

Lectio praecursoria, Rintasyövän arviointi mammografia-aineistosta lääketieteellisen kuva-analyysin ja laskennan avulla

Antti Isosalo

Arvoisa kustos, arvoisa vastaväittäjä, arvoiset kuulijat

Aloitin tämän nyt tarkastettavan väitöstyön Teknologiaateollisuuden 100-vuotissäätiön ja Jane ja Aatos Erkon säätiön Oulun yliopistolle myöntämän rahoituksen turvin. Rahoitus myönnettiin erityisesti Tulvaisuuden tekijät -ohjelmasta, joka tukee ajankohtaisia ja yhteiskunnallisesti merkittäviä hankkeita. Yksi projektin tehtäväelementeistä liittyi rintasyöpään ja sen automaattiseen tunnistamiseen. Aihe kiehtoi minua, koska olin perusopinnoista alkaen ollut kiinnostunut aaltoliikkeestä ja optiikasta, matemaattisesta kuvankäsittelystä sekä konenäöstä.

Rintasyövän yleisyys ja ennuste

Rintasyöpä on maailman yleisimmin diagnosoitu syöpä. Uusia tapauksia todetaan vuosittain yli kaksi miljoonaa [1]. Suomessa todettiin 4 872 uutta rintasyöpää vuonna 2020 [2], ja ne muodostavat noin 30 % kaikista syöpätapauksista. Koko maailmassa uusia rintasyöpätapauksia vuonna 2020 oli arviolta noin 2,3 miljoonaa [1].

Rintasyövän hoitoennuste on hyvä. Suomessa sairastuneiden suhteellinen elossaololuku viisi vuotta syövän toteamisesta on Suomen syöpärekisterin [3] mukaan korkea, yli 90 %. Vaikka rintasyövän hoitoennuste on hyvä, kyse on silti vakavasta sairaudesta, jonka ennaltaehkäisy ja varhainen toteaminen on inhimillisestä ja yhteiskunnallisesta näkökulmasta tärkeää. Mikäli varhainen toteaminen johtaa kevyempiin hoitotoimiin, se auttaa osaltaan vapauttamaan terveydenhuollon resursseja.

Rintasyövän havaitseminen

Kansalliset rintasyöpäseulonnat ovat varhaisen toteamisen menetelmiä, joissa merkkejä rintasyövästä etsitään järjestelmällisesti väestöstä. Rintasyöpäseulonnoissa tutkimusmenetelmänä rintasyövän havaitsemiseksi käytetään mammografiaa, jossa tutkittavan rintakudoksesta muodostetaan kuvia röntgensäteilyn avulla. Vaikka kaikki rintasyövän ilmenemismuodot eivät ole näkyvissä mammografiakuvissa [4], se on ensisijainen keino rintojen kuvantamiseen. Lisäksi on kehitetty mm. röntgensäteitä hyödyntävä rinnan tomosynteesi, monileikekuvaus, jota käytetään tyypillisesti lisätutkimuksena silloin kun halutaan tarkempaa tietoa rinnan rakenteesta.

Rintojen anatomia ja rintakudoksen ulkonäkö ovat olennaisia muuttujia normaalin vaihtelun ja patologisen tilan erottamisessa [5]. Se mitä radiologit kuvista etsivät ei ole merkittävästi muuttunut 1960-luvulta nykypäivään siirryttäessä. Perinteisessä rintasyövän havainnoinnissa radiologi etsii mammografiakuvista pehmytkudosmassoja, pieniä kalkkihiukkasia ja rakennevääristymiä sekä antaa kuvien perusteella alustavan arvion. Löydösten visuaaliset piirteet, kuten massojen muoto ja reunat antavat viitteitä löydöksen