suuntautumiseksi** (eng. Artificial intelligence misalignment) (Dung, 2023). Yksi yleinen esimerkki tällaisesta tilanteesta on **ylikouluttautuminen** (eng. Overfitting), jossa malli oppii ulkoa käytetyn opetusaineiston, ja saavuttaa hyvin korkean tarkkuuden sillä, mutta pärjää uudenlaisten esimerkkien kanssa vain heikosti (Alom ym., 2019; Alzubaidi ym., 2021; Li ym., 2023). **Inferenssillä** tarkoitetaan ennusteiden tekemistä aiemmin koulutetulla mallilla (Alzubaidi ym., 2021). **Epookka** puolestaan tarkoittaa koulutusjaksoa, jonka aikana koulutettava malli näkee koko koulutusaineiston kertaalleen (Alzubaidi ym. 2021).

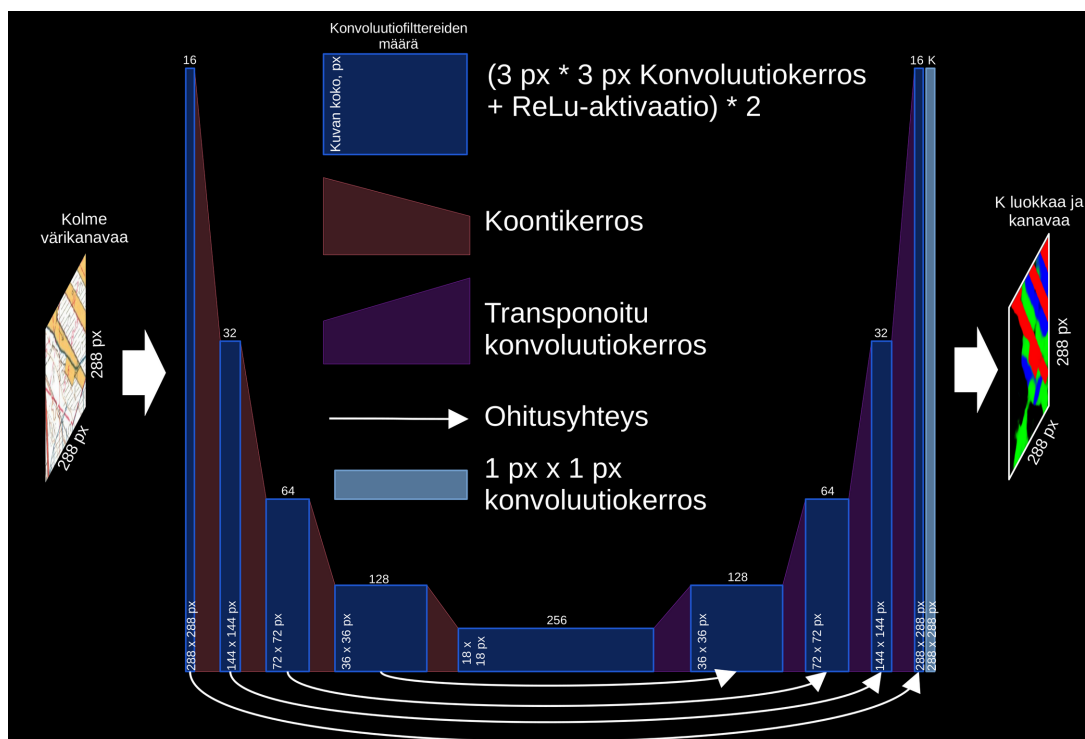
2.2 Syväoppiva konenäkö

Yleisin kuvantunnistustehtäviin käytettävä syväoppiva neuroverkkotyyppi on **konvoluutioneuroverkot** (eng Convolution Neural Network, CNN) (Alom ym., 2019; Alzubaidi ym., 2021), vaikka viime vuosina myös **näkömuuntimet** (eng. Vision Transformer, ViT) ovat tulleet kilpailukykyisiksi etenkin suurilla aineistoilla koulutettuna (Dosovitskiy ym., 2021; Han ym., 2023). Konvoluutioneuroverkot koostuvat tasoista, joista tärkeimmät ovat nimen mukaisesti konvoluutio-operaatioita suorittavia konvoluutiotasoja. Ne ovat liikkuvan ikkunan menetelmiä, joissa kuvan yli liu'utetaan pikseli pikseliltä **ydintä** (eng. Kernel) (Alom ym., 2019; Alzubaidi ym., 2021). Tyypillinen ytimen koko on 3 x 3 pikseliä, jolloin tarkastelussa on aina kerrallaan mukana keskuspikseli ja kaikki siihen reunoistaan tai kulmistaan koskevat pikselit. Ytimen jokaisella solulla on painotuskerroin, joiden perusteella lasketaan painotettu tulojen summa ikkunassa näkyvien pikselien arvojen kanssa. Tämä voidaan myös muuttaa keskiarvoksi jakamalla ytimen pikseleiden määrällä. Tämä keskiarvo annetaan käsitellyn kuvan pikselin arvoksi (Kuva 2). Kaikki tähän asti kuvattu koskee myös tavanomaisia konvoluutiofilttereitä, jotka ovat yleisiä kuvankäsittelyssä (Oberholzer ym., 1996).



Kuva 2: Esimerkki konvoluutiofilterin toiminnasta. Ylimpänä on leike skannatusta peruskarttalehdestä. Siihen on merkattu vihreällä pieni osaleike, jonka yksinkertaistettu pikseliarvojen esitys on suurennettuna. Kuvan yli liikutetaan 3 x 3 pikselin ikkunaa (punaisella) ja ikkunan arvoista lasketaan ytimen (sinisellä) arvoilla pikselikohtaisesti painotettu tulojen summa (violetilla). Tästä summasta tulee ikkunan keskipikselin uusi arvo, ja ikkuna siirtyy yhden pikselin oikealle. Alimpana on esimerkissä näytetyn ytimen tulos koko kuvalle. Esimerkin ydin on vaakasuuntaisia viivoja tunnistava Sobel-filteri. Karttalehti: (Peruskarttalehti 2042 01 Karisjärvi, 1958)

Konvoluutioneuroverkoista tekee erityisiä se, että niissä ytimen solujen painotukset opitaan koulutuksen aikana. Konvoluutioneuroverkkojen arkkitehtuurit vaihtelevat paljon, mutta yleisesti voidaan sanoa että niiden tärkeimpiä osia ovat **konvoluutiokerrokset** (eng. Convolution layer), joissa tehdään rinnakkain lukuisia konvoluutio-operaatioita sekä **koontikerrokset** (eng. Pooling layer), joissa käsiteltävien kuvien kokoa pienennetään, jotta laskenta-aika ja koulutettavien muuttujien määrä eli koulutukseen vaadittavan aineiston määrä ei kasva liian suureksi (Alom ym., 2019; Alzubaidi ym., 2021). Konvoluutio- ja koontikerroksia asetetaan peräkkäin useita, ja niiden väliin voidaan asettaa myös muita kerroksia, kuten **aktivaatio**- (eng. Activation) **normalisointi**- (eng. Normalization) tai **tiputuskerroksia** (eng. Dropout). Syvemmälle verkossa edetessä kuvien resoluutio heikkenee, mutta informaation määrä kuvan sisällöstä kasvaa. Tämän takia tilanteissa, joissa kuva halutaan osittaa eri luokkiin, liitetään neuroverkon loppupuolelle resoluutiota kasvattavia kerroksia, kuten **transponoituja konvoluutioita** (eng. Transposed convolution) (Alom ym., 2019; Alzubaidi ym., 2021; Ji ym., 2021). Usein tässä käytetään hyödyksi aiempien tasojen tuloksia, jotka sisältävät enemmän tilallista informaatiota mutta vähemmän merkitysinformaatiota. Ne liitetään myöhempiin tasoihin, joiden koko on kasvatettu vastaamaan alkupään tasoa. Näitä kutsutaan ohitusyhteyksiksi (eng. Skip connection). Tunnetuin esimerkki edellä mainittua rakennetta käyttävistä syväoppivista kuvantunnistusarkkitehtuureista on U-net (Kuva 3; Ronneberger ym., 2015).



Kuva 3: U-netin rakenne. Kuvan koko ja filterien määrä voi vaihdella aineiston tarpeiden mukaan.

Syväoppimisen nopeasti kasvaneelle suosiolle kuvantunnistuksessa on monia syitä. Ensinnäkin niiden tarkkuus on suurilla opetusaineistoilla usein parempi kuin aiempien tekniikoiden (Alom ym., 2019; Alzubaidi ym., 2021; Wu ym., 2022). Toiseksi niiden arkkitehtuurit ovat yleiskäyttöisiä, eli samaa koodia pystyy käyttämään mihin tahansa kuvantunnistukseen opetusaineistoa vaihtamalla (Alom ym., 2019; Alzubaidi ym., 2021). Tämä tarkoittaa että uusien arkkitehtuurien kehittäminen hyödyttää melkein kaikkia kuvantunnistuksen alalajeja. Kolmanneksi eduksi voidaan katsoa mahdollisuus siirto-oppimiseen ja koulutettujen mallien kyky hyväksyä samanaikaisesti aineistoja erityyppisistä lähteistä, kunhan neuroverkko on koulutettu siihen (Thung & Wee, 2018; Alom ym., 2019; Alzubaidi ym., 2021). Yleisesti voidaan todeta, että syväoppivat menetelmät toimivat sitä paremmin, mitä enemmän monipuolista ja laadukasta aineistoa on saatavilla (Alom ym., 2019; Alzubaidi ym., 2021; Li ym., 2023). Kaikissa aihepiireissä riittävän suuren aineiston tuottaminen ei kuitenkaan ole realistista tai mahdollista. **Siirto-oppimisella** tarkoitetaan prosessia, jossa laajalla aineistolla koulutettua neuroverkkoa opetetaan uuteen tehtävään, mutta uusi aineisto ainoastaan hienosäätää neuroverkon viimeisiä tasoja (Bansal ym., 2022). Tämä toimii, koska yleensä neuroverkkojen ensimmäiset tasot tunnistavat matalan tason ominaisuuksia, kuten muotoja tai tekstuureita (Alzubaidi ym., 2021). Siirto-oppiminen on hyödyllistä, koska syväoppiminen vaati runsaasti opetusdataa, ja siirto-oppiminen vähentää tarvittua uuden opetusaineiston määrää huomattavasti. Esikoulutusaineiston olisi kuitenkin hyvä olla mahdollisimman samankaltaista kuin varsinainen koulutusaineisto (Raghu ym., 2019).

Yleisesti voidaan sanoa, että konvoluutioneuroverkoilla luokittelu on yleensä helpointa, kohteentunnistus ja luokkaositus hieman vaikeampia ja tapausositus vaikeinta (Tian ym., 2021). Tämä johtuu siitä, että konvoluutioneuroverkot pystyvät tuottamaan hyvätason luokittelijan suhteellisen yksinkertaisilla arkkitehtuureilla (esim. Krizhevsky ym., 2012). Tämän jälkeen muut kuvantunnistusmenetelmät pystytään johtamaan luokittelijasta. Kohteentunnistus voidaan toteuttaa ennustamalla rajaava suorakaide ja sitten luokittelemalla sen sisältö (Ren ym., 2017). Luokkaositus on puolestaan mahdollista toteuttaa liu'uttamalla pientä ikkunaa kuvan yli ja luokittelemalla kuvan keskipikseli (Pinheiro & Collobert, 2014). Tällainen lähestymistapa on kuitenkin suoritusajaltaan hidas, koska luokittelijaa täytyy käyttää erikseen pikseleiden määrän verran (Pinheiro & Collobert, 2014; Shelhamer ym., 2017). Tapausositus voidaan puolestaan suorittaa toteuttamalla kohteentunnistus ja luokkaosittamalla tunnistetut kohteita rajaavat suorakaiteet (He ym., 2020).

Huomautettakoon, että nämä eivät kuitenkaan ole ainoita tai välttämättä parhaiten toimivia neuroverkkoarkkitehtuureja näihin tehtäviin (Tian ym., 2021). Muille kuin syväoppiville menetelmille (tässä tutkielmassa niistä käytetään tästä lähtien nimitystä klassiset menetelmät) tämä helppousjärjestys ei myöskään päde, koska niiden avulla ei yleensä ole helpompaa tuottaa luokittelua kuin luokkaositusta. Näin ollen useimmat klassiset karttaprosessointimenetelmät pyrkivät tuottamaan suoraan luokkaosituksen, joka voidaan muuttaa tarvittaessa tapausositukseksi jälkikäsittelemällä (Chiang ym., 2014; Liu ym., 2019).

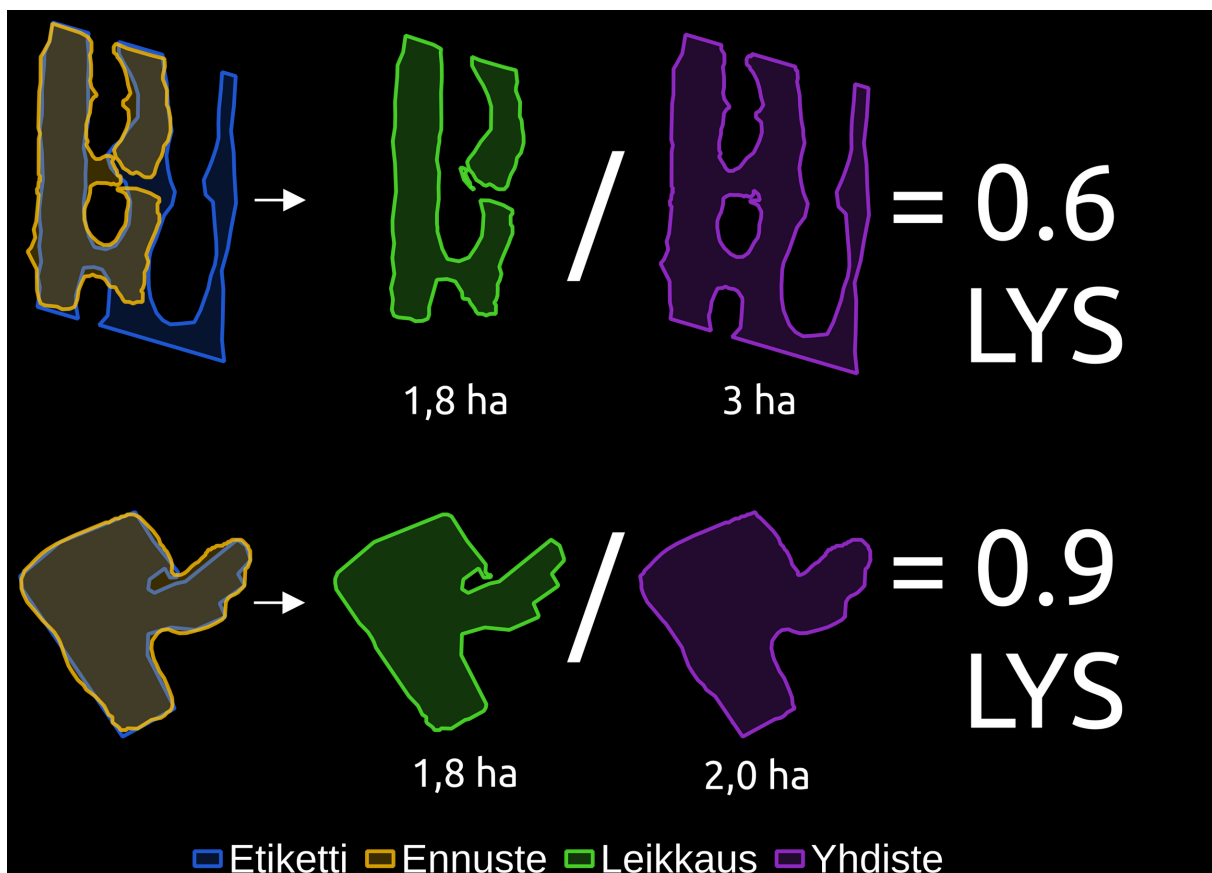
2.3 Luokkaosituksen laatumittarit

Yleisiä luokkaosituksen laatumittareita ovat **sekaannusmatriisi** (eng. Confusion matrix), **Cohenin kapp**a (eng. Cohen's kappa), **tuottajan tarkkuus** (eng. producer's accuracy, recall) **käyttäjän tarkkuus** (eng. user's accuracy, precision) sekä **leikkauksen ja yhdisteen suhde** (**LYS**, eng. Intersection over union, IoU). Kuvantunnistuksessa ja maankäyttöluokittelussa käytetään yleisesti myös muita mittareita, joita ei kuitenkaan esitellä tässä.

Sekaannusmatriisi on yleinen laatumittari maankäyttöluokitusten arvioinnissa, mutta sitä käytetään jonkin verran myös kuvanosituksessa (Maxwell ym., 2021; Tariq ym., 2021). Siinä osituksen tulosta verrataan ristiintaulukoimalla ennustetta ja laadultaan hyvää referenssiaineistoa samoissa kohdissa (Esim. liite 1). Sekaannusmatriisista on mahdollista laskea tuottajan ja käyttäjän tarkkuudet jokaiselle luokalle, sekä kokonaistarkkuus koko aineistolle. Tuottajan tarkkuus kertoo millä todennäköisyydellä referenssiaineistossa tiettyä luokkaa oleva kohta luokituu ennusteessa oikeaksi luokaksi. Tuottajan tarkkuus saadaan jakamalla sekä ennusteessa, että referenssiaineistossa tiettyä luokkaa oleva ala kyseisen luokan pinta-alalla referenssiaineistossa. Käyttäjän tarkkuus puolestaan kertoo, millä todennäköisyydellä ennusteessa jotain luokkaa oleva kohta on samaa luokkaa myös referenssiaineistossa. Se lasketaan jakamalla sekä ennusteessa, että referenssiaineistossa tiettyä luokkaa oleva ala kyseisen luokan pinta-alalla ennusteessa. Aineiston kokonaistarkkuus kertoo, kuinka suuri osa ennusteesta vastaa referenssiaineistoa luokaltaan. Kokonaistarkkuus ei ole kuitenkaan hyvä mittari yksinään. Jos luokkien pinta-alat ovat merkittävästi erisuuruisia, ennustamalla pelkästään enemmistöluokkaa kaikkialle voi saavuttaa korkean tarkkuuden. Yksi ratkaisu tähän ongelmaan on Cohenin kapp (Cohen, 1960). Se on mittari sille, kuinka paljon parempi luokittelu on kuin satunnaisesti luokkajakaumien mukaan tuotettu arvaus. Nolla tarkoittaa ettei tulos eroa merkittävästi sattumasta, lähellä yhtä olevat arvot taas osoittavat luokittelijan olevan paljon sattumaa

parempi. Cohenin kappan tulkinta ei kuitenkaan ole yksiselitteistä, ja riippuu vahvasti aihepiiristä ja aineistosta (Banerjee ym., 1999).

LYS, joka tunnetaan myös nimellä Jaccardin indeksi, on yleinen laatumittari kuvaosituksessa (Tariq ym., 2021; Tian ym., 2021). Se lasketaan jakamalla enusteen ja referenssietiketin geometrioiden leikkauksen pinta-ala niiden yhdisteen pinta-alalla (Kuva 4). LYS huomioi sekaannusmatriisia paremmin sen, että pinta-alaltaan samankokoiset virheet vaikuttavat kohteentunnistuksen tarkkuuteen hyvin eri tavoin erikokoisilla kohteilla. Puolen hehtaarin virhe on merkittävä kohteella, jonka todellinen pinta-ala on hehtaarin, mutta mitätön kohteessa, jonka pinta-ala on sata hehtaaria. Onnistuneen kohteentunnistuksen raja-arvona pidetään yleensä yli 0,5 LYS:iä (Tong ym., 2020; Tian ym., 2021). LYSiä laskettaessa on tyypillistä laskea LYS sekä kaikille kohteille, että erikokoisille kohteille. Yksi yleinen kokojaottelu on MS COCO-aineiston käyttämä jako pieniin, eli alle 32^2 pikselin, suuriin eli yli 96^2 pikselin ja keskisuuriin kohteisiin eli edellisten väliin jääviin kokoluokkiin (Lin ym., 2015; Tong ym., 2020).



Kuva 4: Esimerkki LYS:in laskemisesta. Kumpaakin esimerkkiä voitaisiin pitää onnistuneena kohteentunnistuksena, koska LYS on yli 0,5. Ositus 0,9:n LYS:in tarkkuudella on kuitenkin huomattavasti käyttökelpoisempi kuin 0,6:n LYS:in.

2.4 Haasteet ja epävarmuuden lähteet karttaprosessoinnissa

Historiallisten karttojen tunnistamiseen liittyy muutamia erityispiirteitä, joiden yhteisvaikutusten takia perinteiset dokumenttien prosessointityökalut eivät sovellu sellaisenaan karttojen käsittelyyn (Chiang ym., 2014). Ensinnäkin historialliset kartat sisältävät lukuisia erityyppisiä, ja usein päällekkäisiä symboleita, kuten korkeuskäyriä, teitä ja nimistöä (Liu ym., 2019). Tämä johtaa värien sekoittumiseen, symbolien tiiviiseen vierekkäin oloon ja kohteiden peittymiseen tai katkeamiseen. Viiva- ja aluesymbolit esiintyvät hyvin vaihtelevissa ja epäsäännöllisissä muodoissa, joten tarkasti määritettyjen muotojen tunnistamisesta ei ole apua. Erityisen haastavia kohteita ovat pienet, vain muutaman pikselin kokoiset kohteet ja tekstuuriltaan vaihtelevat kuviot (Uhl ym., 2020). Samoin nimistöt voivat myös olla suuntautuneet monin tavoin tai olla jopa mutkalla, toisin kuin normaaleissa teksteissä. Tämä tarkoittaa että suurinta osaa valmiista kirjaimentunnistusohjelmista ei ole pystytty käyttämään suoraan karttalehtiin (Li ym., 2018). Yhtenä suurimmista karttaprosessoinnin haasteista on se, että usein lukuisat eri symbolit jakavat samat värit, mikä tekee useiden symboleiden erottamisesta pelkästään värin perusteella mahdotonta (Iosifescu ym., 2016; Uhl ym., 2020). Usein vanhojen karttojen symbolit ovat käsinpiirrettyjä, joten samojenkin symboleiden tarkassa ulkonäössä on vaihtelua.

Lisäksi monet fyysisen maailman prosessit heikentävät historiallisten karttojen digitoinnin laatua. Vanhat kartat kärsivät ikääntymisen aiheuttamista ongelmista, kuten haalistumisesta tai värien muuttumisesta (Ostafin ym., 2017; Liu ym., 2019). Huolimaton käsittely on voinut johtaa taitoksiin ja rypistymiseen (Liu ym., 2019). Eri-ikäiset kartat on myös voitu painaa eri värisävyillä ja tekniikoilla. Myös skanneri ja mahdollinen kuvien häviöllinen tiivistäminen aiheuttavat värihäiriöitä tai kohteiden sumentumista (Khotanzad & Zink, 2003).

On hyvä muistaa, että karttaprosessointi itsessään ei ole ainoa epävarmuuksien lähde vanhojen karttojen hyödyntämisessä. Yksi teoreettinen lähestymistapa määrittää epävarmuuden lähteet aineiston käsittelyn lisäksi vielä kahteen muuhun luokkaan: kartan tuotantoon ja aineiston sopivuuteen käyttökohteeseen (Leyk ym., 2005).

Tuotantoon liittyvät epävarmuudet ovat moninaisia. Ensinnäkin sijaintitarkkuuden mittaaminen tapahtuu rajallisella, käytetyn tekniikan määräämällä tarkkuudella, kuten kaikki mittaussprosessit. Toisekseen geodesia ei ole ollut historiallisina aikoina yhtä tarkkaa kuin nykyään. Merkittäväksi ongelmaksi sijaintitarkkuus voi muodostua etenkin huomattavasti ennen 1900-luvun alkua tehtyjen karttojen kanssa, jolloin voidaan puhua monenkymmenen

metrin keskimääräisistä sijaintivirheistä (Vuorela ym., 2002; Pavelková ym., 2016; Ostafin ym., 2017). 1900-luvun kartoituksissa sijaintivirheet ovat usein pienempiä, keskimäärin kymmenkunta metriä tai alle (Ginzler ym., 2011; Mäyrä ym., 2023). Sijaintivirheet eivät myöskään ole samanlaisia kaikkialla. Esimerkiksi niiden on havaittu lisääntyvän maankäyttöluokkien rajoilla, vuoristossa ja heikommin saavutettavilla tai näkyvillä alueilla (Gimmi ym., 2016; Frajer & Fiedor, 2021). Toisaalta nämä havaitut epävarmuudet ovat usein riippuvaisia tutkitusta kartastosta ja epävarmuuksien havaitsemiseen käytetystä metodologiasta. Esimerkiksi Habsburgien valtakunnan toisessa sotilaskartoituksessa kohteiden kulkukelpoisuuden vaikutus kartoituspisteiden jakaumaan havaittiin vähäiseksi ja näkyvyydenkin vaikutus kartoituksen tarkkuuteen oli suhteellisen pieni (Ostafin ym., 2021).

Sijaintivirheitä suurempi ja haastavampi ongelma syntyy karttojen todellisuutta yksinkertaistavasta luonteesta. Osa havaituista vanhojen karttojen maankäyttöluokkien rajojen ”sijaintivirheistä” on todennäköisesti peräisin määritelmällisistä eroista (Leyk ym., 2005). Historiallisissa ja moderneissa kartoissa on erittäin yleistä, että maasto on jaoteltu kartalla selvärajaisiin luokkiin, kuten peltoihin, niittyihin, metsiin ja soihin. Nämä luokat esitetään kauttaaltaan samalla aluesymbolilla, ikään kuin ne olisivat selvärajaisia, täysin yhtenäisiä ja koostumukseltaan samankaltaisia. Todellisuus on kuitenkin huomattavasti monimutkaisempaa. Usein maanpeite vaihettuu toisentyyppiseksi asteittain, eikä selkein hyppäyksiin. Esimerkiksi suon ja metsän välissä ei usein ole selkeää rajaa, vaan niiden välissä on eriasteisesti soistunutta metsää (Kaakinen ym., 2018). Tämän tyyppinen raja voitaisiin esittää konseptuaalisesti paremmin sumeana joukkona, jossa yksittäinen kohta voi kuulua osittain useampaan luokkaan (Halás ym., 2019; Jana & Mahanta, 2023). Historiallisissa topografisissa kartoissa sumeita joukkoja hyvin kuvaavat liukuvärjäysratkaisut ovat kuitenkin hyvin harvinaisia, todennäköisesti ainakin niiden työläyden ja painattamisen vaikeuden takia. Muun tyyppisiä ratkaisuita epävarmuuksien esittämiseen on tosin keksitty. Esimerkiksi Siegfried-kartoissa epäselvät metsänrajat on merkattu jättämällä metsien aluesymbolin reunaviiva piirtämättä (Gimmi ym., 2016). Historiallisissa Peruskartoissa taas esimerkiksi epäselvät rantojen ja niittyjen rajat merkataan katkoviivalla (‘Peruskartta 1:20 000.’, 1948).

Yksi kartalla esitetty maankäyttöluokka voi myös sisällyttää itseensä varsin erilaisia kohteita. Otetaan esimerkiksi Suomen vanhojen Peruskarttojen suoluokat. Ne on jaoteltu kulkukelpoisuuden ja puustonkasvun perusteella (‘Peruskartta 1:20 000.’, 1948). Tämä tarkoittaa sitä, että saman puuttoman suoluokan sisällä voi olla niin nevoja, lettoja, luhtia kuin lähteiköitä, jotka ovat toki nekin vain erityyppisiä yksinkertaistuksia. Yksinkertaistukset ovat

kuitenkin välttämättömiä, koska kaikkia todellisuuden aspekteja ei ole mahdollista saada kuvattua kartalle kerralla (Sester, 2020). Käytännön rajoitteiden lisäksi arvovalinnat siitä, mitkä kartoitettavat ominaisuudet ovat merkittäviä, aiheuttavat siten määritelmiin liittyviä epävarmuuksia historiallisiin karttoihin (Leyk ym., 2005). Huomionarvoista on, että tämäntyyppisten epävarmuuksien arviointi joko puuttuu tai on hyvin pintapuolista monissa historiallisia karttoja käyttävissä tutkimuksissa. Eri kartastoissa käytetyt maankäyttöluokkien määritelmät saattavat erota toisistaan merkittävästi, vaikka kyse olisi samoista alueista (Kaim ym., 2014). Samassa tutkimuksessa myös esitettiin kartastojen välisten laatuerojen mittaamiseen kehityssuuntien todenmukaisuuden arviointia. Eli esimerkiksi metsän kehityskulkua pidettäisiin epätodennäköisenä, jos alue vaihtuisi muutaman kymmenen vuoden välein tehdyissä kartoissa metsästä metsättömäksi ja takaisin tai toisinpäin. Tämäntyyppinen analyysi vaatii kuitenkin hyvin tarkkaa ymmärrystä eri maankäyttöluokkien muutosten aikajänteistä ja syklisyyksistä.

Karttaprosessoidun aineiston käyttökelpoisuus riippuu käyttökohteesta. Maankäytön muutoksia tutkittaessa modernien aineistojen ja vanhojen karttojen vertailukelpoisuutta voidaan arvioida vertailemalla maankäyttöluokkien määritelmien ja laatutietojen eroja (Leyk ym., 2005; Kaim ym., 2014). Vain riittävän samankaltaisia aineistoja tulisi verrata keskenään. ”Riittävä” sisältää tosin runsaasti subjektiivisuutta ja tutkijan omaa harkintaa, joskus myös modernien aineistojen muokkaamista vastaamaan historiallista määritelmää mahdollisimman tarkasti (Leyk ym., 2005). On myös otettava huomioon, että mahdolliset epämääräisyydet määritelmässä ja eroavuudet koulutuksessa ovat todennäköisesti johtaneet eriäviin kartoitustulkintoihin jo alkuperäisillä kartoittajilla (Frajer & Fiedor, 2021). Lisäksi on hyvä tiedostaa, että joidenkin vanhojen karttojen metadata voi olla puutteellista, jolloin tutkijan voi olla tarpeen tehdä tai löytää historiallinen katsaus eri maankäyttöluokkien historiallisesta ymmärtämisestä ja sääntelystä (Leyk ym., 2005).

2.5 Karttaprosessoinnin menetelmät

2.5.1 Georeferointi

Ennen kuin skannattua karttaa on mahdollista hyödyntää, täytyy kartta georeferoida eli kiinnittää todelliseen sijaintiin. Tämä tapahtuu kiintopisteiden avulla. Ne ovat kohteita, joiden sijainti on tunnistettavissa sekä kartalta pikseleissä, että tosimaailman koordinaatteina. Tällaisia kohteita voivat olla esimerkiksi teiden risteykset, koordinaattiruudukot tai

rakennukset. On kuitenkin hyvä muistaa, että vanhimmissa kartoissa esiintyvistä kohteista voi olla jäljellä vain vähän varmasti tunnistettavia kohteita, ja kartoitustarkkuus on ollut tuolloin myös heikompi (Vuorela ym., 2002). Koska karttaprosessoinnin tavoitteena on minimoida karttalehteä kohden tarvittu ihmistyön määrä, myös (puoli)automatisoitu georeferointi on ollut viimeaikaisen tutkimuksen kohteena (Dong ym., 2018; Heitzler ym., 2018; Burt ym., 2020; Bahgat & Runfola, 2021). Nämä tutkimukset ovat ehdottaneet erityyppisiä kohteita kiintopisteiksi, ja monimuotoisia menetelmiä niiden tunnistamiseen.

Tunnettuun projektiosysteemin sekä riittävän tarkkaan geodesiaan ja kartoitukseen pohjautuvien karttojen kohdalla voidaan kiintopisteinä käyttää suoraan kartan koordinaattiruudukon risteys- tai kulmapisteiden koordinaatteja (Dong ym., 2018; Heitzler ym., 2018; Burt ym., 2020). Etenkin suorakaiteen muotoiset ja säännöllisen ruudukon sisältävät kartat tarjoavat vahvat mahdollisuudet pitkälle viedylle automaatiolle (Heitzler ym., 2018; Burt ym., 2020). Myös muiden kuin suorakaiteisten karttojen koordinaatistoon perustuvassa georeferoinnissa voidaan vähentää ihmistyön määrää, mutta täysi automatisointi ei ole aivan vielä mahdollista nykyisillä menetelmillä (Burt ym., 2020). Suorakaiteisten karttojen georeferointi koostuu useista vaiheista, joista keskeisimmät ovat koordinaattiruudukon tunnistaminen kuvasta, koordinaattien määrittäminen tunnistetuille risteyksille, validointi ja lopullinen georeferointi (Heitzler ym., 2018; Burt ym., 2020).

Ruudukon määrittäminen kuvasta on mahdollista toteuttaa hyvällä tarkkuudella etsimällä ruudukon risteyspisteiden malliin verrattuna samankaltaisia kohteita (Heitzler ym., 2018; Burt ym., 2020). Apuna voidaan käyttää myös oletusta, että kartan kulmissa on vaaleaa taustaa kolmella neljänneksellä, joten ne on helpompi tunnistaa kuin keskellä karttaa olevat koordinaattiruudukon risteykset (Burt ym., 2020). Jos myös koordinaattiviivojen tiheys ja yksittäisen pikselin koko metreissä tunnetaan, voidaan myös muiden risteyspisteiden likimääräinen sijainti arvioida, mikä vähentää valekohteiden tunnistuksen riskiä ja nopeuttaa etsintää.

Koordinaattiruudukon kulmissa sijaitsevien kiintopisteiden koordinaattien selvittämiseen on kaksi vartenotettavaa erilaista lähestymistapaa: tiedon lukeminen kartalta tai koordinaattien määrittäminen tunnetun tai takaisinmallinnetun karttalehtijaon perusteella. Heitzler ja kumppanit (2018) opettivat yhden konvoluutioneuroverkot tunnistamaan koordinaatteja ja toisen symboleita koordinaattiehdoista. Neuroverkot saavuttivat kummassakin tapauksessa yli 99% tarkkuuden, ja tunnistetuille koordinaateille tehtiin myös jälkitarkastus, jossa

tarkistettiin niiden realistisuus. Burt ja kumppanit (2020) sen sijaan oletivat, että georeferoinnin käyttäjä pystyisi tuottamaan tai löytämään järkevällä vaivalla karttalehtijakoruudukon niiden säännöllisyyden ansiosta. Näin karttalehtien kulmakoordinaatit saataisiin suoraan. Seuraavaksi koordinaattiruudukon risteyspisteiden koordinaatit kartan sisällä voidaan laskea kulmapisteiden koordinaattien perusteella, koska ne ovat säännöllisin, oletetusti tunnetuin välimatkoin (Heitzler ym., 2018; Burt ym., 2020). Huomautettakoon, että pääasiassa säännöllisiä, suorakulmaisia karttoja sisältävissä karttasarjoissa voi myös olla kartan tai projektiokaistojen raja-alueilla epäsäännöllisen muotoisia tai jopa kahta projektiota yhdisteleviä karttalehtiä (esim. Peruskarttalehdet 1034 12+2012 14 Dragsfjärd, 1968 ja 4112 12+4112 11+4114 03 Rautjärvi, 1971). Tällaiset poikkeustapaukset voivat estää karttalehtijaon täysin automaattisen takaisinmallinnuksen. Lisäksi ne voivat vaatia muokkauksia eri kohtiin automaattisen georeferoinnin prosessia.

Paikannimet pysyvät usein muuttumattomina yllättävän pitkiä aikoja, vaikkakin poliittiset muutokset voivat muuttaa merkittävien sijaintien nimiä (Rajić, 2012; Seidl, 2019). Pienimittakaavaisissa kartoissa niihin liittyy usein pistemäinen symboli ja joihinkin kartastoihin liittyy myös paikannimihakemistoja, joissa nimien sijainti on määritelty tarkemmin. Näiden seikkojen takia myös paikannimiä on käytetty onnistuneesti pienimittakaavaisten karttojen kiintopisteinä (Bahgat & Runfola, 2021). Bahgat ja Runfola (2021) hyödynsivät avoimen lähdekoodin tekstintunnistusohjelma Tesseractia nimien tunnistamiseen (Tesseract documentation, 2023). Tämän jälkeen heidän menetelmänsä etsii nimien läheltä tummia pistemäisiä kohteita, jotka merkitsevät nimen tarkoittamaa sijaintia. Paikannimen sijainti selvitettiin paikannimihakemistosta. Kartoilla on usein myös samannimisiä kohteita eri paikoissa, mikä on tosin ratkaistavissa tarkastelemalla useampaa paikannimeä samalta alueelta (Bahgat & Runfola, 2021). Nimien yleisempään käyttöön georeferoinnissa liittyy kuitenkin suuria ongelmia. Kaikki karttasarjat eivät merkitse nimen tarkkaa sijaintia millään symbolilla ja nimi voi viitata pistemäisen kohteen sijasta alueeseen etenkin suurimittakaavaisilla kartoilla. Samalla paikalla voi myös olla useita nimiä eri kielillä, eivätkä kaikki kartoille merkityt nimet välttämättä löydy paikannimihakemistosta (Seidl, 2019).

Validointi on tärkeää, koska parhaatkään nykyiset automaattisen georeferoinnin menetelmät eivät ole virheettömiä ja siten käyttäjän tarvitsee korjata virheitä. Muutoin virheet välittyvät eteenpäin kaikkeen kartalla tapahtuvaan prosessointiin. Yhdeksi ratkaisuksi on tarjottu ehdotetun georeferoinnin avaamista paikkatieto-ohjelmaan tunnistettujen koordinaattiarvojen

ja kiintopisteiden kanssa (Heitzler ym., 2018). Jos kaikki vaikuttaa olevan kunnossa, voi käyttäjä vain sulkea ikkunan, muutoin käyttäjän pitää itse tehdä korjaukset. Myös validoinnin automatisointia on kehitetty. Jos kiintopisteitä on riittävästi, voidaan yksittäiset virheet huomata laskemalla koordinaattimuunnos siten, että yksi kiintopiste kerrallaan on pois tarkastelusta (Burt ym., 2020; Bahgat & Runfola, 2021). Tällöin on mahdollista mitata etäisyys kiintopisteen koordinaateista georeferoinnin tulokseen. Liian suuret etäisyydet merkataan virheiksi, jolloin kyseinen kiintopiste voidaan poistaa (Bahgat & Runfola, 2021). Vaihtoehtoisesti lista epäillyistä virheistä palautetaan käyttäjälle tarkastettavaksi (Burt ym., 2020).

2.5.2 Skannattujen karttojen luokkaositus

Karttojen digitoinnin automatisointiin on kehitetty menetelmiä jo yli 40 vuoden ajan, eli melkein yhtä kauan kuin digitaalista paikkatietoa on ollut (Boyle, 1980; Chiang ym., 2014). Ensimmäiset karttaprosessointijärjestelmät olivat hyvin kalliita, hitaita sekä toiminnallisuudeltaan rajattuja lähinnä viivantunnistukseen, minkä lisäksi ne vaativat kartan fyysisistä esikäsittelyä ja ensiluokkaista kartan laatua sekä runsasta manuaalista jälkikäsittelyä (Boyle, 1980). Klassiset karttaprosessointimenetelmät ovat moninainen joukko menetelmiä, joiden tarkkuus ja työläys vaihtelevat suuresti, myös karttasarjasta riippuen (Chiang ym., 2014; Liu ym., 2019). Klassisilla menetelmillä on pystytty tunnistamaan kohtalaisella tai hyvällä tarkkuudella (85–97 %) muusta kartasta väriltään erottuvia kohteita kuten vesistöjä, rakennuksia, metsiä ja korkeuskäyriä (Iosifescu ym., 2016; Liu, Miao, Xu, ym., 2016; Ostafin ym., 2017). Yksi löytämäni tutkimus oli myös onnistunut tunnistamaan mustalla symbolikuvioinnilla merkattuja metsäalueita 94 % tarkkuudella, vaikka kartalla oli myös muita samenvärisiä muita kohteita (Leyk ym., 2006). Myös tekstin tunnistukseen on tehty yrityksiä, mutta tulokset eivät ole klassisilla menetelmillä olleet erityisen tarkkoja (Mello ym., 2012).

Monet klassiset menetelmät vaativat usein kuvien esikäsittelyä ennen varsinaista luokittelua. Tavanomaisia esikäsittelytapoja ovat kontrastin lisääminen, taustan poisto, värialueiden yhtenäistäminen tai reunojen tunnistus (Mello ym., 2012; Liu, Miao, Tian, ym., 2016; Ostafin ym., 2017). Myös värien esittämistapojen muutoksia tavanomaisesta **puna-viher-sini** (eng. Red-green-blue, RGB) esitystavasta muihin esitystapoihin on hyödynnetty (Herrault ym., 2013; Ostafin ym., 2017).

Varsinaisista luokkaositusmenetelmistä suurin osa hyödyntää jossain määrin värien erottelua (Khotanzad & Zink, 2003; Herrault ym., 2013; Chiang ym., 2014; Iosifescu ym., 2016; Ostafin ym., 2017; Liu ym., 2019; Bahgat & Runfola, 2021) Yleisin tapa värierotteluun on eriyttää kartan värijakaumasta arvovälejä, jolle suurin osa tietyn luokan väreistä päätyy (Dhar & Chanda, 2006; Mello ym., 2012; Herrault ym., 2013; Iosifescu ym., 2016; Ostafin ym., 2017). Toinen tavanomainen lähestymistapa värierotteluun on etsiä väreiltään samankaltaisia ja suhteellisen homogeenisiä alueita, ns. superpikseleitä, ja yhdistellä niitä (Liu, Miao, Tian, ym., 2016). Monien vanhojen digitoitujen karttojen värilaatu ei kuitenkaan ole riittävä virheettömän erottelun tekemiseen vain värityksen perusteella (Liu, Miao, Xu, ym., 2016; Ostafin ym., 2017). Myös skannausprosessit tuottavat värihäiriöitä (Leyk ym., 2005). Näin ollen alustavaa luokkaositusta täytyy jälkikäsitellä.

Syväoppivat (konvoluutio)neuroverkot ovat yleistyneet merkittävästi viimeisimpien vuosien aikana karttaprosessoinnissa (Uhl ym., 2017, 2020; Heitzler ym., 2018, 2018; Duan ym., 2020; Maxwell ym., 2020; Saeedimoghaddam & Stepinski, 2020; Garcia-Molsosa ym., 2021; Guo ym., 2021; Ståhl & Weimann, 2022; Wu ym., 2022; Zhang ym., 2022; Petitpierre & Guhenec, 2023; Vaienti ym., 2023). Vielä vuonna 2019 julkaistussa katsausartikkelissa yksikään tarkasteltu artikkeli ei käyttänyt näitä menetelmiä, vaan sen sijaan niihin viitattiin potentiaalisena tulevaisuuden menetelmänä (Liu ym., 2019). Ensimmäinen löytämäni konferenssiartikkeli aiheesta on sekin vasta vuodelta 2017 (Uhl ym., 2017). Pienien neuroverkkojen käyttö ei itse asiassa ole erityisen uutta karttaprosessoinnissa, sillä niitä on esiintynyt jo 1990-luvulla (Wu ym., 1994; Reiher ym., 1996).

U-net ja siitä jatkokehitellyt menetelmät ovat olleet viime vuosina suosittuja karttojen luokkaosituksessa (Heitzler & Hurni, 2020; Maxwell ym., 2020; Garcia-Molsosa ym., 2021; Petitpierre, 2021; Wu ym., 2022; Mäyrä ym., 2023; Vaienti ym., 2023). Näissä tutkimuksissa ositettiin mm. arkeologisesti merkittäviä paikannimiä ja raunioita, rakennuksia, vesistöjä, peltoja, soita, teitä, entisiä kaivoksia ja urbaaneja maankäyttöluokkia. Muokkaamattomalla U-netillä mallien keskimääräinen tarkkuus oli noin 65–95 % luokasta, karttasarjasta ja koulutuksen hyperparametreista riippuen. Luokkaositukseen on myös hyödynnetty syväoppivaa luokittelijaa, joka kertoo, mitä luokkaa pienen kuvaikkunan keskipiste edustaa (Uhl ym., 2020). Liu'uttamalla tällaista luokittelijaa kartan yli on mahdollista saada koko karttalehti prosessoitua alueellisesti kattavasti.

Neuroverkkojen yhtenä yleisenä haasteena on se, että ne saattavat olla näennäisesti luottavia ennustukseensa, mutta ovat silti väärässä (Abdar ym., 2021). Tätä ratkomaan on kehitetty lukuisia menetelmiä, jotka lisäävät tai parantavat neuroverkkojen epävarmuusarvioita (Abdar ym., 2021). Monet näistä menetelmistä ovat melko tuoreita, minkä takia löysin vain yhden karttaprosessointimenetelmän, jossa niitä oli hyödynnetty. Wu ja kumppanit (2022) luokkaosittivat erityyppisiä vesistökohteita U-netillä, johon oli lisätty **väljä tilallinen pyramidikoontikerros** (eng. Atrous spatial pyramid pooling, ASPP; Chen ym., 2018) U-netin keskelle. Epävarmuutta arvioitiin lisäämällä malliin **epävarmuutta arvioimaan oppiva ulostulo** (eng. Learned confidence estimate, LCE) ja tappiofunktio, joka rankaisee epävarmemmista väärin menneistä ennustuksista vähemmän kuin varmemmista (DeVries & Taylor, 2018; Wu ym., 2022). Käytetty tappiofunktio pohjautui Sørensen-Dicen kertoimeen, koska se toimii tehokkaasti kuvilla, joista suuri osa on taustaa (Milletari ym., 2016). Jottei malli ennusta kaikkialle suurta epävarmuutta, tappiofunktioista vähennetään rankaisuterminä epävarmuuden logaritmi (DeVries & Taylor, 2018). Näillä muutoksilla mallin tuottajan tarkkuus nousi 20–30 prosenttiyksikköä verrattuna muokkaamattomaan U-nettiin (Wu ym., 2022). Ennustetut epävarmuudet olivat suurimpia symbolien reunoilla ja alueilla, joilla kuvassa oli erityyppisiä häiriöitä.

Pistemäisten karttasymbolien tunnistukseen sopivat kohteentunnistuksessa käytettävät neuroverkot. Zhang ja kumppanit (2022) vertailivat eri kohteentunnistusmallien (You Only Look Once version 3 with Convolutional Block Attention Module eli YOLOv3 CBAM-moduulilla; Faster Regional Convolutional Neural Network eli R-CNN ja Single Shot Detector eli SSD) suorituskykyä modernilla kiinalaisella kartalla ja totesivat, että tasalaatuisella kartalla kaikki mallit pystyivät yli 97,5 % kokonaistarkkuuteen, YOLOn saavuttaessa 99,5 %:n kokonaistarkkuuden. Myös moniosaisia geologisia symboleita on onnistuttu tunnistamaan hyvällä tarkkuudella (Guo ym., 2021). Edellämäinittujen tulosten tulkinnassa on otettava huomioon, että nämä molemmat tutkimukset oli tehty moderneilla kartoilla, joiden symbolit ovat käytännössä identtisiä. Saavutus on kuitenkin sinänsä onnistunut, koska pienien kohteiden tunnistaminen on neuroverkoille yleisesti vaikeampaa kuin suurien (Tong ym., 2020; Tian ym., 2021). Kohteentunnistusta on hyödynnetty onnistuneesti myös historiallisilla kartoilla esimerkiksi tienristeysten tunnistamiseen (Saeedimoghaddam & Stepinski, 2020).

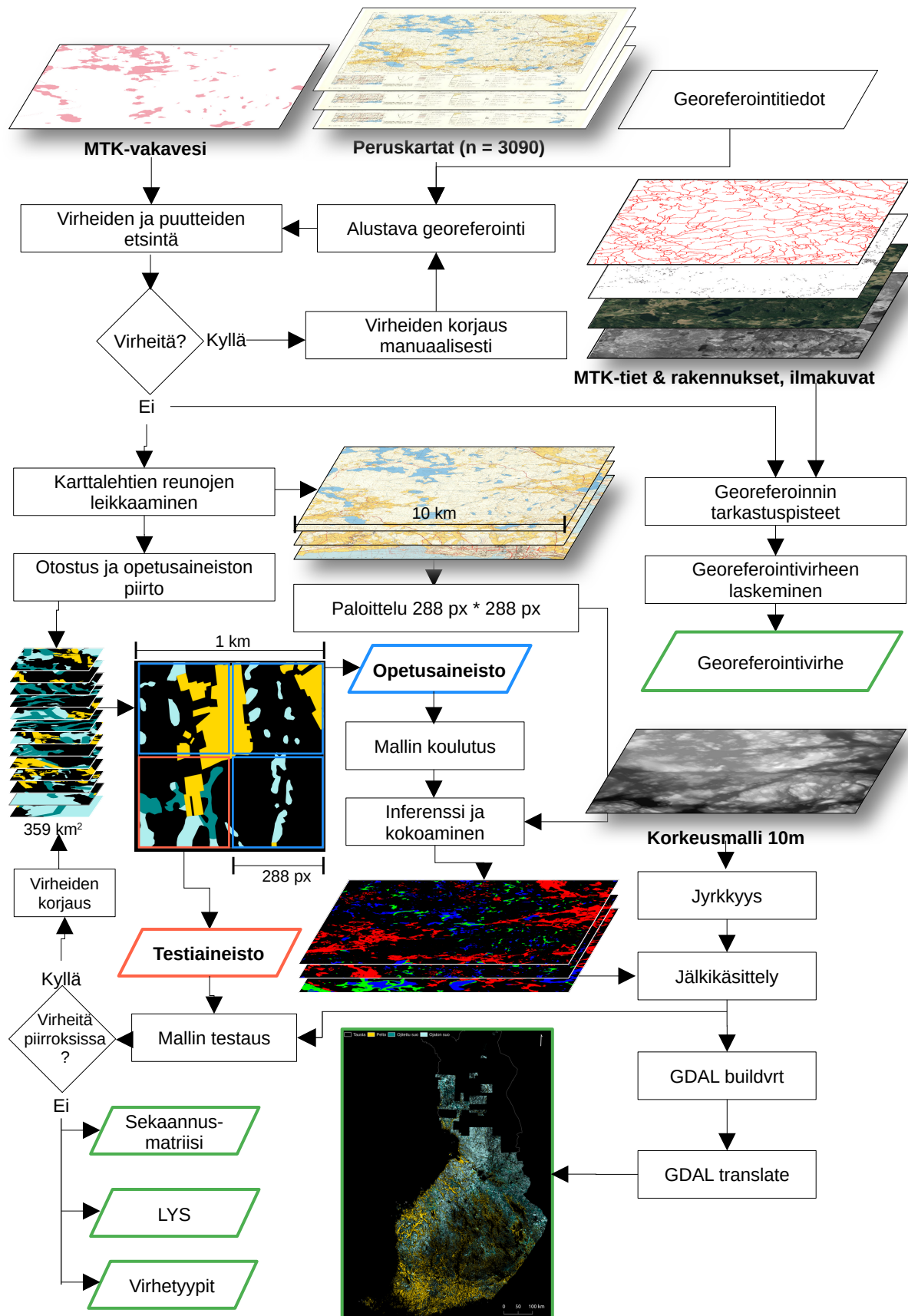
Syväoppivaa kohteentunnistusta voi hyödyntää tekstien lukemiseen vanhoista kartoista tunnistamalla tekstialueet, jotka voidaan sitten kääntää vaakatasoon ja syöttää tavanomaiselle

tekstintunnistusohjelmalle, kuten Tesseractille (Li ym., 2018; Tesseract documentation, 2023). Myös viimeaikainen valtava kehitys **CAPTCHA**-tekstien automaattisessa murtamisessa (Täysin automatisoitu julkinen Turinginin testi ihmisten ja tietokoneiden erottamiseksi; eng. *Completely Automated Public Turing test to tell Computers and Humans Apart*) on mahdollistanut erittäin vääntyneiden ja vääristyneiden tekstien automaattisen lukemisen hyvällä tarkkuudella (Chen ym., 2023).

3 Aineistot ja menetelmät

3.1 Tutkielman metodologinen kokonaisuus

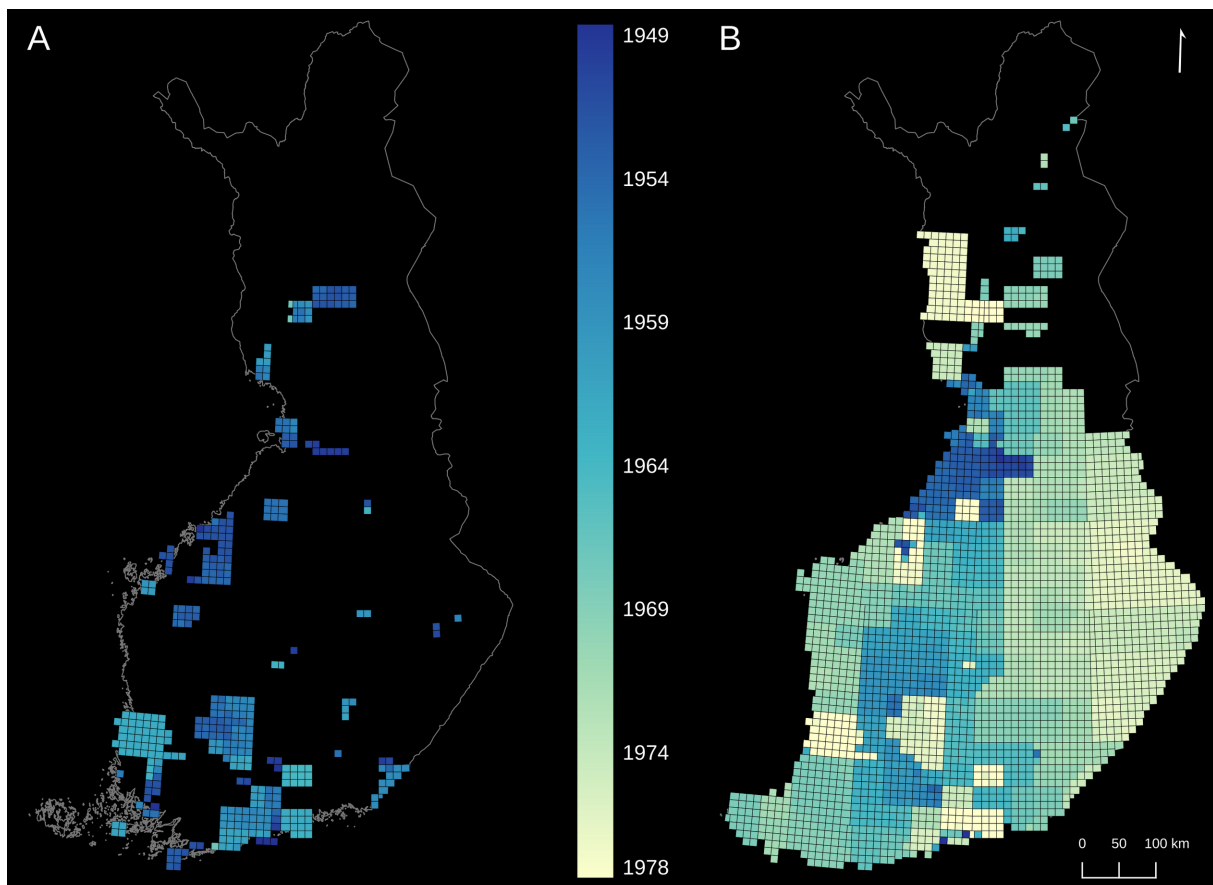
Tutkielman metodologinen kokonaisuus oli monivaiheinen, ja hyödynsi sekä historiallisia skannattuja Peruskarttoja, että uudempia maanmittauslaitoksen aineistoja (Kuva 5). Prosessi alkoi skannattujen Peruskarttojen alustavalla georeferoinnilla, sekä sen virheiden ja puutteiden etsimisellä ja korjaamisella. Tämän jälkeen karttalehdistä leikattiin karttainformaatiota sisältämättömät reunat pois, ja georeferointivirhe mitattiin. Seuraavaksi syväoppimista varten luotiin käsin digitoimalla opetus- ja testiaineisto, jossa skannattujen karttalehtien sisältö jaoteltiin ojitettuihin ja ojittamattomiin soihin, peltoihin sekä taustaan, joka käsitti kaikki muut maankäyttöluokat. Opetusaineistolla koulutettiin syväoppiva kuvantunnistaja, jolla luokkaosoitettiin kaikki karttalehdet. Tämän jälkeen ositus jälkikäsiteltiin, sen laatu tarkastettiin testiaineistolla ja karttalehdet yhdistettiin yhtenäiseksi maankäyttöaineistoksi.



Kuva 5: Tutkimuksen menetelmät yleistetysti. Näkyvillä Peruskarttalehti 2042 01 Karisjärvi, 1958

3.2 Skannatut peruskarttalehdet 1949–1978

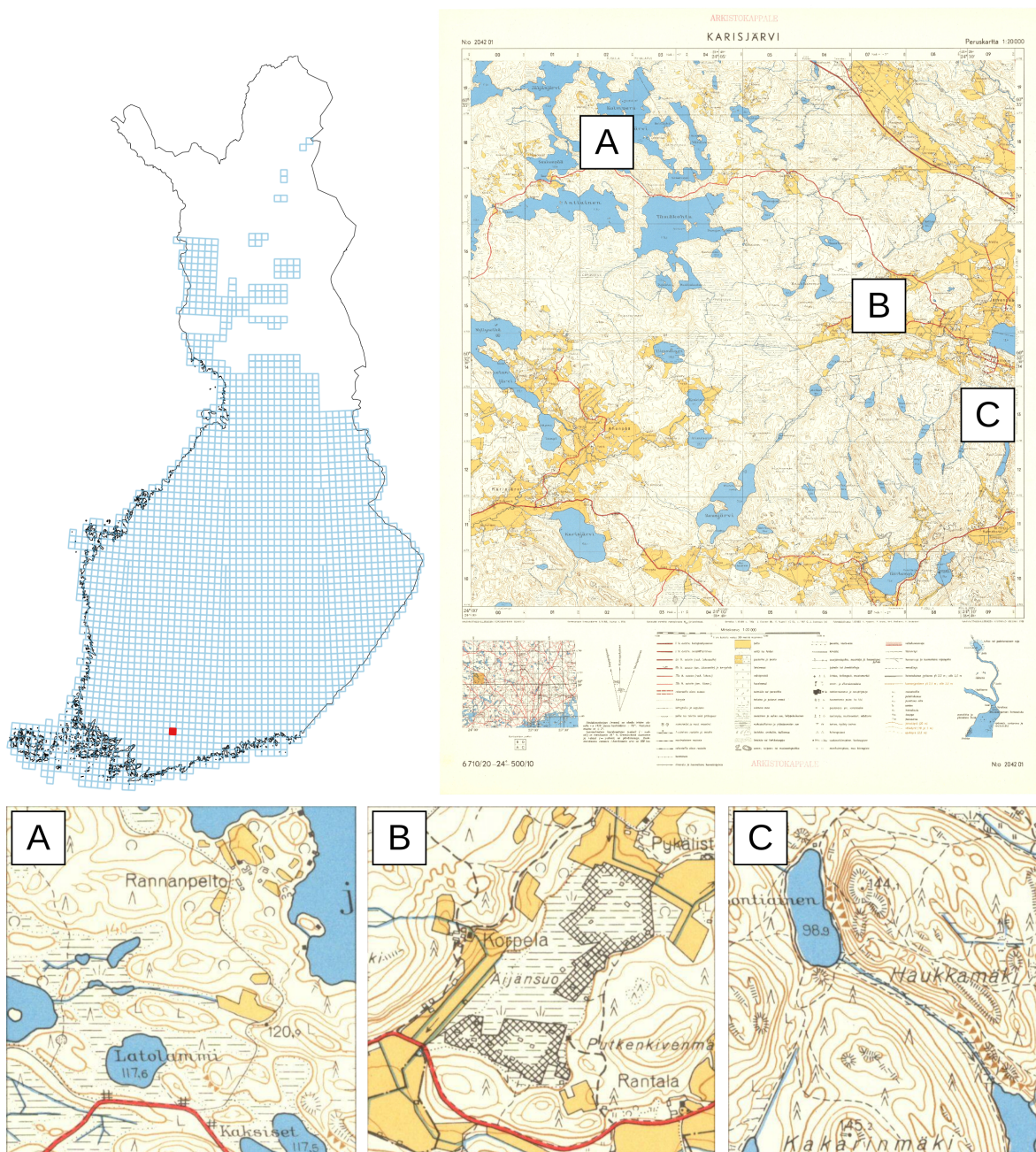
Tutkimuksen pääaineistona ovat vuosien 1949-1978 skannatut peruskarttalehdet (n = 3090) (Kuva 6). Skannausresoluutio on 1,69 m / px. Yleensä peruskartan lehdet ovat 10 km x 10 km neliöitä, mutta siihen on kuitenkin poikkeuksia projektiokaistojen reunoilla sekä raja-alueilla (esim. Peruskarttalehdet 1034 12+2012 14 Dragsfjärd, 1968 ja 4112 12+4112 11+4114 03 Rautjärvi, 1971). Peruskartan pohjana käytettiin ilmakuvia, joiden päälle varsinainen maastokartoitus piirrettiin mittakaavaan 1:10 000, mutta kartta pienennettiin painovaiheessa 1:20 000:een (Hirva ym., 1975). Näin tarvittavien karttalehtien määrä saatiin pienennettyä kohtuullisemmaksi, mutta kartan tarkkuus säilyi korkeana. Historialliset peruskartat ovat ladattavissa skannattuina, mutta georeferoimattomina valokuvina maanmittauslaitoksen vanhojen painettujen karttojen latauspalvelusta ('Vanhat painetut kartat', 2024).



Kuva 6: Tutkimuksessa käytetyt karttalehdet ja niiden valmistumisvuodet A) Tutkimusajanjaksona uudistettujen karttalehtien vanhemmat versiot. B) Tutkimusajanjaksona kertaalleen kartoitetut sekä uudistetut karttalehdet. Valtionrajat tässä ja kaikissa muissa kuvaajissa: Hallintorajat teemakartoille, ei merialueita 2023, 1:1 000 000 (Maanmittauslaitos, 2023).

Ajanjakso 1949–1978 valittiin, koska silloin tehtiin ensimmäinen liki koko Suomen kattava peruskartoitus (Hirva ym., 1975). Itä- ja Pohjois-Lapista valmistettiin tuolloin vain karkeampi ja vähävärisempi Topografinen kartta 1:20 000, joten ne on rajattu tämän tutkimuksen ulkopuolelle. Peruskartat suunniteltiin siten, että niistä olisi hyötyä mahdollisimman monelle toimialalle, sekä sotilas-, että siviilikäyttöön (Hirva ym., 1975). Siksi kartoissa on tietoja niin maastonmuodoista, maankäytöstä, nimistöstä kuin myös hallinto- ja tonttirajoista (Kuva 7). Aiemmissa suomalaisissa karttasarjoissa taloudelliset Pitäjänkartat 1:20 000 ja maastoa kuvaavat Topografiset kartat 1:20 000 ja 1:100 000 olivat olleet erillisiä. Tämän lisäksi Pitäjän- ja Topografisia karttoja oli tehty vain osasta Suomea ja koko suomen kattanut Suomen yleiskartta oli paljon karkeampi, vain 1:400 000 (Hirva ym., 1975). Peruskartoituksesta vastasivat sekä maanmittaushallitus, että puolustusvoimien pääesikunnan topografikunta, joka kartoitti 150 peruskarttalehteä lähinnä varuskuntien läheltä. Ilmakuvaus oli ensimmäisen vuosikymmenen täysin topografikunnan ja ilmavoimien vastuulla ja sen jälkeenkin osittain (Hirva ym., 1975). Työn nopeuttamiseksi 78 karttalehteä pohjanmaalta valmistettiin ilman korkeuskäyriä (Hirva ym., 1975).


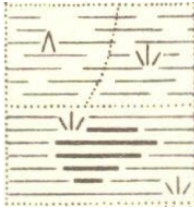
Monista kartoista on kiinteistörajalliset ja -rajattomat versiot, joista valittiin rajallinen aina kun se oli saatavilla. 224 tapauksessa käytettiin kuitenkin kiinteistörajatonta versiota, koska rajallista ei ollut saatavilla samalta vuodelta. Monista kartoista on tehty myös uusintapainoksia, joissa kiinteistörajoja päivitettiin mutta varsinaista uudelleenkartoitusta ei tehty. Myös nämä poistettiin, koska ne eivät sisältäneet uutta maankäyttötietoa. Karttalehdistä 368:lle oli ehditty tekemään tutkimuksen aikarajauksen aikana maastotietojen päivitys. Helsingin ja Tampereen seudun karttalehdille oli ehditty tekemään kolme kartan päivitystä. Koska tämä ala oli suhteellisen pieni, jätettiin päivityksistä keskimäinen pois. Jäljelle jääneet karttalehdet jaettiin kahteen luokkaan: Vanhemmat karttalehdet ($n = 368$, $37\,200\text{ km}^2$), joista löytyi myös päivitetty versio sekä loput karttalehdet, sisältäen päivitettyjen karttalehtien uudemmat versiot ($n = 2722$, $272\,600\text{ km}^2$) (Kuva 6). Tähän jaotteluun päädyttiin, koska päivitettyjen karttalehtien vanhemmat versiot olivat pääosin 1950-luvulta, kun taas kertaalleen kartoitettujen, sekä päivitettyjen karttojen uudemmat versiot olivat valtaosin 1960- ja 70-luvulta (Kuva 6, Hirva ym., 1975). Tutkimusalue määräytyy täysin aineistolähtöisesti.

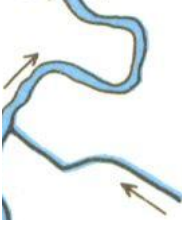


Kuva 7: Karisjärven vuoden 1958 peruskarttalehti ja sijainti peruskartan vuoteen 1978 mennessä valmistuneiden karttojen ruutujaossa. Lisäksi suurennettuna on kolme neliökilometrin kokoista esimerkkiä, joiden sijainti on merkitty karttalehdelle. Karttalehti: (Kosonen ym., 1958).

Peruskarttojen alkuperäisiä kuvausohjeita säilytetään kansallisarkistossa, ja ne kuuluvat maanmittaushallituksen II arkistoon (‘Peruskartta 1:20 000.’, 1948; ‘Peruskartta 1:10 000 /1:20 000.’, 1960). Taulukossa 1 esitellään tämän tutkielman kannalta oleelliset peltojen, soiden ja purojen sekä ojien kuvausohjeet.

Taulukko 1: Tutkielmassa hyödynnettävien peruskartan luokkien, eli peltojen, soiden ja purojen sekä ojien alkuperäiset kuvausohjeet.

Luokka	Kuvaus (suoria lainauksia peruskartan alkuperäisistä kuvausohjeista)	Symboli	Käyttötarkoitus
Pelлот	<p>“ Pelto on viljeltyä kivennäis- tai suomaata, jota säännöllisesti uudistuvasti muokataan ja jossa viljellään pääasiassa peltokasveja. Uudisviljely katsotaan kartoituksen yhteydessä pelloksi silloin, kun siinä pintaraivauksen lisäksi on suoritettu tai suoritettavana ojitustyö. Siihen asti se katsotaan entiseen maankäyttölajiinsa kuuluvaksi.</p> <p>Jos pelto on jäänyt tai jätetty viljelemättä, se on kuitenkin katsottava pelloksi niin kauan kuin se voidaan varsinaiseen uudisviljelykseen ryhtymättä muokata peltokasveille. Esim. vanhat heinäpellot, jotka etäisen sijaintinsa vuoksi ovat olleet pitkät ajat muokkaamatta ja muistuttavat luonnonniittyjä, on katsottava pelloiksi, elleivät kuivatus- tai muokkausvaikeudet estä käyttämästä aluetta peltona, jolloin se luetaan käyttönsä mukaan niityksi tai laitumeksi.</p> <p><u>Suoviljelykset</u> luetaan kuuluviksi peltoihin.</p> <p>Pellot piirretään kantapiirroksen ilman pintamerkintää ja ehjällä ääri viivalla. Sarkaojia ei kuvata, mutta kylläkin viemärit ja valtaojat”</p> <p>(‘Peruskartta 1:20 000.’, 1948)</p>		Ositetaan sellaisenaan
Suot	<p>”Vedenvaihaamista maista kuvataan peruskartalle soistuvat maat ja varsinaiset suot.</p> <p>Ensiksimmäiseen ryhmään luetaan kuuluviksi lehtokorvet, kangaskorvet ja kangasrämeet. Niissä turvekerroksen vahvuus on yleensä alle 0,3 metriä ja niissä kasvaa pääasiallisesti lehti- ja sekämetsää.” ...</p> <p>”Varsinaisista soista otetaan peruskartalle rämeet ja korvet sekä nevat ym. aukeat suot. Soiden</p>		Sualueet ositetaan, mutta ne jaetaan ojitettuihin ja ojitamattomiin soihin

	<p>kuvauksessa käytetään turvamerkkiä, vesiviivoitusta ja metsämerkkejä siten, että <u>rämeet</u>, jotka ovat mäntymetsää kasvavia rahkasoita, esitetään mäntymetsän merkeillä yhdessä turvamerkkien ja vesiviivoituksen kanssa, <u>korvet</u> kuusi- ja lehtipuun merkeillä yhdessä turvamerkkien ja vesiviivoituksen kanssa ja <u>nevat ym. muut aukeat suot</u> ilman metsämerkkejä yksinomaan turvamerkeillä ja vesiviivoituksella. Jos kuvattava suo on pieni, piirretään siihen vesiviivojen lisäksi vain toinen pintamerkeistä, nimittäin metsämerkki silloin, kun suo on metsää kasvava ja turvamerkki silloin, kun suo on aukea. Samoin menetellään isojen soiden kapeissa osissa, pyrkimyksenä saada nevat erottumaan räme- ja korpialueista.</p> <p>Kulkukelpoisuuteensa nähden suot jaetaan helppokulkuisiin ja vaikeakulkuisiin tai ylitsepääsemättömiin soihin. Edelliset kuvataan niissä siten, että vesiviivat piirretään yhtenäisiksi merkkien selityksen osoittamalla tavalla. Milloin jännemuodostumat helpottavat ylikulkua, ne osoitetaan katkaisemalla vesiviivat jänteiden suunnassa sopivin välein.”</p> <p>(‘Peruskartta 1:20 000.’, 1948)</p>		
<p>Purot ja ojat</p>	<p>”Joet, purot, viemärit ja ojat kuvataan erilaisilla merkeillä niiden leveydestä riippuen. Ne varustetaan virtaamissuuntaa sekä koskipaikkoja osoittavilla merkeillä. Selvästi näkyvän puron epämääräiset kohdat, tai sen osittainen kulku maanalaisena, osoitetaan aaltomaisella katkoviivalla. Pelloilla merkitään kokoojaojat ja viemät jättäen varsinaiset sarkaojat pois kartasta” (‘Peruskartta 1:10 000 /1:20 000.’, 1960)</p>		<p>Suoalueet jaetaan ojitettuihin ja ojittamattomiin sen perusteella, kuinka lähellä ne ovat puroja ja ojia.</p>

Valtaosa peruskarttojen georeferointitiedoista saatiin vanhatkartat.fi palvelun ylläpitäjältä Shingle Oy:ltä. Käytännössä kyseessä oli json-tiedosto, joka sisältää karttalehden kulmapisteiden pikselitiedot, maantieteelliset koordinaatit, nimen, valmistumisvuoden sekä latauslinkin kartan tiedostoon.

3.3 Muut aineistot

Maanmittauslaitoksen maastotietokantaa sekä historiallisia ja moderneneja ilmakuvia hyödynnettiin georeferoinnin laaduntarkastuksessa (Taulukko 2) . Lisäksi Maanmittauslaitoksen 10 m korkeusmallia hyödynnettiin luonnottoman jyrkkien suo-ositusten poistamiseen.

Taulukko 2: Tutkimuksen muut aineistot

Aineisto	Tuottaja	Vuosi	Sisältö	Käyttötarkoitus
Maastotietokanta: vedet, rakennukset, tiet, suot, ojat, pellot	Maanmittauslaitos	2023	Suomen alueella sijaitsevien merialueiden, järvien ja jokien, rakennusten, peltojen, ojien, soiden ja teiden sijainti ja muoto.	Georeferoinnin tarkastuspisteiden määrittäminen, soita, oja ja peltoja verrattiin esimerkissä tuloksiin.
Korkeusmalli 10 m	Maanmittauslaitos	2019	Maanpinnan korkeus N2000-järjestelmän merenpinnan tasosta 10 m x 10 m tilaresoluutiolla. Johdettu laserkeilauksella tuotetusta korkeusmalli 2m -aineistosta.	Liian jyrkkien suennusteiden poistaminen
Maanmittauslaitoksen ilmakuva	Maanmittauslaitos	2023	Georeferoituja ortoilmakuvia 0,5 metrin resoluutiolla, kuvattu digikameralla.	Visuaalinen apu georeferoinnin tarkastuspisteiden sijoittamiseen
Maanmittauslaitoksen historialliset ilmakuvat	Maanmittauslaitos	1950-1975	Georeferoituja mustavalkoisia ortoilmakuvia 0,5 metrin resoluutiolla, kuvattu fillmille.	Visuaalinen apu georeferoinnin tarkastuspisteiden sijoittamiseen

3.4 Peruskarttojen esikäsittely

Historiallisten karttojen määritelmät maankäyttöluokista voivat erota merkittävästi nykyisten aineistojen määritelmistä (Leyk ym., 2005; Kaim ym., 2014). Tämän takia vanhojen peruskarttojen ja maastotietokannan kuvausohjeita verrattiin, jotta voitiin varmistaa tuotettavan aineiston vertailukelpoisuus sen moderniin vastineeseen.

Valmiiksi saatujen peruskarttojen georeferointitietojen kattavuus tarkastettiin. Kaikki puuttuvat tai virheelliset karttalehdet lisättiin tai korjattiin json-tiedostoon käsityönä.

Kun selkeimmät virheet oli korjattu ja aineistorajaukset tehty, toteutettiin varsinainen georeferointi, karttalehtien leikkaaminen muotoon ja yhdistäminen virtuaaliseksi rasteriksi. Työhön käytettiin projektin GitHubista (<https://github.com/iioseppa/finnish-historical-base-map-processing>) löytyviä Python-skriptejä, jotka hyödynsivät pääasiassa Rasterio (1.3.8)-, Geopandas (0.13.2)- ja PyQGIS (3.32–3.36)-kirjastoja (Gillies & et. al, 2013; Jordahl ym., 2020; QGIS Development Team, 2024). Virtuaalirasterin luontiin käytettiin GDAL buildvrt -työkalua. Tämän jälkeen georeferoinnin onnistuminen tarkastettiin visuaalisesti jokaiselle karttalehdelle maastotietokannan vesistöjen avulla. Maastotietokannan vesistöt värjättiin vaaleanpunaiseksi, ja asetettiin karttatasojen päälle. Mahdolliset suuret, yli monenkymmenen metrin georeferoinnin epäonnistumiset tulevat tällöin ilmeisiksi myös kaukaa katsottuna (kts. Kuva 9 tuloksista). Tällä tavoin kaikki karttalehdet oli mahdollista käydä läpi yleisellä tasolla suhteellisen nopeasti. Löydetyt virheet korjattiin ja karttalehdet uudelleentarkastettiin.

Lopullisen georeferoinnin tarkkuus varmistettiin valitsemalla satunnaisotannalla 51 karttalehteä. Kuudesta karttalehdestä löytyi sekä 1950-, että 1970-luvun versiot. Jotta myös vanhempia karttoja olisi riittävästi, otettiin 17 manuaalista lisäotosta karttalehdiltä, jotka uudistettiin tutkimusajanjakson aikana. Jokaisesta karttalehdestä etsittiin kolme selkeää kiintopistettä, ja uudistetuille karttalehdille mittaukset tehtiin molemmille versioille, joten yhteensä kiintopisteitä oli 279. Otosten määrä vastaa suuruusluokaltaan aiempaa tutkimusta (Mäyrä ym., 2023). Kiintopisteiksi valittiin rakennuksia ja teidenristeyksiä, joita verrattiin vuoden 2023 maastotietokantaan, ja vain selkeästi samanlaisena pysyneet kohteet otettiin mukaan. Tarvittaessa vanhoja ja uusia ilmakuvia käytettiin tulkinta-apuna. Koska jokaisesta karttalehdestä oli kolme virhemittausta, jaettiin virheet kolmeen luokkaan: pienimmät, keskimmäiset ja suurimmat sijaintivirheet. Pienimpiin virheisiin sijoitettiin kunkin karttalehden pienin virhe, ja samoin toimittiin keskimmäisten ja suurimpien virheiden kohdalla. Sijaintivirheen suuruutta arvioitiin laskemalla **keskineliövirheen neliöjuuri** (eng.

Root-mean-square error, RMSE) ja muita tilastollisia tunnuslukuja sekä kaikille kiintopisteille, että erikseen karttalehtien pienimmille, keskimmäisille ja suurimmille virheille. Lisäksi sijaintivirheiden suuntautuneisuutta tarkasteltiin laskemalla samalla karttalehdellä olevien sijaintivirheiden suuntien välisten kulmien kosinien keskiarvo.

3.5 Opetusaineiston digitointi ja käsittely

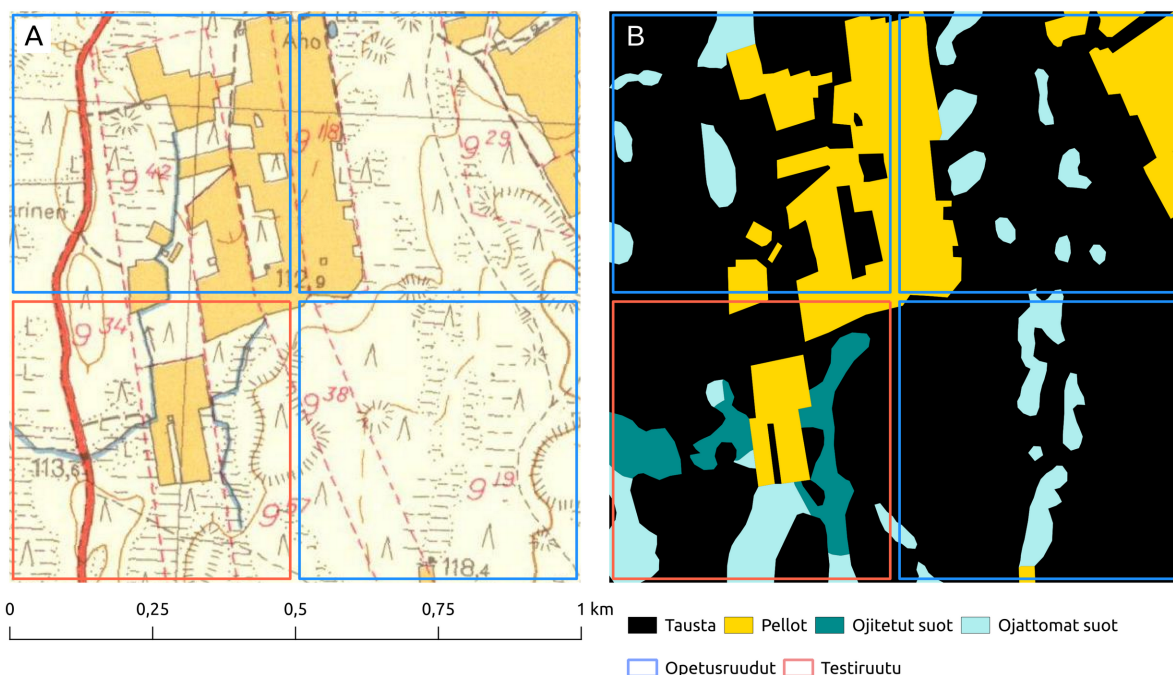
Syväoppivat neuroverkot, jotka esitellään luvussa 2.2, tarvitsevat suuria opetusaineistoja (Alom ym., 2019; Alzubaidi ym., 2021). Tässä tutkielmassa opetusaineisto piirrettiin käsityönä.

Digitointi aloitettiin jakamalla koko tutkimusalue kilometrin ruutukoon ruudukkoon, jonka ruutujen sisään digitoinnit tehtiin. Ruutukoon valintaan vaikutti useampi asia. Ensinnäkin kilometrin ruutu on historiallisten peruskarttojen kohdalla riittävän pieni, jotta visuaalisessa tarkastelussa on mahdollista huomata melko luotettavasti myös pienet kohteet. Toiseksi kilometrin ruutu osoittautui pikselimäärältään sopivaksi. Tässä tutkielmassa käytettävä koneoppimismalli (muokattu U-net) tarvitsee neljän koontikerroksensa takia kuvia joiden leveys ja korkeus voidaan jakaa tasan kahtia vähintään neljä kertaa. Skannattujen peruskarttojen yhden pikselin leveys on noin 1,69 metriä, jolloin neliökilometrin ruudun koko on 591×591 pikseliä. Tämän voisi jakaa yhdeksi 576×576 , neljäksi 288×288 tai kuudeksitoista 144×144 pikselin kuvaksi, jotka täyttävät mallin vaatimukset. Aiemmissä vastaavissa tutkimuksissa on hyödynnetty tyypillisesti 256 pikselin levyisiä kuvia, joten tässä tutkielmassa mallin kuvakooksi valittiin 288×288 pikseliä (Wu ym., 2022; Mäyrä ym., 2023). Kolmanneksi kilometrin ruutukoko on riittävän pieni siihen, että ruutuja on realistista tehdä käsin satoja. Tämä varmistaa, että kuviin päätyy riittävästi peruskarttojen kirjoja.

Kilometrin ruutukoon ruudukosta valittiin satunnaisotannalla 301 ruutua käsin digitoitavaksi. Lisäksi mukaan otettiin karttaruudut Helsingin ja Turun keskustoista, koska otosten joukkoon ei osunut yhtään tiheästi rakennettua kaupunkikeskustaa, joilla on oma aluesymbolinsa vanhoissa peruskartoissa. Ruuduista 56 oli karttalehdellä, joka oli uudistettu. Tällöin sekä uudempi että vanhempi karttalehti digitoitiin, joten yhteensä opetusaineistoa kerättiin 359 km^2 alalta. Digitoitavia peruskartan kohdeluokkia olivat pellot, suot ja ojat (Taulukko 1). Ojia ei eroteltu luonnollisiin ja ihmisen tekemiin, koska peruskarttojen symboliikka ei tee merkittävää eroa näiden välille ('Peruskartta 1:20 000.', 1948).

Suodigitoinnit jaettiin jälkikäsitellyssä ojitetuiksi ja ojittamattomiksi soiksi sen perusteella, olivatko ne 50 m etäisyydellä ojasta tai purosta, mikä vastaa arvioita suo-ojan vaikutusetäisyydestä (Toivonen ym., 2022). Kaikki muut kuin suo- tai peltokohteet luokiteltiin taustaksi.

Jokainen digitoitu neliökilometrin ruutu muutettiin rasterimuotoiseksi ja jaettiin neljäksi 288 x 288 pikselin etikettirasteriksi, joiden alalta otettiin talteen myös kuva vanhasta peruskartasta. Näistä lounaisnurkan kuva ja etiketti otettiin erikseen testaamista varten, loput käytettiin opetukseen (Kuva 8).



Kuva 8: Esimerkki opetus- ja testiruutuihin jakamisesta yhdellä kilometrin ruudulla Lounaisnurkan ruutu säästetään vain testaamiseen, loput ruudut ovat opetuskäyttöön. A) Alkuperäinen karttakuva B) Käsin digitoitu etiketti

3.6 Koneoppimismallin kouluttaminen

Kuvantunnistusmallin perustana käytettiin U-nettiä (Ronneberger ym., 2015). Se on suhteellisen kevyt, suorituskykyinen ja yleisesti käytetty malli kuvien ositukseen (Alzubaidi ym., 2021). Mallia muokattiin asettamalla sen keskelle tilallisesti väljä pyramidikoontikerros, joka parantaa mallin kykyä tunnistaa eri kokoluokkien kohteita (Chen ym., 2018). Tarkka mallin implementaatio on peräisin Wun ja kumppanien artikkelista (2022), joskin Sørensen-Dicen kertoimeen perustuva tappiofunktio vaihdettiin luokkakohtaisesti painotettuun luokalliseen ristientropiaan (Eng. Categorical cross-entropy) (Implementaatio Fischer, 2020).

Samalla luovuttiin myös alkuperäiseen malliin kuuluneen epävarmuusarvion tuottamisesta analyysiin kuluvaan työmäärän vähentämiseksi. Koska digitoitujen luokkien sekä taustan pinta-alat vaihtelivat suuresti, mallin tappiofunktiossa pienempialaisia luokkia painotettiin seuraavilla kertoimilla, jotka saatiin jakamalla maankäyttöluokan pinta-ala opetusaineistossa taustan pinta-alalla opetusaineistossa: tausta: 1,0; pellot: 7,3; ojitetut suot: 11,5 ja ojattomat suot: 7,2. Näin eri luokkien todellinen painoarvo mallille muutetaan samanlaiseksi niiden eriävistä pinta-aloista huolimatta.

Malli koulutettiin käyttäen Keras- ja Tensorflow- kirjastojen versioita 2.11 sekä yhtä Nvidia RTX 4080 näytönohjainta (Abadi ym., 2015; Chollet & ym., 2015). Neljännes opetusaineistosta varattiin opetuksen aikaiseen validointikäyttöön. Mallia koulutettiin 72 epookkia, kunnes validointiaineiston tappiofunktio ei ollut parantunut viiteentoista epookkiin. Aineiston keinotekoisia lisäämismenetelmiä ei käytetty, koska malli saavutti nopealla kouluttamisella erinomaisen tarkkuuden ilman niitä. Lisäämismenetelmiä testattiin, mutta ne tekivät oppimisesta huomattavasti epävakampaa ja hitaampaa. Koulutetun mallin toimivuus tarkistettiin testiaineistolla. Ne tarkkuuden mittaukset, jotka tässä tutkielmassa esitellään, tehtiin kuitenkin vasta jälkikäsitteilyiden jälkeen.

3.7 Karttalehtien luokkaositus ja jälkikäsitteily

Luokkaositus suoritettiin erikseen jokaiselle karttalehdelle. Ensin karttalehti jaettiin 288 x 288 pikselin ruuduiksi siten, että vierekkäiset ruudut olivat reunoistaan päällekkäisiä 64 pikselin levyisellä kaistalla. Karttalehden oikeaan- ja alalaitaan lisättiin tarvittava määrä mustalla täytettyjä rivejä tai sarakkeita, jotta karttalehti olisi jaettavissa tasan 288 x 288 pikselin kuviksi. Nämä kuvat syötettiin koulutetulle mallille, joka ennusti jokaiselle pikselille, kuuluuko se taustaan, peltoihin, ojitetuihin soihin vai ojattomiin soihin. Kuvien ennusteen laatu oli reunoilla hieman keskialuetta heikompi, joten reunimmaisesta 32 pikseliä poistettiin, paitsi karttalehden reunoilla. Leikattavat kuva-alueet lomitettiin siten, että kaikkialta karttalehden alueelta on saatavilla laadukkaita aineistoja, jotka ovat olleet käsitellyn kuvan keskialueella. Näin syntyneet 224 x 224 pikselin laadukkaat kuvat parsittiin takaisin yhdeksi karttalehdeksi. Vastaavaa menetelmää on käyttänyt osituksen laadun parantamiseen esim. Mäyrä ym. (2023). Mallin monikanavainen desimaalimuotoinen ulostulo muutettiin yksikanavaiseksi kokonaislukuluokitteluksi NumPYN Argmax-funktiolla (Harris ym., 2020).

Mallin havaittiin ennustavan avokallioiden reunoja tai louhikoita toisinaan soiksi niiden tiheän mustan kuvionnin takia (kts. Kuva 16B ja D). Tämän takia lopputuloksista poistettiin

suoalueet, joiden jyrkkyys on yli viisi astetta, koska soiden kehittyminen on hyvin rajoittunutta niin suurissa jyrkkyyksissä (Zhao ym., 2014; Sun ym., 2023). Erottelu aloitettiin interpoloimalla maanmittauslaitoksen 10 m x 10 m korkeusmalli samaan resoluutioon kuin alustava ositus, eli 1,69 m x 1,69 m. Kymmenen metrin resoluution korkeusmallia käytettiin, vaikka myös kahden metrin resoluution korkeusmalli on saatavilla. Tähän oli kolme syytä: Ensinnäkin kahden metrin resoluutioinen korkeusmalli melkein koko Suomen alueelta vaatii noin 300 GT:a säilytystilaa ja sen prosessointi vaatisi paljon aikaa. Toiseksi eroja eri resoluutioiden lopputuloksissa testattiin, ja ne havaittiin suhteellisen pieniksi. Kolmanneksi interpoloitu kymmenen metrin resoluutio tuotti tuloksia, jotka olivat yleistyskaalaltaan lähempänä vanhoja peruskarttoja.

Peruskarttaan kuvatut aluekohteet olivat mittauksien perusteella kapeimmillaan 20-30 metriä leveitä. Kaikki tätä pienemmät kohteet ovat suurella todennäköisyydellä virheellisiä luokituksia. Tämän takia kaikki alle 13 x 13 pikselin eli 22 m x 22 m kokoiset yksittäiset kohteet poistettiin. Raja-arvo valittiin sillä perusteella, että se poisti luokittelusta melkein kaikki virheellisesti pelloiksi luokitellut yli kaksimetriset luonnonjyrkänteet (Kts. Kuva 16C). Poistoja ei kuitenkaan tehty taustaan kuuluville pikseleille.

Erilliset karttalehdet muutettiin ensin yhdeksi virtuaalirasteriksi GDAL buildvrt-työkalulla, joka muutettiin yhdeksi Geotiff-tiedostoksi GDAL translate-työkalulla.

Lopuksi valmiista aineistosta etsittiin visuaalisella silmäilyllä alueita, jotka ovat muuttuneet huomattavasti. Näistä alueista valittiin yksi esimerkinomaiseen muutostarkasteluun, joka demonstroi aineiston potentiaalia maankäytön muutosten tutkimukseen.

3.8 Luokittelun tarkkuuden testaus

Jälkikäsitellyn osituksen laatu varmistettiin vertaamalla ositusta testiruutujen käsin piirrettyihin etiketteihin. Ensin testiaineiston käsinpiirrettyjä etikettejä verrattiin visuaalisesti mallin tekemään ositukseen, ja löydetyt, pääasiassa pienet puutteet käsintehdyissä piirroksissa korjattiin. Ennusteille laskettiin sekaannusmatriisi, siitä johdettu Cohenin kappa sekä keskimääräinen LYS. LYS laskettiin erikseen pienille, keskisuurille ja suurikokoisille kohteille. Näiden luokkien raja-arvoina käytettiin kooltaan alle 32^2 px, 32^2 px – 96^2 px ja yli 96^2 px kuva-alueita, mikä vastaa laajalti käytetyn MS COCO- konenäköaineiston jaottelua (Lin ym., 2015). Lisäksi LYS laskettiin ilman kokoerottelua sekä kaikille kuville, että

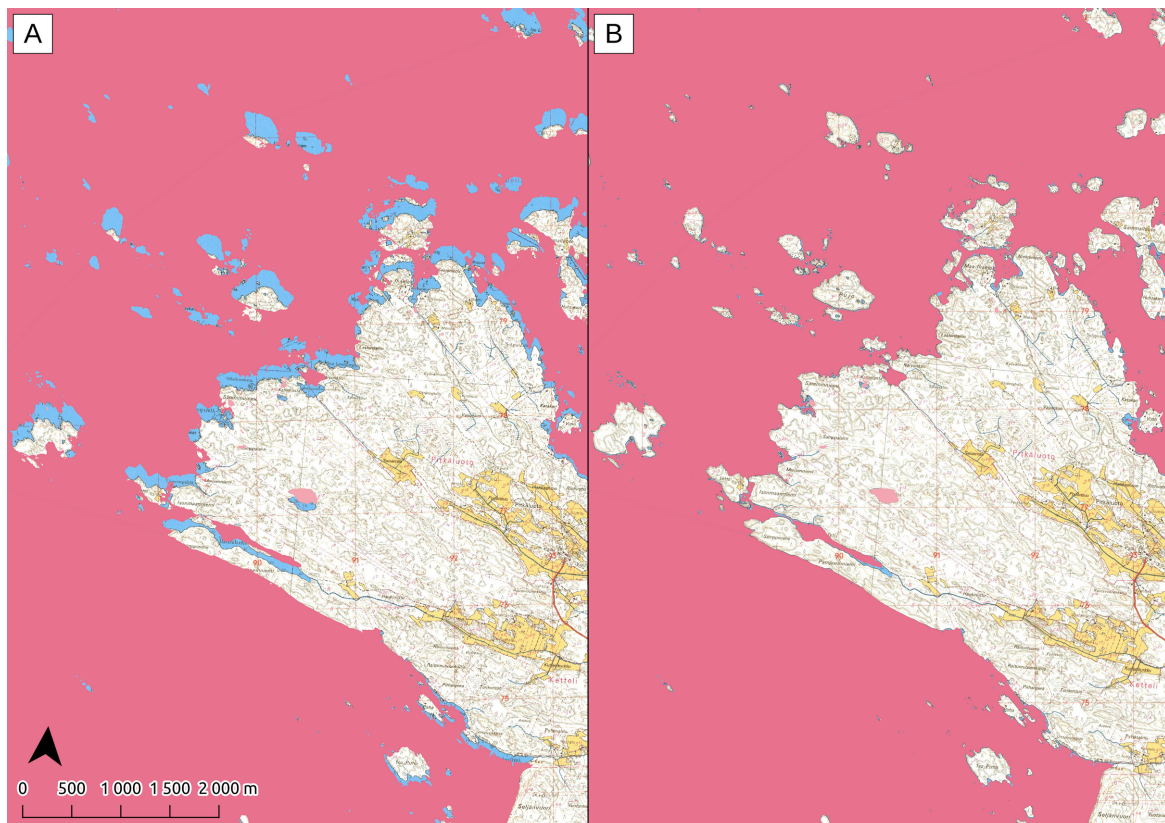
pelkästään niille kuville jotka eivät olleet pelkkää taustaa. Näin tehtiin, koska pelkkää taustaa olevat kuvat poistettiin erikokoisten kohteiden tarkastelusta.

4 Tulokset

4.1 Esikäsittely ja digitointi

Soiden, peltojen ja ojien määritelmät olivat käytännössä identtiset maastotietokannan ja vanhojen peruskarttojen kuvausohjeiden välillä.

Noin 100 karttalehteä joko puuttui tai oli virheellisesti georeferoitu saaduissa georeferointitiedoissa karttalehden epäsäännöllisen muodon tai inhimillisten virheiden vuoksi (Kuva 9). Valtaosa puuttuvista georeferointitiedoista johtui karttalehden epäsäännöllisestä muodosta. Mukana oli myös Maanmittauslaitoksesta lähtöisin olevia virheitä, kuten kuvatiedostoja, jotka on nimetty väärällä karttalehtikoodilla tai tiedostoja joiden päätte on ”JPG” sen sijaan että se olisi ollut ”jpg”. Havaitut virheet korjattiin esikäsittelyssä.



Kuva 9: Vaaleanpunaiseksi värjätty maastotietokannan vesistöt paljastavat historiallisen peruskartan päälle asetettuna selkeät georeferointivirheet jo kaukaa. A) Karttalehden sijainti on vääristynyt kohdistuspisteen pikselikoordinaattien virheellisen kirjaamisen takia, minkä takia karttalehden siniset vesistöt paljastuvat esiin. B) Korjatun georeferoinnin alta ei pilkota juurikaan sinistä. Vesistöt: (Maanmittauslaitos, 2023c) Kartta: (Peruskarttalehti 113106 + 113103 Ketteli, 1968)

Digitoinnissa tuotettiin 1077 opetuskuva ja 359 testakuva noin 120:ssä työtunnissa.

Opetuskuvista 881 ja testikuvista 298 sisälsi jotain kohdeluokkaa, loput sisälsivät pelkkää taustaa. Yhteensä digitoinnissa kuvattiin peltoja 34,3 km², ojitettuja soita 21,8 km² ja

ojittamattomia soita 34,9 km². Taustaan kuului täten 268 km². Vertaamalla testiaineistoa mallin ennusteeseen testiaineistosta löydettiin ja korjattiin noin kolmekymmentä pääasiassa pienikokoista kohdetta, jotka olivat jääneet huomaamatta ja digitoimatta ensimmäisellä digitointikerralla. Huomaamatta jääneiden kohteiden lisäksi löytyi yksittäisiä muokkaamista tarvitsevia digitointeja.

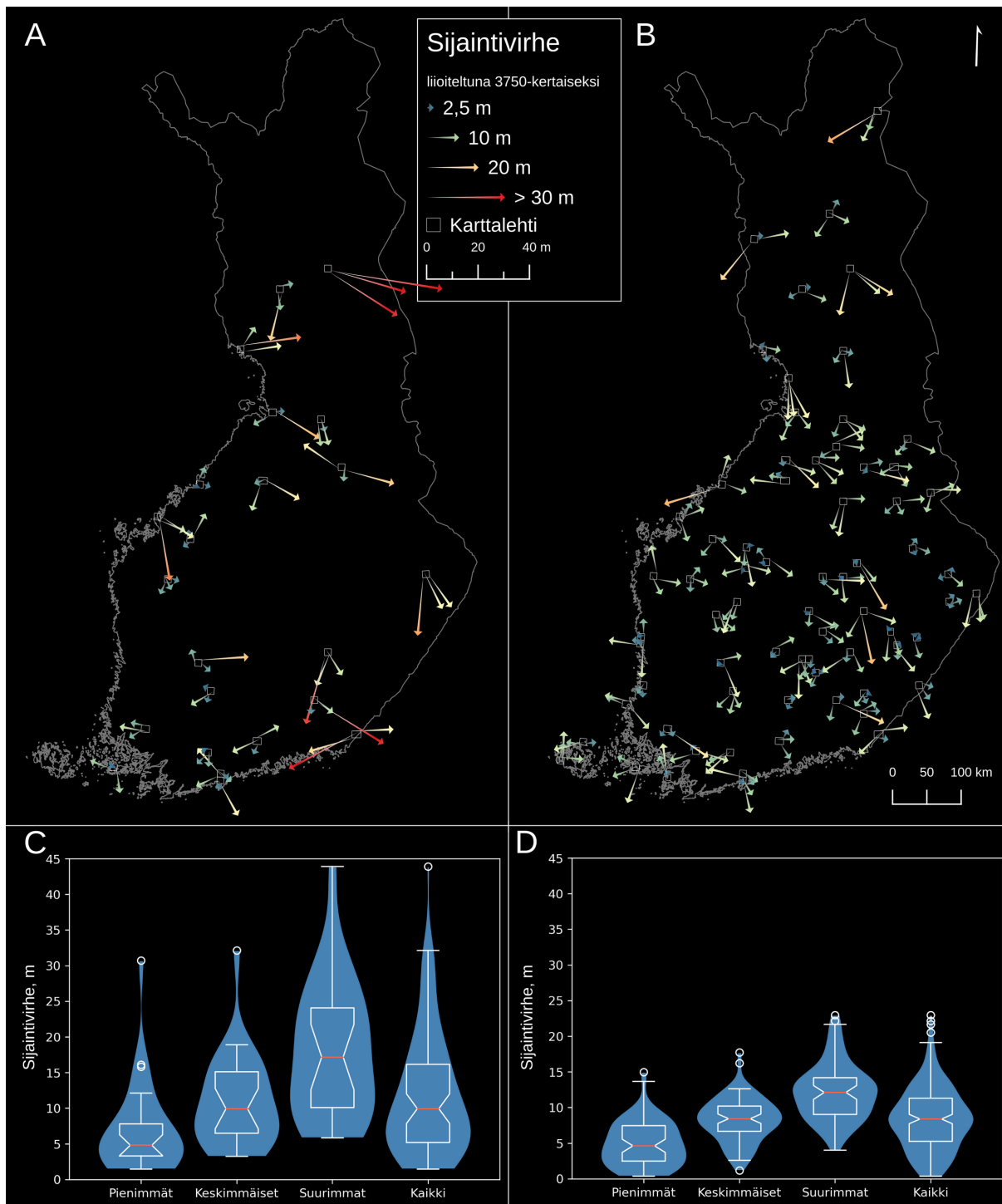
4.2 Rakennusten ja risteysten sijaintivirhe

Keskineliövirheen juuri oli päivitettyjen karttalehtien vanhemmille, noin 1950-luvulla tehdyille versioille noin 15 m ja pääasiassa 1960–1970 -luvuilla päivitettyjen karttalehtien uudemmille versioille ja kertaalleen kartoitetuille karttalehdille noin 10 m (taulukko 3). Jokaisella mitatulla karttalehdellä oli kolme mittauskohdetta. Vaihtelu yksittäisten karttalehtien sisällä oli melko suurta, sillä karttalehtien suurimpien mitattujen sijaintivirheiden RMSE oli yli kaksinkertainen pienimpiin verrattuna.

Taulukko 3: Keskineliövirheiden neliöjuuret olivat pääasiassa lähellä kymmentä metriä. Kaikki arvot ovat keskineliön juuria ja yksikkönä on metri. Yhden pikselin leveys oli 1,69 m.

RMSE, metriä	Pienimmät	Keskimmäiset	Suurimmat	Kaikki
Vanhemmat karttalehdet	9,6	12,7	20,4	14,9
Uudemmat karttalehdet	5,8	8,9	12,9	9,6

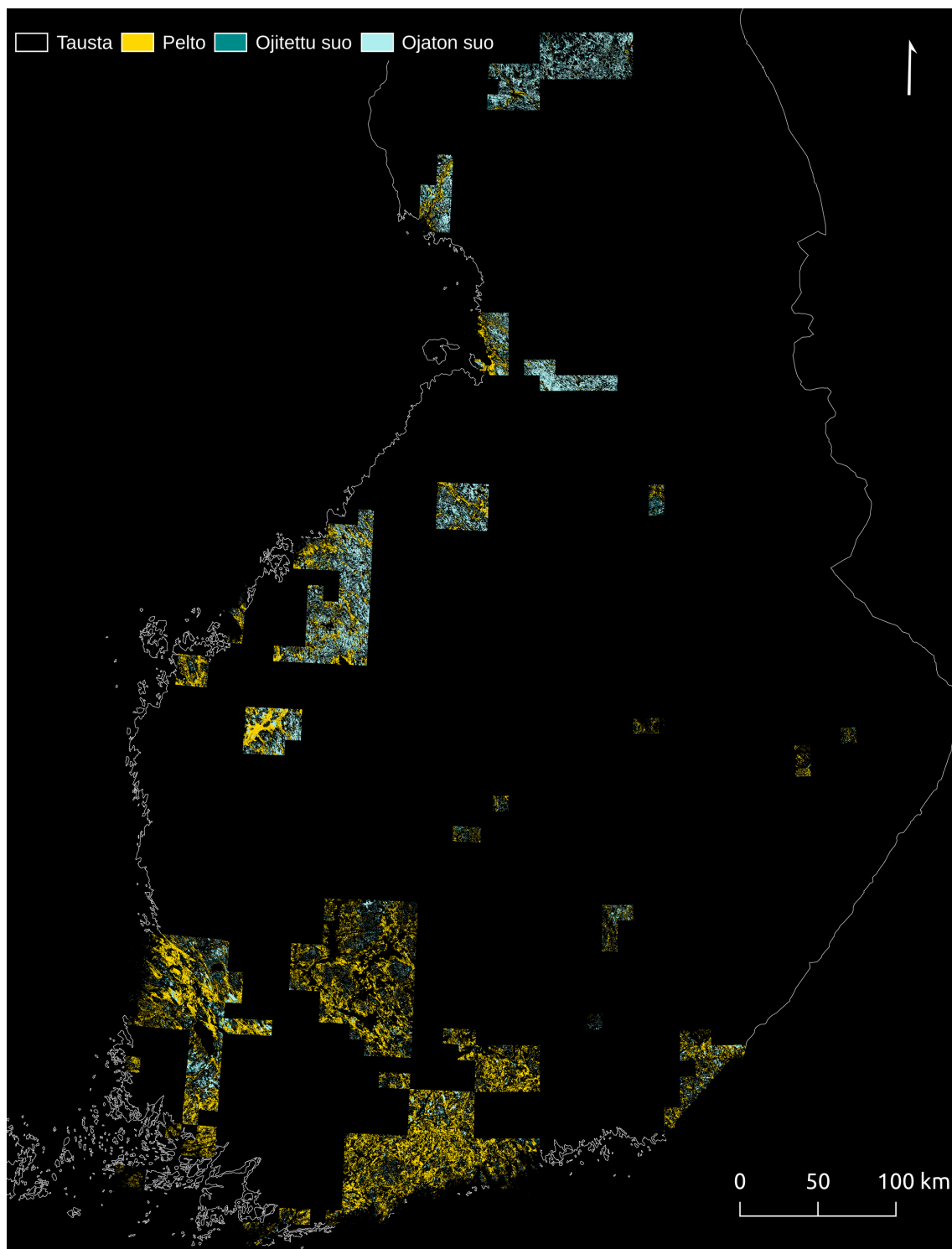
Alkuperäisessä kartoituksessa ja georeferoinnissa syntyneet sijaintivirheet olivat pääasiassa alle kymmenen metrin suuruisia, mutta etenkin vanhempien, 1950-luvun karttalehtien kohdalla hajonta oli suurta (Kuva 10). Suurimmat havaitut sijaintivirheet olivat yli kolmekymmentä metriä, eivätkä ne selittyneet pelkästään yksinkertaisella virheellä georeferoinnissa. Useimmat suurimmat virheet esiintyivät karttalehdillä, joissa muut sijaintivirheet olivat pienempiä. Kokonaisuutena yksittäisen karttalehden sijaintivirheet eivät olleet merkittävästi systemaattisesti samansuuntaisia. Samalta karttalehdeltä mitattujen virheiden suuntien kosinien keskiarvo oli 0,2. Täydellinen samansuuntaisuus tuottaisi arvon 1, ja kolmella sijaintivirhemittauksella suurin mahdollinen keskimääräinen eriävyys on -0,5, jolloin kaikki kulmat eroaisivat toisistaan 120 astetta.



Kuva 10: Mitatun sijaintivirheen suuruus vaihtelee melko suuresti, myös saman karttalehden sisällä. Hajonta on suurempaa vanhemmilla karttalehdillä. Jokaiselta mitatulta karttalehdeltä mitattiin sijaintivirhe kolmesta kiintopisteestä. A ja C sisältävät tutkimusajanjaksona uudistettujen karttalehtien vanhemmat versiot, $n = 69$ B ja D kertaalleen kartoitetut ja päivitettyjen lehtien uudemmat versiot, $n = 210$. A ja B) Georeferointivirheen suunta ja suuruus 3750-kertaiseksi liioiteltuna. Nuoli osoittaa sijainnista maastotietokannassa kohti sijaintia kartalla. C ja D) Yhdistetty viulu- ja laatikkokuvaaja karttalehtien pienimmän, keskimmäisen ja suurimman sijaintivirheen suuruuden jakaumasta sekä kaikkien yhteisestä jakaumasta. Laatikkokuvaajaan on merkitty punaisella mediaani, ja lovet osoittavat mediaanin sijainnin 95 % todennäköisyydellä.

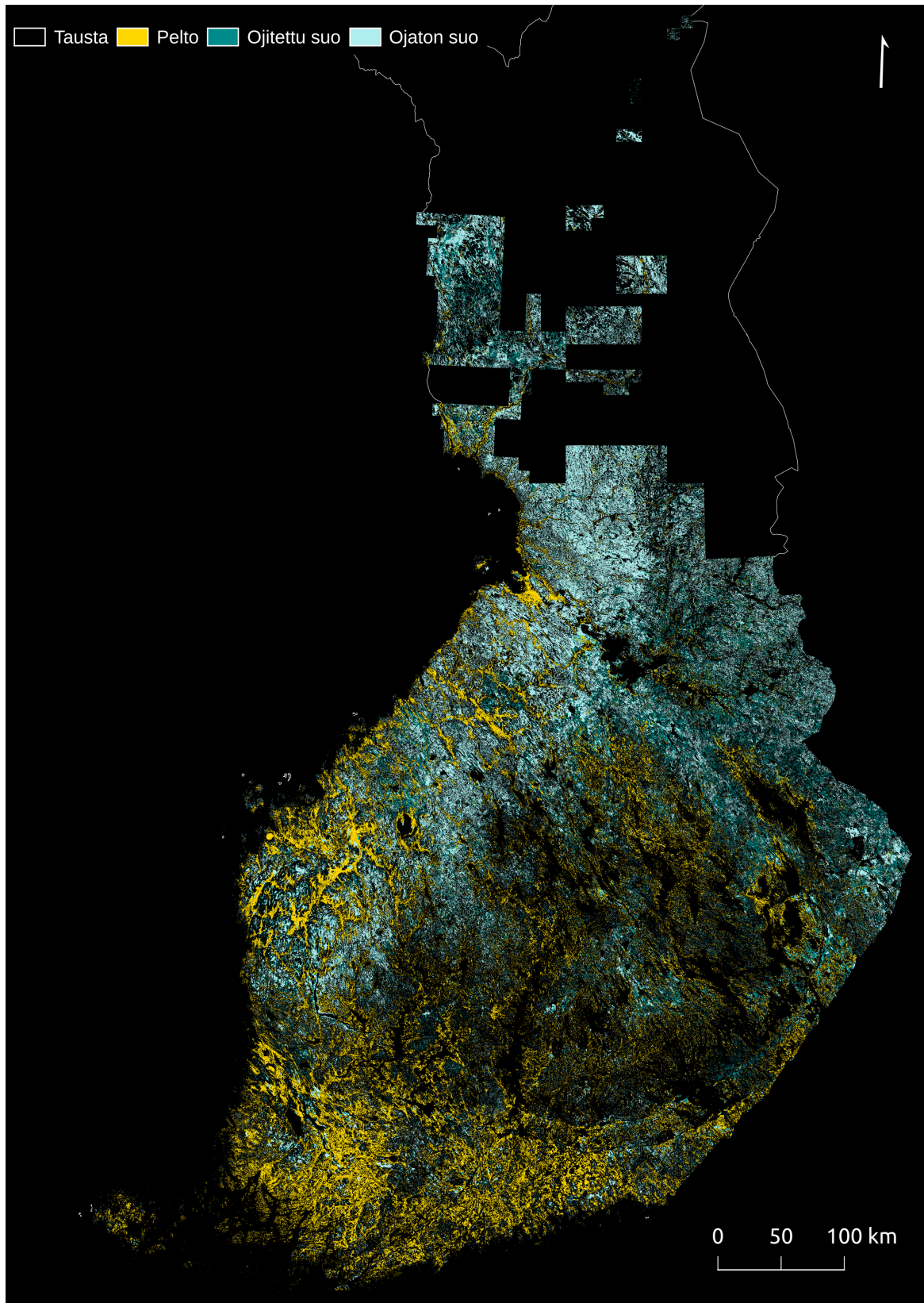
4.3 Karttakohteiden ositus

Lopullisen luokittelun tuloksena peruskarttojen ennen vuotta 1978 uudistettujen karttalehtien vanhemmilta versioilta 6300 km² eli 16,9 % luokittui pelloiksi, 635 km² eli 1,7 % ojitetuiksi soiksi ja 3900 km² eli 10,5 % ojattomiksi soiksi (Kuva 11).



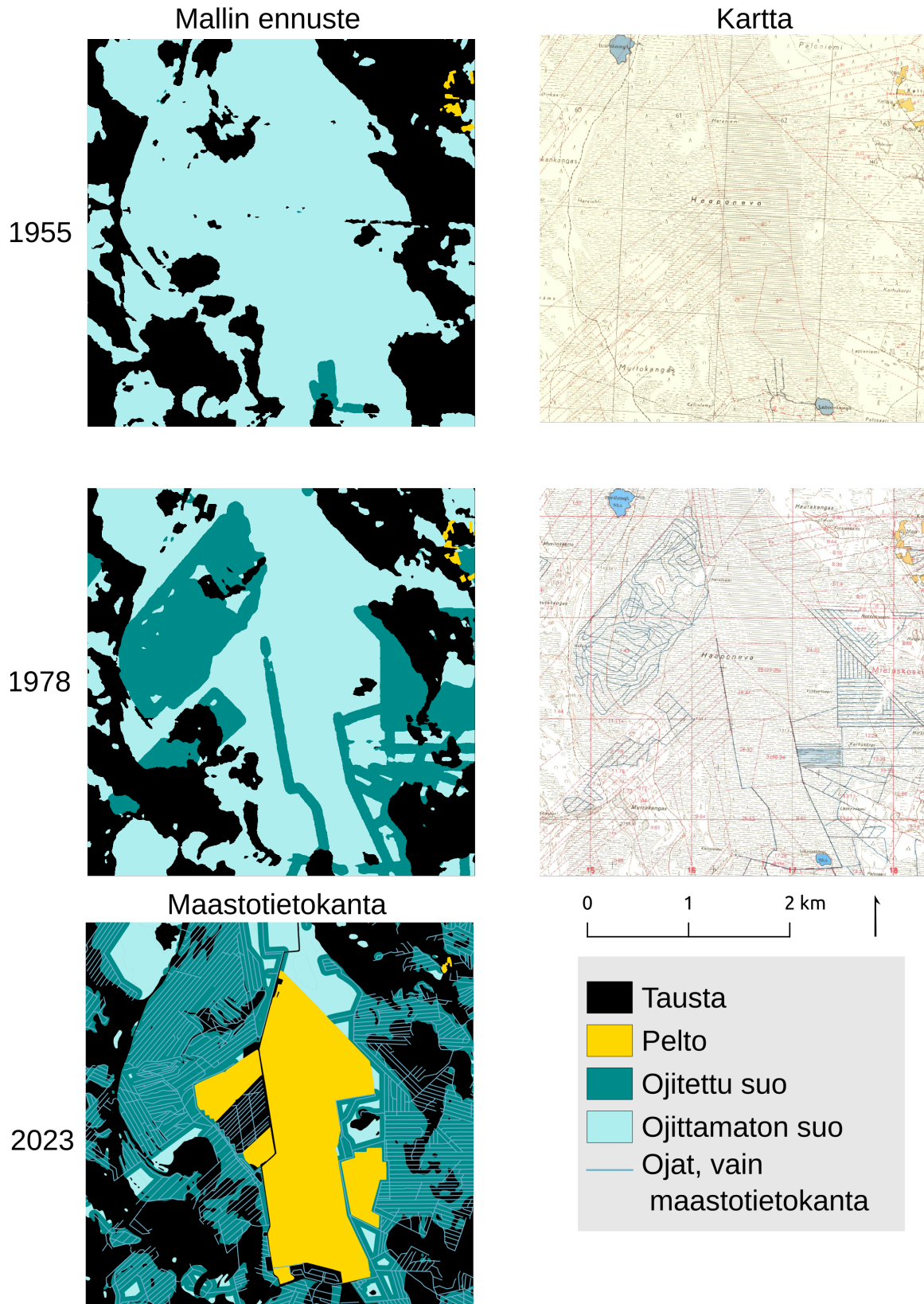
Kuva 11: Luokittelun tulos tutkimusaikana päivitettyjen karttalehtien vanhemmille versioille

Peruskarttojen ennen vuotta 1978 uudistettujen karttalehtien uudemmista versioista ja kertaalleen kartoitetuilta karttalehdiltä 28 640 km², eli 10,5 % luokittui pelloiksi, 16 490 km², eli 6,1 % ojitetuiksi soiksi ja 33 530 km², eli 12,3 % ojittamattomiksi soiksi (Kuva 12).



Kuva 12: Luokittelun tulos ennen vuotta 1978 uudistettujen karttalehtien uudemmille versioille ja kertaalleen kartoitetuille karttalehdille

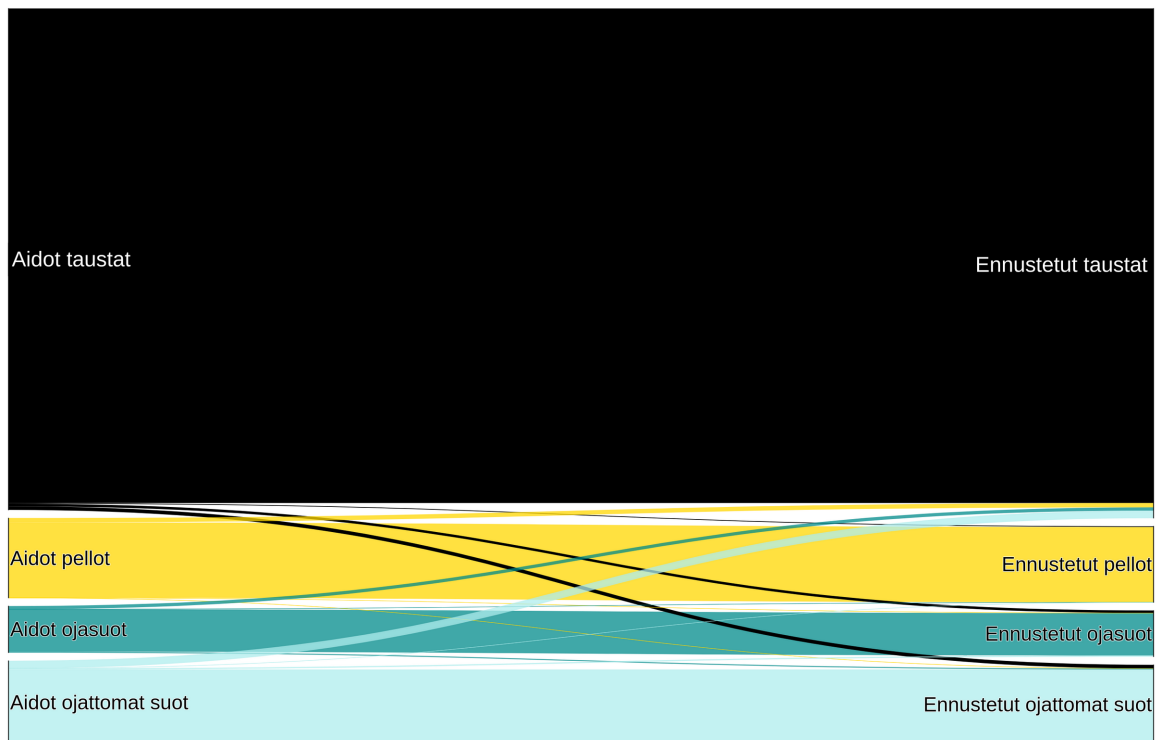
Näiden karttalehtien kohdalla on myös mahdollista huomata selkeä ero Itä- ja Länsi-Suomen ojitusasteen välillä. Ero johtuu ainakin osittain siitä, että monet Pohjanmaan kartat ovat vanhempia kuin Itä-Suomessa (Kuva 6), joten ojitus ei ollut kartoitushetkellä ehtinyt edetä siellä yhtä pitkälle kuin idässä. Tutkimusajanjaksolla pohjanmaalla kahteen kertaan kartoitettujen karttojen uudemmat versiot sisältävät huomattavasti vanhempia versioita enemmän ojitettuja soita (Kuva 12). Otetaan esimerkiksi muutosten laajuudesta Haaponeva Haapajärvellä (Kuva 13). Se on ollut vielä 1950-luvulla käytännössä luonnontilainen, mutta 1978 alueella on ollut jo laaja ojitusverkosto. Verrattaessa näitä vuoden 2023 maastotietokantaan huomataan, että ojitusverkosto on laajentunut kattamaan liki koko alueen, ja merkittävät osat aiemmasta suoalueesta on muutettu viljelysmaaksi. Toiset osat entistä suota taas ovat kuivuneet ja muuttuneet niin paljon, että niitä ei enää lasketa suoksi.



Kuva 13: Esimerkki suon kuivaamisen aiheuttamista muutoksista Haaponevalla, Haapaveden kunnassa. Vuonna 1955 suo on ollut liki luonnontilainen etelälaidassa näkyviä ensimmäisiä oja lukuunottamatta. Vuonna 1978 laajoja osia alueesta on ojitettu. Vuodesta 1987 vuoteen 2018 alueella on ollut useita turpeentuotantoalueita, jotka on muutettu viime vuosina pelloiksi (Orjasniemi & Alapuranen, 2023). Aineistot: (Peruskarttalehti 2433 09 Ainali, 1955 ja 1978; Maastotietokanta 2023)

4.4 Osituksen laadunarviointi

Osituksen kokonaistarkkuus oli 96,6 % ja Cohenin kappa oli 0,93. Malli ei myöskään ylikouluttautunut merkittävästi, sillä validointitappio erosi vain suhteellisen vähän opetustappiosta. Eri maankäyttöluokkien tarkkuudessa oli kuitenkin eroja (Kuva 14).



Kuva 14: Sekaannusmatriisi visualisoituna Sankey-diagrammina. Malli sekoittaa eniten soita taustaan tai toisinpäin, mutta peltojen ja suoluokkien sekaantumiset ovat harvinaisia.

Sekaannusmatriisi löytyy kokonaisuudessaan liitteestä 1.1. Suoluokkien tarkkuus oli alempi kuin peltojen tai taustan, mutta silti varsin hyvä (Taulukko 4). Jos ojitettuja ja ojittamattomia soita tarkastellaan yhdessä pelkästään suoalueina, on niiden tarkkuus parempi kuin erikseen.

Taulukko 4: Tuottajan ja käyttäjän tarkkuudet tutkituille luokille. Yksiköt ovat prosentteja.

Luokka	Tausta	Pellot	Ojasuot	Ojittamattomat suot	Suot yhdessä
Tuottajan tarkkuus	98,6	94,7	91,5	89,9	92,1
Käyttäjän tarkkuus	97,1	99,0	91,7	94,3	95,0

Taustaan kuuluvat pikselit luokittuivat virheellisesti useimmiten ojattomiksi soiksi ja suhteessa pinta-alaan useimmiten ojasoiksi. Huomionarvoista on, että malli sekoittaa myös oikeasti suota olevia alueita herkemmin taustaan, kuin suoluokat sekoittuvat toisiinsa. Tämä kielii siitä, että mallilla on enemmän pieniä haasteita löytää suoalueiden täsmälliset rajat taustaan kuin arvioida ojan läheisyyttä suohon.

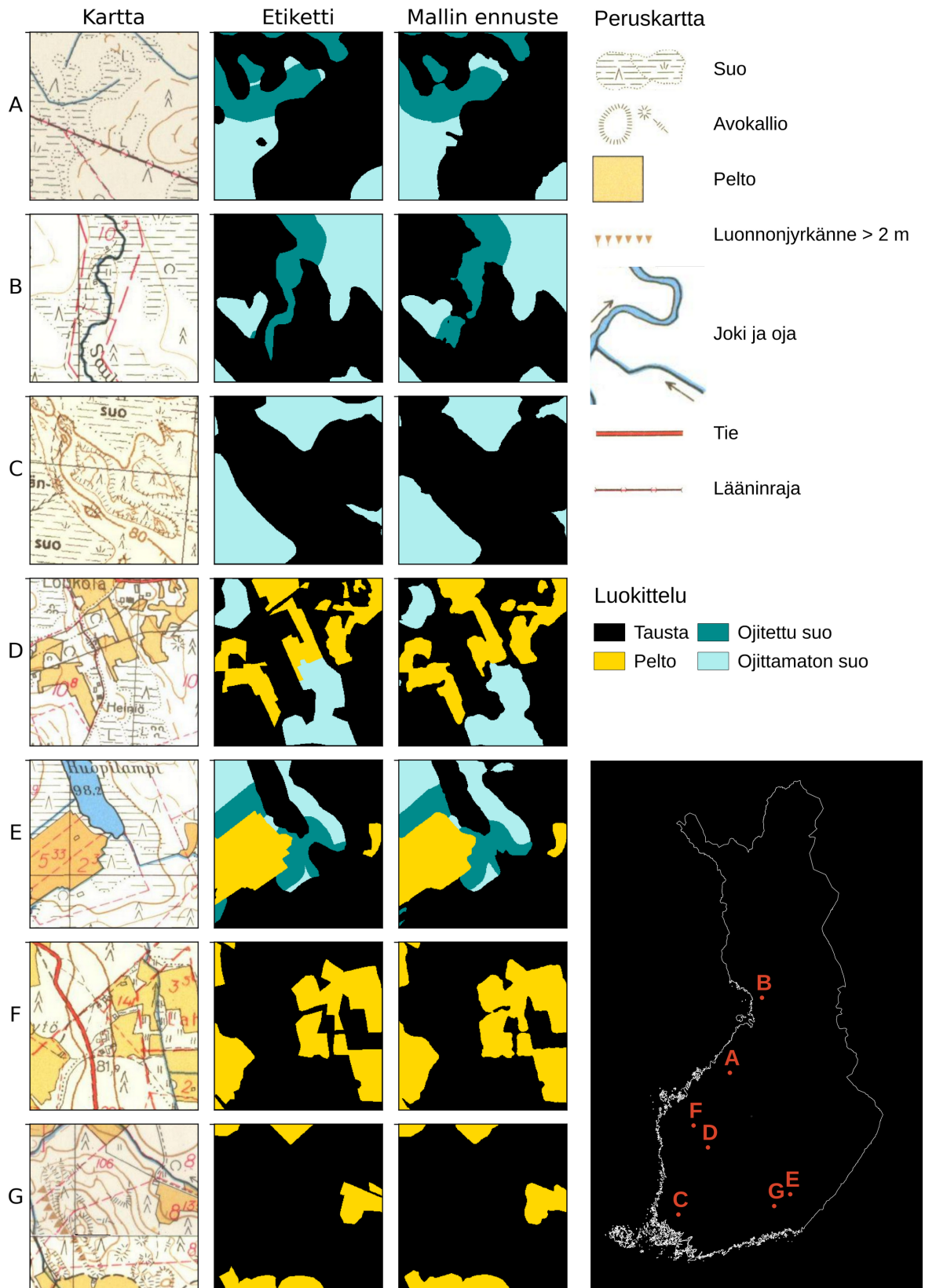
Luokkakohtainen LYS oli selkeästi parempi suurille, kuin pienille kohteille (Taulukko 5, liite 2). LYSien mediaanit osoittavat, että valtaosa kohteista tunnistettiin. Peltojen keskimääräinen LYS oli 0,94, ojitetujen soiden 0,85 ja ojittamattomien soiden 0,76, eli keskimäärin soiden LYS oli 0,81. Suuret kohteet löytyivät liki aina, ja ongelmia esiintyi lähinnä niiden tarkan reunan kanssa. Etenkin pienien suokohteiden tunnistaminen oli kuitenkin pääasiassa epäonnistunutta.

Taulukko 5: Mallin tarkkuus oli erinomainen suurien ja hyvä keskisuurien kohteiden kohdalla, mutta pienet kohteet jäivät usein tunnistamatta. Erityisen merkittävää tämä on suoluokkien kanssa. Kokoluokkien määritelmät ovat peräisin MSCOCO-aineistosta (Lin ym., 2015).

LYS:in mediaani	Pienet (< 32 ² px)	Keskisuuret (>32 ² px, <96 ² px)	Suuret > 96 ² px	Kaikki ilman tyhjiä	Kaikki tyhjien kanssa
Pellot	0,66	0,89	0,94	0,92	1,0
Ojitetut suot	0,0	0,76	0,86	0,77	1,0
Ojittamattomat suot	0,22	0,69	0,89	0,73	0,85

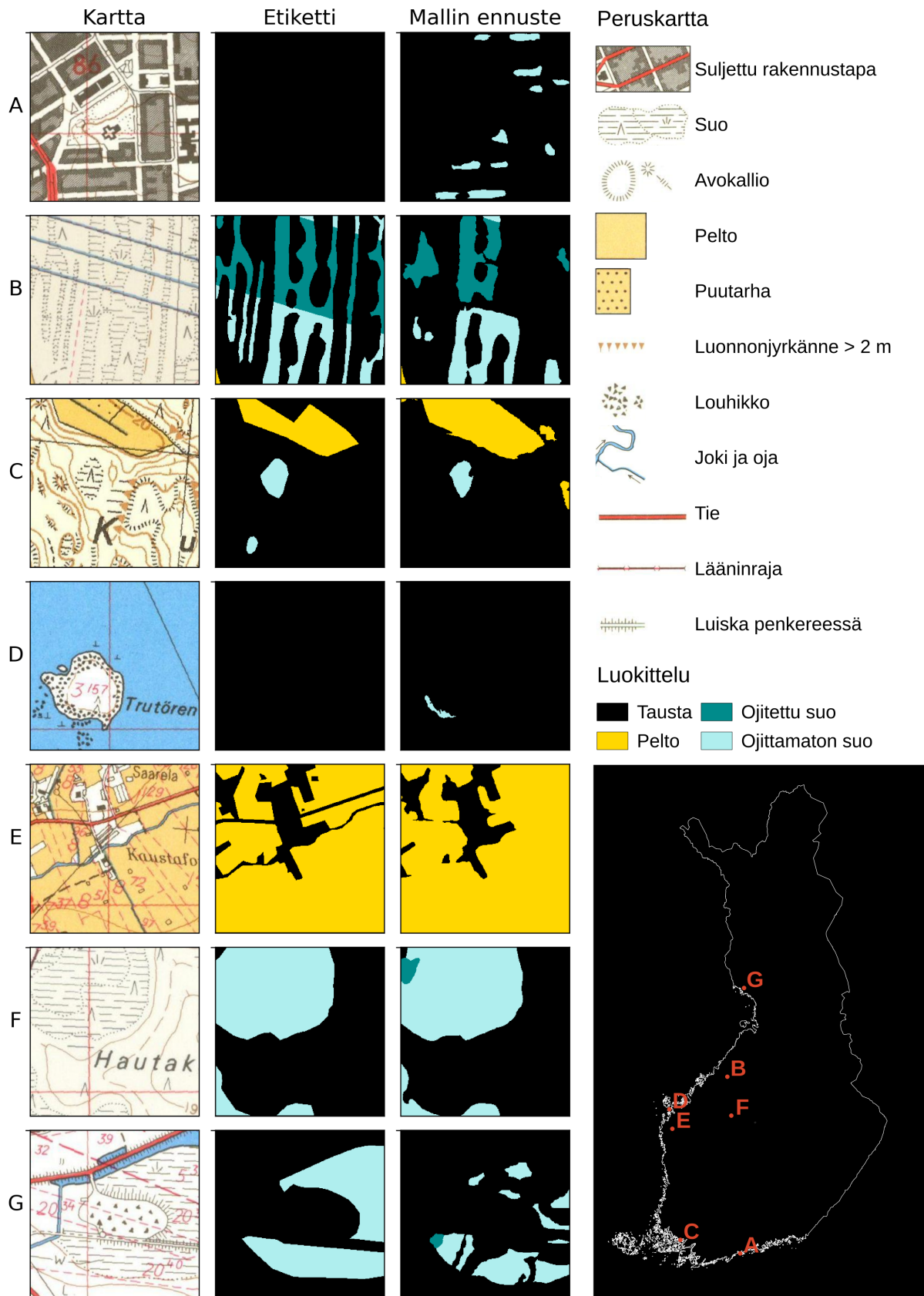
* Käsiteltävässä aineistossa 32² px = 0,29 ha ja 96² px = 2,65 ha, yhden testiruudun ollessa 288² px = 23,8 ha. Tyhjiillä tarkoitetaan ruutuja, joissa sekä digitointi että ennuste ovat pelkkää taustaa.

Pääasiassa malli ennusti järkeviä tuloksia, joissa suurimmat ongelmat koskivat vain reunojen tarkkaa sijaintia (Kuva 15). Pääasiassa tummat viivamaiset kohteet eivät sekoittuneet ojiin, ja malli osasi arvioida tarkasti opetuksessa käytetyn 50 m etäisyyden ojasta (Kuvat 15A, B ja E). Myöskään tiheästi tummia symboleita sisältävät alueet eivät pääasiassa sekoitu soihin (Kuvat 15 C, D, F ja G). Peltojen tunnistus on yleisesti liki moitteetonta pientä reunojen pyöreyttä lukuunottamatta (Kuvat 15 D, E, F ja G).



Kuva 15: Pääasiassa mallin ennusteet olivat tarkkoja, eivätkä sekoittaneet merkittävästi luokkia toisiinsa.

Malli ei kuitenkaan ole täysi ongelmaton. Osittain taustan sekoittuminen suoksi sekä pienien kohteiden alhainen LYS johtuu myös siitä, että malli ennustaa toisinaan soiksi tiettyjä symboleita, jotka eivät ole oikeasti soita, kuten avokallioiden reunoja, kivikoita tai kaupunkialueiden viivoituksia (Kuvat 16A ja D). Suoluokkien suurimmat ongelmat johtuvat siitä, että pienet tai ohuet kohteet jäivät tunnistamatta (Kuvat 16B ja C). Malli voi sekoittaa myös muita kartalla esiintyviä viivakohteita ojiksi, mikä voi aiheuttaa virheellisiä ojasuoluokituksia (Kuva 16F). Peltojen tarkkuus on yleisesti hyvin korkea. Selkeimmät virheluokittelut pienien kohteiden pois jäämisen lisäksi ovat peltoja halkovien teiden sisällyttäminen pelloiksi ja tiettyjen jyrkännetyyppien sekoittuminen pelloiksi (Kuvat 16C ja E). Malli ei myöskään erota puutarhoja pelloista (Kuva 16C). Jyrkkyyden perusteella tehty suorajaus voi aiheuttaa myös virheitä alueilla, joilla on tehty mittavia maanmuokkaustöitä kartoitushetken ja nykyhetken välillä, koska käytetty korkeusmalli perustuu 2010-luvulla tehtyihin mittauksiin (Kuva 16G). Tämän takia tuotetusta aineistosta julkaistaan myös korkeusfiltteröimätön malli.



Kuva 16: Esimerkkejä tilanteista, joissa malli tekee virheitä. Kaikkien ruutujen sivunpituus on 288 px eli 487 m. Malli ei ole nähnyt näitä kuvia koulutuksen aikana.

A) Tiheä kaupunkirakentaminen sekoittuu ojittomaksi suoksi.

- B) Ohuet suokaistaleet jäävät toisinaan tunnistamatta.
- C) Malli ei osaa erottaa puutarhaa ja peltoa toisistaan. Myös pienikokoinen suo jää huomaamatta. Lisäksi liian lähelle toisiaan piirretyt yli kaksimetriset luonnonjyrkänteet kuvautuvat toisinaan pelloiksi.
- D) Malli ennustaa toisinaan myös lohkarikkoja soiksi.
- E) Kapeat yksityiskohdat katoavat usein ennusteesta.
- F) Toisinaan malli erehtyy pitämään viivamaisia kohteita ojina.
- G) Kaivosalueen sivukivikasat ovat muuttuneet kartoitusajankohdan ja modernin korkeusmallin välillä, minkä takia jyrkkyyksifiltteröinti poistaa kohteita virheellisesti.

5 Tulosten tarkastelu

5.1 Mallin ositustarkkuus

Tässä tutkielmassa mitattu historiallisten peruskarttojen kohteiden sijaintitarkkuus vastasi varsin hyvin ainoaa löytämääni Suomen historiallisia peruskarttoja karttaprosessoivaa julkaisua (Mäyrä ym., 2023). Mäyrä ja kumppanit (2023) käyttivät historiallisista peruskartoista kiinteistörajoja ja uusina referenssiaineistoina metsikkötietoja. He mittasivat vuoden 1965 karttojen sijaintivirheen keskineliön juureksi 12 m ja 1980-luvun kartoilla 10,8 m, kun tässä tutkielmassa vanhempien, 1950-1961 valmistuneiden peruskarttojen sijaintivirheen keskineliön juuri oli 12,8 m ja uudempien, vuoteen 1978 mennessä valmistuneiden 8,9 m. Erityyppisten kohteiden sijaintitarkkuus vaikuttaisi siis olevan suurinpiirtein samankaltaista historiallisten peruskarttojen erityyppisten kohteiden välillä.

Tässä tutkielmassa saavutettu ositustarkkuus vastasi läheisesti Mäyrän ja kumppaneiden (2023) tuloksia historiallisilla peruskartoilla, ja samankaltaisella U-net arkkitehtuurilla. Heidän tutkimuksensa tosin rajoittui paljon pienemmälle tutkimusalalle, vain yhdeksään karttalehteen kahtena eri aikana. Tässä tutkielmassa peltojen keskimääräinen LYS oli 0,94 kun se Mäyrän ja kumppaneiden (2023) tuloksissa oli 0,95. Suoluokkien yhteinen LYS oli tässä tutkielmassa 0,81, kun se Mäyrällä ja kumppaneilla (2023) oli 0,80. Mäyrän ja kumppaneiden tuottajan ja käyttäjän tarkkuudet olivat hyvin samankaltaisia pelloille, mutta suoalueitten tarkkuus oli tässä tutkielmassa useamman prosenttiyksikön verran parempi. Tulokset myös ovat joko hieman parempia tai vastaavia kuin muille karttasarjoille tehdyissä syväoppivissa soiden karttaprosessoinneissa, joskin niissä käytettyjen karttojen graafinen laatu oli myös heikompi (Ståhl & Weimann, 2022; Wu ym., 2022). Tiedossani ei ole klassisen karttaprosessoinnin menetelmiä, jotka olisivat keskittyneet tämän tutkielman luokkiin. Kuitenkin niiden yksiväristen luokkien kohdalla, jotka eivät jaa väriään muiden luokkien kanssa tarkkuus on kuitenkin ollut tämän tutkielman peltojen tarkkuuteen verrattavissa tai hieman alhaisempi (Leyk & Boesch, 2009; Herrault ym., 2013; Ostafin ym., 2017). Mielenkiintoinen vertailukohta ovat myös tässä tutkielmassa tarkastellut ojitetut suot ja Mäyrän ja kumppaneiden (2023) osittamat ojat ja purot. Ojat ja purot ovat määritelmällisesti kapeita, minkä seurauksena niitä on vaikeampi tunnistaa kuin laaja-alaisempia kohteita, ja niiden ositustarkkuus onkin heikointa Mäyrän ja kumppaneiden (2023) osittamista luokista. Visuaalinen tarkastelu kuitenkin paljastaa, että Mäyrän ja kumppaneiden (2023) malli ei kuitenkaan kadota juurikaan oja, vaan virhe syntyy lähinnä reunojen tarkasta sijainnista

(Mäyrä ym., 2023). Tässä tutkielmassa kapeiden kohteiden haasteita haluttiin välttää opettamalla malli tunnistamaan ojitetut suoalueet 50 metrin etäisyydelle ojista. Näin saavutettiin parempi tarkkuus, mutta kummatkin tavat ovat selkeästi käyttökelpoisia moneen tarkoitukseen.

Toinen mielenkiintoinen tarkkuuteen liittyvä vertailukohta on se, että Mäyrä ja kumppanit (2023) olivat korjanneet karttalehtien valkotasapainoa, käyttäneet esiopetettua luokittelijaa ja lisänneet opetusaineistoa keinotekoisesti. Näistä etenkin esiopetusta ja lisäämistä pidetään yleisesti syväoppimismallien tarkkuutta parantavina keinoina (Bansal ym., 2022). Tässä tutkielmassa malli saavutti kuitenkin hyvän tarkkuuden ilman esiopetusta ja koulutusaineiston lisääminen todettiin tässä tapauksessa tarpeettomaksi, sillä se hidasti mallin oppimista merkittävästi, ja teki kouluttamisesta epävakaampaa. Tämän tutkielman suhteellisen hyvään laatuun vaikutti myös ainakin se, että tuotettu opetusaineisto oli pinta-alaltaan ja pikselimäärältään 2,4-kertainen Mäyrään ja kumppaneihin (2023) verrattuna. Yhteenvetona osituksen laadusta tässä tutkielmassa voidaan todeta, että tuotettu aineisto joko vastaa tai ylittää aiemman suomalaisille peruskartoille tehdyn karttaprosessoinnin laadun, mutta kattaa paljon aiempaa suuremman alan, liki koko Suomen.

5.2 Tuotetun aineiston käyttökelpoisuus

Kuten luvuissa 4.4 ja 5.1 todettiin, tuotettu aineisto vastaa varsin tarkasti historiallisten peruskarttojen sisältöä. Tässä luvussa puolestaan tarkastellaan, millaisia hyödyllisiä tai toisaalta epätoivottuja ominaisuuksia karttaprosessoiduilla skannatuilla peruskartoilla on maankäytön muutoksien tutkimuskäytössä. Ensinnäkin historiallisten peruskarttojen määritelmät pelloille, soille ja ojille vastaavat joko täysin tai hyvin pitkälle niiden nykyistä vastinetta eli maastotietokantaa. Tämä tarkoittaa että peruskarttoja on varsin mielekästä verrata parhaaseen nykyiseen maankäyttötietoon.

Tutkielman tutkimusajanjakso eli 1949–1978 sijoittuu yhteen Suomen maankäytön muutosten kiihkeimpään ajanjaksoon (Voutilainen ym., 2012; Korhonen ym., 2017). Etenkin ne karttalehdet, jotka ehdittiin päivittää tutkimusajanjaksona, tarjoavat mahdollisuuden tarkastella tuon ajan nopeita muutoksia (Kuva 13). Niiden pinta-ala kattaa kuitenkin vain noin 14 % kaikkien tutkielmassa käsiteltyjen skannattujen karttalehtien pinta-alasta, ja on lisäksi keskittynyt vahvasti Turun, Helsingin, Tampereen ja Rovaniemen seuduille sekä hajanaisesti Pohjanmaalle (Kts. Kuva 6A). Tämän lisäksi valtaosa kartoista on tuotettu 1960-luvun loppupuolella tai 1970-luvulla, jolloin peruskartoista ei ole apua nopeita muutoksia edeltävän

ajanjakson maankäytön selvittämiseen. Nyt prosessoitujen karttojen ja ajantasaisen maastotietokannan väliin jää usean vuosikymmenen aukko, jonka aikana ehtii tapahtua monenlaisia muutoksia. Esimerkiksi ymmärrys luvussa 4.3 ja kuvassa 13 esitettyjä Haaponevan maankäytön muutoksista jää vajaaksi, kun käytetään pelkästään modernia maastotietokantaa ja ennen vuotta 1979 tehtyjä peruskarttoja alueelta. Alue oli nimittäin vuosina 1987–2018 turpeentuotantokäytössä (Orjasniemi & Alapuranen, 2023). Pelkästään edellämainittujen aineistojen perusteella olisi ollut helppoa tehdä virheellinen johtopäätös, että suo on kuivattu suoraan peltokäyttöön. Myös vuoden 1978 jälkeen valmistetut peruskartat olisi tärkeä saada karttaprosessoitua tulevaisuudessa, jotta aikaresoluutiota saataisiin parannettua. Historialliset maastotietokannat ovat saatavilla vuodesta 2005 lähtien (Historialliset Maastotietokannat saatavilla nyt yhtenäisinä versioina, 2023), ja vuosien 1978–2005 välille mahtuu noin kahdesta kolmeen uudelleenkartoitusta. Lisäksi tulevissa peruskarttojen karttaprosessoinneissa olisi syytä laajentaa maankäyttöluokkien määrää. Tässä tutkielmassa tunnistettiin vain peltoja, soita ja epäsuorasti oja. Aiemmin peruskartoista on ositettu näiden lisäksi vielä teitä ja vesistöjä (Mäyrä ym., 2023). Tämä kuitenkin tarkoittaa, että monet aspektit peruskarttojen rikkaasta tietosisällöstä ovat täysin karttaprosessoimatta.

Peruskarttalehtien yhtenä laaja-alaisen hyödyntämisen lisähaasteena on myös se, että ne eivät tarjoa yhtenäistä aikakerrosta, vaan jokaisella karttalehdellä on oma maastokartoitusvuotensa. Koska uudelleenkartoituksia tehtiin tyypillisesti korkeintaan kerran kymmeneen vuoteen (Hirva ym., 1975), täytyy aineistoa käyttäessä luopua joko karttalehtien samanaikaisuudesta tai valtakunnallisesta alueellisesta kattavuudesta. Maastokartoituksessa edettiin alueittain, minkä takia karttalehdet muodostavat aikamosaiikin, jossa lähellä toisiaan olevat karttalehdet ovat yleensä myös suurin piirtein samanaikaisia (Kuva 6). Näin ollen peruskartat saattavatkin sopia valtakunnallista tasoa paremmin esimerkiksi maakuntatasolle tai sitä suurimittakaavaisempaan tarkasteluun. Valtion metsien inventoinnit tarjoavat tässä mielenkiintoisen vertailukohdan valtakunnalliseen tarkasteluun, sillä niitä on tehty useammin, vain viidestä kymmeneen vuoden välein, ja ne sisältävät myös tietoa soiden ojituksesta (Metsävarat maakunnittain, 2023). Ojitettujen soiden osuus on sekä tämän tutkielman uudemmilla päivitetyillä tai kertaalleen kartoitetuilla karttalehdillä, että valtakunnan metsien viidennessä inventoinnissa (VMI 5 1964–1970) noin hieman vajaa kolmannes (‘Valtakunnan metsien inventointi 5’, 1970). Vaikka VMI:en aikaresoluutio onkin parempi, niiden tulokset on tilastollisesti otostettu vain maakunnittain. Peruskarttojen ja VMI:n yhteiskäyttö voikin johtaa parempaan ajallisten ja tilallisten muutosten holistiseen ymmärtämiseen.

5.3 Syväoppivan karttaprosessoinnin haasteet ja mahdollisuudet

Syväoppivat kuvantunnistusmenetelmät ovat lyöneet läpi karttaprosessoinnissa viimeisen neljän vuoden aikana. Tämän syinä voidaan nähdä näiden menetelmien kyky oppia tunnistamaan monimutkaisia symboleita ja tekstuureita sekä samanvärisiä kohteita toisistaan suurella tarkkuudella (esim. Ekim ym., 2021; Uhl ym., 2022; Wu ym., 2022; Mäyrä ym., 2023), kuten myös tässä tutkielmassa havaittiin empiirisesti. Syväoppimisen suurena etuna on myös se, että jokaiselle tunnistettavalle kohteelle ei tarvitse räätälöidä tunnistusalgoritmia erikseen, vaan syväoppivat kuvantunnistussmallit ovat käytettävissä lähtökohtaisesti mihin tahansa kuvatyyppiin. Tämän suuren potentiaalın hyödyntäminen ei kuitenkaan ole vailla haasteita.

Ensinnäkin syväoppiminen vaatii laajoja etiketöityjä aineistoja oppiakseen hyvin (Bansal ym., 2022). Tällaisten kartta-aineistojen tuottaminen käsin asiantuntijatyönä on hidasta (Leyk & Boesch, 2009; Chiang ym., 2013), kuten myös tässä tutkielmassa havaittiin. Ratkaisuja tähän voivat olla mapathonit ja muu vapaaehtoisten hyödyntäminen (Koskinen ym., 2019; Sobotkova ym., 2023). Toisaalta myös puoliautomaattinen opetusaineiston tuottaminen tai lisääminen voivat olla ratkaisu datan vähyyteen. Se voi perustua esimerkiksi kartoitettavan kohdeluokan tunnettuihin nykysijainteihin (Uhl ym., 2020), tai synteettisen, alkuperäistä muistuttavan aineiston tuotantoon (Bansal ym., 2022). Kummankin tuovat kuitenkin myös uusia haasteita. Mapathonit vaativat usein markkinointia vapaaehtoisten saamiseksi, laitteita ja ohjelmistoja joilla kartoittaa ja osallistujien koulutusta sekä aihepiiriin, että käytettävään tekniikkaan (Koskinen ym., 2019; Sobotkova ym., 2023).

Myös aineiston laadulla on väliä. Nykyisten aineistojen hyödyntäminen automatisoituun aineistonkeräämisen, esimerkiksi keräämällä esimerkkejä asutuksesta nykyisten asutusten seudulta, kärsii puolestaan karttojen iän myötä kasvavista tarkkuusongelmista (Uhl ym., 2020). Synteettisesti uusimmillakin tekniikoilla tuotetut aineistot taas ovat usein vähemmän monipuolisia kuin aidot aineistot, joten nekään eivät ole täydellinen ratkaisu ongelmaan (Hataya ym., 2023). Asiantuntijan käsin piirtämää digitointia pidetään yleisesti korkealaatuisimpana vektorointitapana ja useissa tutkimuksessa siihen suhtaudutaan implisiittisesti kuin se olisi virheetöntä (Wu ym., 2022; Mäyrä ym., 2023; Sobotkova ym., 2023). Tämä ei kuitenkaan ole täysin ongelmaton oletamus. Tässä tutkielmassa noin kahdeksasta prosentista käsinpiirretyistä testausetiketeistä löytyi virheitä, jotka paljastuivat kun niitä verrattiin koulutetun mallin ennusteisiin. Käytännössä ne olivat lähinnä puuttuvia

pieniä kohteita. Myös muissa tutkimuksissa, joissa asiantuntijan tekemää digitointia on arvioitu, asiantuntijoiden työ on havaittu korkealaatuiseksi, muttei virheettömäksi (Koskinen ym., 2019; Maduekwe, 2021). Hyvin tarkasti toimivien mallien kanssa pienetkin virheet testiaineistossa vaikuttavat arvioituun tarkkuuteen heikentävästi. Täten myös asiantuntijan digitoinnin tarkastelu ja validointi on myös tärkeää opetusaineiston laadun varmistamiseksi. Tässä tutkielmassa käsin tehtyjen piirrosten iteratiivinen uudelleentarkastelu osoittautui toimivaksi laadunparannusmenetelmäksi.

Toiseksi syväoppivien menetelmien käyttö koodaamalla vaatii toistaiseksi huomattavan määrän teknistä osaamista ja ymmärrystä, sekä joko kalliita näytönohjaimia tai maksullisia pilvipalveluita. Viime aikoina saataville on kuitenkin tullut helppokäyttöisempää työkaluja, joiden tarkoituksena on madaltaa kynnystä syväoppimisen hyödyntämiseen geospaatialisissa tehtävissä, kuten OTBTF, tietyt QGIS-pluginit tai ArcGIS:in syväoppimistyökalut (Cresson, 2018; Deep learning in ArcGIS Pro—ArcGIS Pro | Documentation, 2024; Home — QGIS: Deepness: Deep Neural Remote Sensing 0.6.3 documentation, 2024). Nämä eivät kuitenkaan täysin poista tarvetta ymmärtää mallien toiminnan peruseriaatteita saatikka tarvetta laskentateholle.

Kolmanneksi on pidettävä mielessä, että tarkatkin koneoppimisjärjestelmät tekevät virheitä. Osa virheistä saattaa olla samantyyppisiä joita ihmisellekin käy, kuten pienien kohteiden huomaamatta jääminen (Sobotkova ym., 2023). Toisaalta kone voi myös tuottaa väärää positiivisia, jotka näyttävät ihmissilmään varsin älyttömiltä, kuten tässäkin tutkielmassa huomattiin. Tämän takia malleja tulee testata huolellisesti ja sekä numeerisesti, että laadullisesti. Yhteenvetona voidaan todeta, että karttaprosessointi on mullistunut syväoppivien menetelmien ansiosta, minkä ansiosta historiallisia karttoja on paljon helpompaa muuttaa nykypäivän aineistojen kanssa vertailtavissa oleviksi, korkealaatuiseksi paikkatietoaineistoiksi.

6 Aineistojen saatavuus

Tutkielmassa tuotettu opetus- ja testiaineisto, alusta lähtien tässä tutkielmassa tuotetut georeferointitiedot, käytetty karttalehtijako ja luokittelun tulos koko maasta tullaan lataamaan UTU Geonode –palveluun geonode.utu.fi . Luokittelusta julkaistaan sekä korkeusfilteröity, että -filteröimätön versio.

Tutkielmassa käytetty koodi löytyy Githubista: <https://github.com/iioseppa/finnish-historical-base-map-processing>

Peruskartat ovat saatavilla skannattuina valokuvina Maanmittauslaitoksen latauspalvelussa <https://vanhatpainenutkartat.maanmittauslaitos.fi/>

Kiitokset

Kiitän Mikko Kutilaista (Shingle oy) peruskarttojen georeferointitiedoista.

Kiitokset myös ohjaajilleni neuvoista ja ideoista.

Lähteet

Abadi, M., Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, & Xiaoqiang Zheng (2015) 'TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems'.
<https://www.tensorflow.org/>. 11.1.2023

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V. & Nahavandi, S. (2021) 'A review of uncertainty quantification in deep learning: Techniques, applications and challenges', *Information Fusion*, 76, 243–297. <https://doi.org/10.1016/j.inffus.2021.05.008>.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S. & Asari, V.K. (2019) 'A State-of-the-Art Survey on Deep Learning Theory and Architectures', *Electronics*, 8(3),s. 292.
<https://doi.org/10.3390/electronics8030292>.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. & Farhan, L. (2021) 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', *Journal of Big Data*, 8(1), 1–74.
<https://doi.org/10.1186/s40537-021-00444-8>.

Bahgat, K. & Runfola, D. (2021) 'Toponym-assisted map georeferencing: Evaluating the use of toponyms for the digitization of map collections', *PLOS ONE*, 16(11),s. e0260039.
<https://doi.org/10.1371/journal.pone.0260039>.

Banerjee, M., Capozzoli, M., McSweeney, L. & Sinha, D. (1999) 'Beyond kappa: A review of interrater agreement measures', *The Canadian Journal of Statistics*, 27(1).
<https://doi.org/10.2307/3315487>.

Bansal, Ms.A., Sharma, Dr.R. & Kathuria, Dr.M. (2022) 'A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications', *ACM Computing Surveys*, 54(10s), 208, 1-29. <https://doi.org/10.1145/3502287>.

Barnosky, A.D., Hadly, E.A., Bascompte, J., Berlow, E.L., Brown, J.H., Fortelius, M., Getz, W.M., Harte, J., Hastings, A., Marquet, P.A., Martinez, N.D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J.W., Gillespie, R., Kitzes, J., Marshall, C., Matzke, N., Mindell, D.P., Revilla, E. & Smith, A.B. (2012) 'Approaching a state shift in Earth's biosphere', *Nature*, 486(7401), 52–58. <https://doi.org/10.1038/nature11018>.

Bhattacharjee, J., Marttila, H., Haghghi, A.T., Saarimaa, M., Tolvanen, A., Lepistö, A., Futter, M.N. & Kløve, B. (2021) 'Development of Aerial Photos and LIDAR Data Approaches to Map Spatial and Temporal Evolution of Ditch Networks in Peat-Dominated Catchments', *Journal of Irrigation and Drainage Engineering*, 147(4), s. 04021006.
[https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001547](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001547).

Bin, D. & Cheong, W.K. (1998) 'A system for automatic extraction of road network from maps', in *Proceedings. IEEE International Joint Symposia on Intelligence and Systems (Cat.*

No.98EX174). *Proceedings. IEEE International Joint Symposia on Intelligence and Systems (Cat. No.98EX174)*, 359–366. <https://doi.org/10.1109/IJSIS.1998.685476>.

Boyle, R. (1980) 'Scan digitization of cartographic data' *Teoksessa Herber Freeman & Goffredo Pieroni 1980 (toim.) Map Data Processing*. London: Academic Press.

Burt, J.E., White, J., Allord, G., Then, K.M. & Zhu, A.-X. (2020) 'Automated and semi-automated map georeferencing', *Cartography and Geographic Information Science*, 47(1), 46–66. <https://doi.org/10.1080/15230406.2019.1604161>.

Chen, J., Luo, X., Zhu, L., Zhang, Q. & Gan, Y. (2023) 'Handwritten CAPTCHA recognizer: a text CAPTCHA breaking method based on style transfer network', *Multimedia Tools and Applications*, 82(9), 13025–13043. <https://doi.org/10.1007/s11042-021-11485-9>.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018) 'Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation', in V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (eds.) *Computer Vision – ECCV 2018*. Cham: Springer International Publishing (Lecture Notes in Computer Science), 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.

Chiang, Y.-Y., Leyk, S. & Knoblock, C.A. (2013) 'Efficient and Robust Graphics Recognition from Historical Maps', in Y.-B. Kwon & J.-M. Ogier (eds.) *Graphics Recognition. New Trends and Challenges*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), 25–35. https://doi.org/10.1007/978-3-642-36824-0_3.

Chiang, Y.-Y., Leyk, S. & Knoblock, C.A. (2014) 'A Survey of Digital Map Processing Techniques', *ACM Computing Surveys*, 47(1), 1, 1-44. <https://doi.org/10.1145/2557423>.

Chollet, F. & ym. (2015) 'Keras'. <https://keras.io>. 1.11.2023

Cohen, J. (1960) 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, 20(1).

Cresson, R. (2018) 'A framework for remote sensing images processing using deep learning techniques', *IEEE Geoscience and Remote Sensing Letters*, 16(1), 25–29.

Deep learning in ArcGIS Pro—ArcGIS Pro | Documentation (2024). <https://pro.arcgis.com/en/pro-app/latest/help/analysis/deep-learning/deep-learning-in-arcgis-pro.htm> 25.4.2024.

DeVries, T. & Taylor, G.W. (2018) 'Learning Confidence for Out-of-Distribution Detection in Neural Networks'. arXiv. <https://doi.org/10.48550/arXiv.1802.04865>.

Dhar, D.B. & Chanda, B. (2006) 'Extraction and recognition of geographical features from paper maps', *International Journal of Document Analysis and Recognition (IJDAR)*, 8(4), 232–245. <https://doi.org/10.1007/s10032-005-0010-9>.

Dong, L., Zheng, F., Chang, H. & Qin, Y. (2018) 'Corner points localization in electronic topographic maps with deep neural networks', *Earth Science Informatics*, 11(1), 47–57. <https://doi.org/10.1007/s12145-017-0317-3>.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2021) 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale'. arXiv. <https://doi.org/10.48550/arXiv.2010.11929>.

Duan, W., Chiang, Y.-Y., Leyk, S., Uhl, J.H. & Knoblock, C.A. (2020) 'Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning', *International Journal of Geographical Information Science*, 34(4), 824–849. <https://doi.org/10.1080/13658816.2019.1698742>.

Dung, L. (2023) 'Current cases of AI misalignment and their implications for future risks', *Synthese*, 202(5),s. 138. <https://doi.org/10.1007/s11229-023-04367-0>.

Ekim, B., Sertel, E. & Kabadayı, M.E. (2021) 'Automatic Road Extraction from Historical Maps Using Deep Learning Techniques: A Regional Case Study of Turkey in a German World War II Map', *ISPRS International Journal of Geo-Information*, 10(8),s. 492. <https://doi.org/10.3390/ijgi10080492>.

Finér, L., Lepistö, A., Karlsson, K., Räike, A., Tattari, S., Huttunen, M., Härkönen, L., Joensuu, S., Kortelainen, P., Mattsson, T., Piirainen, S., Sarkkola, S., Sallantausta, T. & Ukonmaanaho, L. (2020) *Metsistä ja soilta tuleva vesistökuormitus 2020*. valtioneuvoston kanslia. <https://julkaisut.valtioneuvosto.fi/handle/10024/162009>

Fischer, M. (2020) 'Multiclass_losses.py'. https://github.com/maxvfischer/keras-image-segmentation-loss-functions/blob/master/losses/multiclass_losses.py. 31.10.2023

Frajer, J. & Fiedor, D. (2021) 'A historical curiosity or a source of accurate spatial information on historical land use? The issue of accuracy of old cadastres in the example of Josephian Cadastre from the Habsburg Empire', *Land Use Policy*, 100,s. 104937. <https://doi.org/10.1016/j.landusepol.2020.104937>.

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(2), 187–199. <https://doi.org/10.1002/arp.1807>.

Gillies, S. & et. al (2013) 'Rasterio: geospatial raster I/O for Python programmers'. Mapbox. <https://github.com/rasterio/rasterio>. 7.3.2024

Gimmi, U., Ginzler, C., Müller, M. & Psomas, A. (2016) 'Assessing accuracy of forest cover information on historical maps', *Prace Geograficzne*, 2016(Zeszyt 146), 7–18. <https://doi.org/10.4467/20833113PG.16.014.5544>.

Ginzler, C., Brändli, U.-B. & Hägeli, M. (2011) 'Waldflächenentwicklung der letzten 120 Jahre in der Schweiz', *Schweizerische Zeitschrift für Forstwesen*, 162(9), 337–343. <https://doi.org/10.3188/szf.2011.0337>.

Guo, M., Bei, W., Huang, Y., Chen, Z. & Zhao, X. (2021) 'Deep learning framework for geological symbol detection on geological maps', *Computers & Geosciences*, 157,s. 104943. <https://doi.org/10.1016/j.cageo.2021.104943>.

Halás, M., Klapka, P. & Erlebach, M. (2019) 'Unveiling spatial uncertainty: a method to evaluate the fuzzy nature of functional regions', *Regional Studies*, 53(7), 1029–1041. <https://doi.org/10.1080/00343404.2018.1537483>.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y. & Tao, D. (2023) 'A Survey on Vision Transformer', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>.

Harris, C.R., Millman, K.J., Walt, S.J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H. van, Brett, M., Haldane, A., Río, J.F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. & Oliphant, T.E. (2020) 'Array programming with NumPy', *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.

Hataya, R., Bao, H. & Arai, H. (2023) 'Will Large-scale Generative Models Corrupt Future Datasets?', in. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE Computer Society, 20498–20508. <https://doi.org/10.1109/ICCV51070.2023.01879>.

He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2020) 'Mask R-CNN', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>.

Heitzler, M., Gkonos, C., Tsorlini, A. & Hurni, L. (2018) 'A modular process to improve the georeferencing of the Siegfried map', *e-Perimetreon*, 13(2), 85–100.

Heitzler, M. & Hurni, L. (2020) 'Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map', *Transactions in GIS*, 24(2), 442–461. <https://doi.org/10.1111/tgis.12610>.

Herrault, P.-A., Sheeren, D., Fauvel, M. & Paegelow, M. (2013) 'Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIE Lab Color Space', in. *The 16th AGILE International conference on geographic information science*, Springer, s. 95. https://doi.org/10.1007/978-3-319-00615-4_6.

Hirva, K., Hämäläinen, P., Lukkarinen, S., Luoma, N., Lyytikäinen, H.E., Niemelä, O., Nikunen, M., Nummenmaa, M., Nurmi, V., Paavola, O., Raevaara, H., Ruotoistenmäki, M., Saarentaus, T., Tiainen, H., Vepsäläinen, S. & Viikko, L. (1975) *Suomen Peruskartoitus 1947–1975*. Maanmittauslaitoksen julkaisu 42.

Historialliset Maastotietokannat saatavilla nyt yhtenäisinä versioina | Maanmittauslaitos (2023). <https://www.maanmittauslaitos.fi/ajankohtaista/historialliset-maastotietokannat-saatavilla-nyt-yhtenaisina-versioina>. 22.4.2024.

'Historical Topographic Maps - Preserving the Past' (2024). <https://www.usgs.gov/programs/national-geospatial-program/historical-topographic-maps-preserving-past>. 19.4.2024

'Historische Karten' (2024). <https://www.swisstopo.admin.ch/de/historische-karten>. 26.4.2024

'Historiska kartor' (2024). <https://www.lantmateriet.se/sv/kartor/vara-karttjanster/Historiska-kartor/> 19.4.2024.

Home — QGIS: Deepness: Deep Neural Remote Sensing 0.6.3 documentation (2024). <https://qgis-plugin-deepness.readthedocs.io/en/latest/>. 25.4.2024).

Hotanen, J.-P., Maltamo, M. & Reinikainen, A. (2006) 'Canopy stratification in peatland forests in Finland', *Silva Fennica*, 40(1). <https://www.silvafennica.fi/article/352>

Hotanen, J.-P., Saarinen, M. & Nousiainen, H. (2015) 'Secondary succession of threeless and composite mire site types after drainage', *Suo - Mires and peat*, 66(1). <https://www.suo.fi/article/9896>

- Iosifescu, I., Tsorlini, A. & Hurni, L. (2016) 'Towards a comprehensive methodology for automatic vectorization of raster historical maps', *e-Perimetron*, 11(2), 57–76.
- Jana, S. & Mahanta, J. (2023) 'Boundary of a fuzzy set and its application in GIS: a review', *Artificial Intelligence Review*, 56(7), 6477–6507. <https://doi.org/10.1007/s10462-022-10331-0>.
- Ji, Y., Zhang, H., Zhang, Z. & Liu, M. (2021) 'CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances', *Information Sciences*, 546, 835–857. <https://doi.org/10.1016/j.ins.2020.09.003>.
- Jordahl, K., Bossche, J.V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A.G., Farmer, C., Hjelle, G.A., Snow, A.D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L.J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe & Leblanc, F. (2020) 'geopandas: v0.12.0'. Zenodo. <https://doi.org/10.5281/zenodo.3946761>. 7.3.2024.
- Kaakinen, E., Kokko, A., Aapala, K., Autio, O., Eurola, S., Hotanen, J.-P., Kondelin, H., Lindholm, T., Nousiainen, H., Rehell, S., Ruuhijärvi, R., Sallantausta, T., Salminen, P., Tahvanainen, T., Tuominen, S., Turunen, J., Vasander, H. & Virtanen, K. (2018) 'Suot', *Teoksessa Ojanen ym (2018): Suomen luontotyyppien uhanalaisuus 2018 : Luontotyyppien punainen kirja. Osa 1: Tulokset ja arvioinnin perusteet*. Ympäristöministeriö. <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161233/Suomen%20luontotyyppien%20uhanalaisuus%202018%20OSA1.pdf?sequence=1&isAllowed=y>.
- Kaim, D., Kozak, J., Ostafin, K., Dobosz, M., Ostapowicz, K., Kolecka, N. & Gimmi, U. (2014) 'Uncertainty in Historical Land-Use Reconstructions with Topographic Maps', *Quaestiones Geographicae*, 33(3), 55–63. <https://doi.org/10.2478/quageo-2014-0029>.
- Khotanzad, A. & Zink, E. (2003) 'Contour line and geographic feature extraction from USGS color topographical paper maps', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1), 18–31. <https://doi.org/10.1109/TPAMI.2003.1159943>.
- Korhonen, K.T., Ihalainen, A., Ahola, A., Heikkinen, J., Henttonen, H.M., Hotanen, J.-P., Nevalainen, S., Pitkänen, J., Strandström, M. & Viiri, H. (2017) *Suomen metsät 2009–2013 ja niiden kehitys 1921–2013*. Luonnonvarakeskus (Luke). <https://jukuri.luke.fi/handle/10024/540537>.
- Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., Pekkarinen, A. & Käyhkö, N. (2019) 'Participatory mapping of forest plantations with Open Foris and Google Earth Engine', *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 63–74. <https://doi.org/10.1016/j.isprsjprs.2018.12.011>.
- Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- Lång, K., Aro, L., Assmuth, A., Haltia, E., Hellsten, S., Larmola, T., Lempinen, H., Lindfors, L., Lohila, A., Miettinen, A., Minkkinen, K., Nieminen, M., Ollikainen, M., Ojanen, P., Sarkkola, S., Sorvali, J., Seppälä, J., Tolvanen, A., Vainio, A., Wall, A. & Vesala, T. (2022) *Turvemaiden käytön vaihtoehdot Hiilineutraalissa Suomessa*. 2.

<https://www.ilmastopaneeli.fi/wp-content/uploads/2022/04/ilmastopaneelin-raportti-2-2022-turvemaiden-kayton-vaihtoehdot-hiilineutraalissa-suomessa.pdf>.

Leyk, S. & Boesch, R. (2009) 'Extracting Composite Cartographic Area Features in Low-Quality Maps', *Cartography and Geographic Information Science*, 36(1), 71–79. <https://doi.org/10.1559/152304009787340115>.

Leyk, S., Boesch, R. & Weibel, R. (2006) 'Saliency and semantic processing: Extracting forest cover from historical topographic maps', *Pattern Recognition*, 39(5), 953–968. <https://doi.org/10.1016/j.patcog.2005.10.018>.

Leyk, S., Weibel, R. & Boesch, R. (2005) 'A Conceptual Framework for Uncertainty Investigation in Map-based Land Cover Change Modelling', *Transactions in GIS*, 9(3), 291–322. <https://doi.org/10.1111/j.1467-9671.2005.00220.x>.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J. & Zhou, B. (2023) 'Trustworthy AI: From Principles to Practices', *ACM Computing Surveys*, 55(9), 177, 1-46. <https://doi.org/10.1145/3555803>.

Li, H., Liu, J. & Zhou, X. (2018) 'Intelligent map reader: A framework for topographic map understanding with deep learning and gazetteer', *IEEE Access*, 6, 25363–25376.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L. & Dollár, P. (2015) 'Microsoft COCO: Common Objects in Context'. arXiv. <https://doi.org/10.48550/arXiv.1405.0312>.

Liu, T., Miao, Q., Tian, K., Song, J., Yang, Y. & Qi, Y. (2016) 'SCTMS: Superpixel based color topographic map segmentation method', *Journal of Visual Communication and Image Representation*, 35, 78–90. <https://doi.org/10.1016/j.jvcir.2015.12.004>.

Liu, T., Miao, Q., Xu, P., Tong, Y., Song, J., Xia, G., Yang, Y. & Zhai, X. (2016) 'A contour-line color layer separation algorithm based on fuzzy clustering and region growing', *Computers & Geosciences*, 88, 41–53. <https://doi.org/10.1016/j.cageo.2015.12.017>.

Liu, T., Xu, P. & Zhang, S. (2019) 'A review of recent advances in scanned topographic map processing', *Neurocomputing*, 328, 75–87. <https://doi.org/10.1016/j.neucom.2018.02.102>.

Maanmittauslaitos (2019) 'Korkeusmalli 10 m'. <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/korkeusmalli-10-m>. 12.12.2023

Maanmittauslaitos (2023) 'Hallintorajat, teemakartoille, ei merialueita 2023, 1:1 000 000'. <https://etsin.fairdata.fi/dataset/b7d422fa-bcfa-44f1-89aa-9e1971264e31>. 12.12.2023

Maanmittauslaitos (2023) 'Maastotietokanta'. <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/maastotietokanta>. 12.12.2023

Maanmittauslaitos (2023) 'Maanmittauslaitoksen ilmakekuva'. <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/aineistot-ja-rajapinnat/tuotekuvaukset/ilmakekuva>. 26.4.2024.

Maduekwe, N.I. (2021) 'A GIS-Based Methodology for Extracting Historical Land Cover Data from Topographical Maps: Illustration with the Nigerian Topographical Map Series', *KN - Journal of Cartography and Geographic Information*, 71(2), 105–120. <https://doi.org/10.1007/s42489-020-00070-z>.

'Mapy archiwalne Polski i Europy Środkowej' (2024). <http://igrek.amzp.pl/index.php> 26.4.2024.

Maxwell, A., Bester, M., Guillen, L., Ramezan, C., Carpinello, D., Fan, Y., Hartley, F., Maynard, S. & Pyron, J. (2020) 'Semantic Segmentation Deep Learning for Extracting Surface Mine Extents from Historic Topographic Maps', *Remote Sensing*, 12(24), s. 4145. <https://doi.org/10.3390/rs12244145>.

Maxwell, A.E., Warner, T.A. & Guillén, L.A. (2021) 'Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review', *Remote Sensing*, 13(13), s. 2450. <https://doi.org/10.3390/rs13132450>.

Mäyrä, J., Kivinen, S., Keski-Saari, S., Poikolainen, L. & Kumpula, T. (2023) 'Utilizing historical maps in identification of long-term land use and land cover changes', *Ambio*, 52(11), 1777–1792. <https://doi.org/10.1007/s13280-023-01838-z>.

Mello, C.A.B., Costa, D.C. & Santos, T.J. dos (2012) 'Automatic image segmentation of old topographic maps and floor plans', in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 132–137. <https://doi.org/10.1109/ICSMC.2012.6377689>.

Metsävarat maakunnittain (2023) *Luonnonvarakeskus*.

<https://www.luke.fi/fi/tilastot/metsavarat/metsavarat-maakunnittain-6>. 24.4.2024.

Milletari, F., Navab, N. & Ahmadi, S.-A. (2016) 'V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation', in *2016 Fourth International Conference on 3D Vision (3DV). 2016 Fourth International Conference on 3D Vision (3DV)*, 565–571. <https://doi.org/10.1109/3DV.2016.79>.

Myllys, M. (1996) 'Agriculture on peatlands', in *Teoksessa Vasander, H. (toim.) (1996) Peatlands in Finland*. Finnish Peatland Society, 64–71.

Oberholzer, M., Östreicher, M., Christen, H. & Brühlmann, M. (1996) 'Methods in quantitative image analysis', *Histochemistry and Cell Biology*, 105(5), 333–355. <https://doi.org/10.1007/BF01463655>.

Orjasniemi, T. & Alapuranen, M. (2023) 'Haaponevan turvetuotantoalueen ympäristöluvan rauettaminen ja lohkojen 1, 2, 6 ja lisäalueen 2B (lohko 7) jälkihoitotoimenpiteiden vahvistaminen', *Pohjois-Suomen Avin päätös nro 140/2023*.

Ostafin, K., Iwanowski, M., Kozak, J., Cacko, A., Gimmi, U., Kaim, D., Psomas, A., Ginzler, C. & Ostapowicz, K. (2017) 'Forest cover mask from historical topographic maps based on image processing', *Geoscience Data Journal*, 4(1), 29–39. <https://doi.org/10.1002/gdj3.46>.

Ostafin, K., Pietrzak, M. & Kaim, D. (2021) 'Impact of the Cartographer's Position and Topographic Accessibility on the Accuracy of Historical Land Use Information: Case of the Second Military Survey Maps of the Habsburg Empire', *ISPRS International Journal of Geo-Information*, 10(12), s. 820. <https://doi.org/10.3390/ijgi10120820>.

Pavelková, R., Frajer, J., Havlíček, M., Netopil, P., Rozkošný, M., David, V., Dzuráková, M. & Šarapatka, B. (2016) 'Historical ponds of the Czech Republic: an example of the interpretation of historic maps', *Journal of Maps*, 12(sup1), 551–559. <https://doi.org/10.1080/17445647.2016.1203830>.

'Peruskartta 1:10 000 /1:20 000.' (1960). Peruskarttatoimikunta. Had:2 Kuvausohjeet 1960-1987, Maanmittaushallituksen II arkisto, Kansallisarkisto.

'Peruskartta 1:20 000.' (1948). Peruskarttatoimikunta. Had:2 Kuvausohjeet 1948-1987, Maanmittaushallituksen II arkisto, Kansallisarkisto.

Peruskartta 1:20 000, lehti 1034 12+2012 14 Dragsfjärd. Helsinki, 1968.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/1/1034/103414/103414_201214_1968.jpg. 24.8.2023

Peruskartta 1:20 000, lehti 113106 + 113103 Ketteli, Helsinki 1968.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/1/1131/113106/113106_113103_1968.jpg. 26.4.2024

Peruskartta 1:20 000, lehti 2042 01 Karisjärvi. Maanmittaushallitus, Helsinki 1958.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/2/2042/204201/204201_1958_ei_rajoja.jpg. 6.2.2024

Peruskartta 1:20 000, lehti 2333 10+331101 Vuorilahti. Helsinki 1966.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/3/3311/331101/233310-331101_1966.jpg. 24.8.2023

Peruskartta 1:20 000, lehti 2433 09 Ainali. Maanmittaushallitus, Helsinki 1955.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/2/2433/243309/243309_1955.jpg. 26.4.2024

Peruskartta 1:20 000, lehti 2433 09 Ainali. Maanmittaushallitus, Helsinki 1978.
https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/2/2433/243309/243309_1978.jpg. 26.4.2024.

Peruskartta 1:20 000, lehti 4112 12+4112 11+4114 03 Rautjärvi. Maanmittaushallitus, Helsinki 1971. https://vanhatpainetutkartat.maanmittauslaitos.fi/mml_vanhat_kartat/Painetut_kartat/21_Peruskartta_20k/4/4112/411212/411212_411211_411403_1971.jpg. 24.8.2023

Petipierre, R. (2021) *Neural networks for semantic segmentation of historical city maps: Cross-cultural performance and the impact of figurative diversity - ProQuest*.
<https://www.proquest.com/docview/2484420081?parentSessionId=MXhs6BJJoKdMs71s3q6ZJmCT8rYu57heEmzXmHZ7yRs%3D&pq-origsite=primo&accountid=14774> 16.5.2023.

Petitpierre, R. & Guhenec, P. (2023) 'Effective annotation for the automatic vectorization of cadastral maps', *Digital Scholarship in the Humanities*, s. fqad006.
<https://doi.org/10.1093/lc/fqad006>.

Pinheiro, P.H.O. & Collobert, R. (eds.) (2014) 'Recurrent Convolutional Neural Networks for Scene Labeling', *Proceedings of the 31st International Conference on Machine Learning*.

QGIS Development Team (2024) 'QGIS Geographic Information System'. QGIS Association.
<https://www.qgis.org>. 7.3.2024.

Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. (2019) 'Transfusion: Understanding Transfer Learning for Medical Imaging', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
https://proceedings.neurips.cc/paper_files/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html (Accessed: 11 August 2023).

Rajić, L. (2012) 'Toponyms and the Political and Ethnic Identity in Serbia', *Oslo Studies in Language*, 4(2). <https://doi.org/10.5617/osla.319>.

- Ratajczak, R., Crispim-Junior, C.F., Faure, E., Fervers, B. & Tougne, L. (2019) 'Automatic Land Cover Reconstruction From Historical Aerial Images: An Evaluation of Features Extraction and Classification Algorithms', *IEEE Transactions on Image Processing*, 28(7), 3357–3371. <https://doi.org/10.1109/TIP.2019.2896492>.
- Reiher, E., Li, Y., Delle Donne, V., Lalonde, M., Hayne, C. & Zhu, C. (1996) 'A system for efficient and robust map symbol recognition', in *Proceedings of 13th International Conference on Pattern Recognition. Proceedings of 13th International Conference on Pattern Recognition*, 783–787 vol.3. <https://doi.org/10.1109/ICPR.1996.547275>.
- 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. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Ronneberger, O., Fischer, P. & Brox, T. (2015) 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in N. Navab, J. Hornegger, W.M. Wells, & A.F. Frangi (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing (Lecture Notes in Computer Science), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Saeedimoghaddam, M. & Stepinski, T.F. (2020) 'Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks', *International Journal of Geographical Information Science*, 34(5), 947–968. <https://doi.org/10.1080/13658816.2019.1696968>.
- Schmidhuber, J. (2015) 'Deep learning in neural networks: An overview', *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
- Seidl, N.P. (2019) 'Engraved in the Landscape: The Study of Spatial and Temporal Characteristics of Field Names in the Changing Landscape', *Names*, 67(1), 16–29. <https://doi.org/10.1080/00277738.2017.1415539>.
- Sester, M. (2020) 'Cartographic generalization', *Journal of Spatial Information Science*, (21), 5–11.
- Shelhamer, E., Long, J. & Darrell, T. (2017) 'Fully Convolutional Networks for Semantic Segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- Sobotkova, A., Ross, S.A., Nassif-Haynes, C. & Ballsun-Stanton, B. (2023) 'Creating large, high-quality geospatial datasets from historical maps using novice volunteers', *Applied Geography*, 155, s. 102967. <https://doi.org/10.1016/j.apgeog.2023.102967>.
- Ståhl, N. & Weimann, L. (2022) 'Identifying wetland areas in historical maps using deep convolutional neural networks', *Ecological Informatics*, 68, s. 101557. <https://doi.org/10.1016/j.ecoinf.2022.101557>.
- Sun, J., Gallego-Sala, A. & Yu, Z. (2023) 'Topographic and climatic controls of peatland distribution on the Tibetan Plateau', *Scientific Reports*, 13(1), s. 14811. <https://doi.org/10.1038/s41598-023-39699-x>.
- Tariq, N., Hamzah, R.A., Ng, T.F., Wang, S.L. & Ibrahim, H. (2021) 'Quality Assessment Methods to Evaluate the Performance of Edge Detection Algorithms for Digital Image: A Systematic Literature Review', *IEEE Access*, 9, 87763–87776. <https://doi.org/10.1109/ACCESS.2021.3089210>.

Tesseract documentation (2023) Tesseract OCR. <https://tesseract-ocr.github.io/>. 18.8.2023.

Thung, K.-H. & Wee, C.-Y. (2018) 'A brief review on multi-task learning', *Multimedia Tools and Applications*, 77(22), 29705–29725. <https://doi.org/10.1007/s11042-018-6463-x>.

Tian, D., Han, Y., Wang, B., Guan, T., Gu, H. & Wei, W. (2021) 'Review of object instance segmentation based on deep learning', *Journal of Electronic Imaging*, 31(4), s. 041205. <https://doi.org/10.1117/1.JEI.31.4.041205>.

Toivonen, T., Herranen, T., Kivilompolo, J., Kujala, H., Laatikainen, M., Suomi, T. & Turunen, J. (2022) *GTK:n tutkimien soiden tutkimustilanne ja luonnontilaisuusluokitukset maakunnittain*. 40/2022. GTK, 40 s.

Tong, K., Wu, Y. & Zhou, F. (2020) 'Recent advances in small object detection based on deep learning: A review', *Image and Vision Computing*, 97, s. 103910. <https://doi.org/10.1016/j.imavis.2020.103910>.

Uhl, J.H., Leyk, S., Chiang, Y.-Y., Duan, W. & Knoblock, C.A. (2017) 'Extracting human settlement footprint from historical topographic map series using context-based machine learning', in *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*. *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*, 1–6. <https://doi.org/10.1049/cp.2017.0144>.

Uhl, J.H., Leyk, S., Chiang, Y.-Y., Duan, W. & Knoblock, C.A. (2020) 'Automated Extraction of Human Settlement Patterns From Historical Topographic Map Series Using Weakly Supervised Convolutional Neural Networks', *IEEE Access*, 8, 6978–6996. <https://doi.org/10.1109/ACCESS.2019.2963213>.

Uhl, J.H., Leyk, S., Chiang, Y.-Y. & Knoblock, C.A. (2022) 'Towards the automated large-scale reconstruction of past road networks from historical maps', *Computers, Environment and Urban Systems*, 94, s. 101794. <https://doi.org/10.1016/j.compenvurbsys.2022.101794>.

Vaianti, B., Petitpierre, R., Lenardo, I. di & Kaplan, F. (2023) 'Machine-Learning-Enhanced Procedural Modeling for 4D Historical Cities Reconstruction', *Remote Sensing*, 15(13), s. 3352. <https://doi.org/10.3390/rs15133352>.

'Valtakunnan metsien inventointi 5' (1970). https://statdb.luke.fi/PxWeb/pxweb/fi/LUKE/LUKE__04%20Metsa__06%20Metsavarat/1.04_Ojitustilanne_metsatalousmaalla.px/ 2.12.2023.

'Vanhat painetut kartat' (2024). <https://vanhatpainetutkartat.maanmittauslaitos.fi/> 26.4.2024.

Vasander, H. (2006) 'The use of mires for agriculture', in *Teoksessa Lindholm, T. ja Heikkilä, R. (toim.) (2006) Finland – land of mires*. Finnish Environment Institute (The Finnish Environment), 173–178.

Voutilainen, O., Wuori, O. & Muilu, T. (2012) *Eriytyvät alue- ja maatalouden rakenteet Suomessa maaseutunäkökulmasta*. 64. MTT. <https://jukuri.luke.fi/bitstream/handle/10024/438271/mtrraportti64.pdf?sequence=1&isAllowed=y>.

Vuorela, N., Alho, P. & Kalliola, R. (2002) 'Systematic Assessment of Maps as Source Information in Landscape-change Research', *Landscape Research*, 27(2), 141–166. <https://doi.org/10.1080/01426390220128631>.

Wang, Yanzhao, Sun, Y., Cao, X., Wang, Yihan, Zhang, W. & Cheng, X. (2023) 'A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing', *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 311–334. <https://doi.org/10.1016/j.isprsjprs.2023.11.014>.

Wise, S. (2002) 'Capturing Raster Data From Scanned Thematic Maps Using Desktop Graphics Software', *Transactions in GIS*, 6(3), 327–337. <https://doi.org/10.1111/1467-9671.00114>.

Wu, J., Yan, H. & Chalmers, A.N. (1994) 'Color image segmentation using fuzzy clustering and supervised learning', *Journal of Electronic Imaging*, 3(4), 397–403. <https://doi.org/10.1117/12.183755>.

Wu, S., Heitzler, M. & Hurni, L. (2022) 'Leveraging uncertainty estimation and spatial pyramid pooling for extracting hydrological features from scanned historical topographic maps', *GIScience & Remote Sensing*, 59(1), 200–214. <https://doi.org/10.1080/15481603.2021.2023840>.

Zhang, H., Zhou, X., Li, Huan, Zhu, G. & Li, Hongwei (2022) 'Machine Recognition of Map Point Symbols Based on YOLOv3 and Automatic Configuration Associated with POI', *ISPRS International Journal of Geo-Information*, 11(11), s. 540. <https://doi.org/10.3390/ijgi11110540>.

Zhao, Y., Tang, Y., Yu, Z., Li, H., Yang, B., Zhao, W., Li, F. & Li, Q. (2014) 'Holocene peatland initiation, lateral expansion, and carbon dynamics in the Zoige Basin of the eastern Tibetan Plateau', *The Holocene*, 24(9), 1137–1145. <https://doi.org/10.1177/0959683614538077>.

Zu Ermgassen, P.S.E., Spalding, M.D., Blake, B., Coen, L.D., Dumbauld, B., Geiger, S., Grabowski, J.H., Grizzle, R., Luckenbach, M., McGraw, K., Rodney, W., Ruesink, J.L., Powers, S.P. & Brumbaugh, R. (2012) 'Historical ecology with real numbers: past and present extent and biomass of an imperilled estuarine habitat', *Proceedings of the Royal Society B: Biological Sciences*, 279(1742), 3393–3400. <https://doi.org/10.1098/rspb.2012.0313>.

Liitteet

Liite 1. Taulukot

Liite 1.1: Sekaannusmatriisi käsin piirrettyjen testiaineiston ja mallin jälkikäsiteltyjen ennusteiden välillä. Yksikkönä on neliökilometri. Vaaleanpunaiset ruudut osoittavat virheelliset ennusteet, muut värit osoittavat onnistuneen luokittelun.

		Ennuste				Tuottajan tarkkuus
		Tausta	Pelto	Ojasuo	Ojaton suo	
Totuus	Tausta	59,500	0,087	0,311	0,435	98,62 %
	Pelto	0,498	9,097	0,007	0,010	94,65 %
	Ojasuo	0,369	0,004	5,138	0,102	91,5 %
	Ojaton suo	0,864	0,002	0,146	8,969	89,9 %
	Käyttäjän tarkkuus	97,2 %	99,0 %	91,7 %	94,3 %	Yhteensä 96,7 %

Liite 1.2: LYS:in keskiarvot eri kokoisille kohteille ja luokille

LYS:in keskiarvot	Pienet (< 32 ² px)	Keskisuuret (>32 ² px, <96 ² px)	Suuret > 96 ² px	Kaikki ilman tyhjiä	Kaikki tyhjien kanssa
Pellot	0,52	0,70	0,84	0,88	0,94
Ojitetut suot	0,19	0,55	0,78	0,66	0,85
Ojittamatomat suot	0,26	0,55	0,84	0,63	0,76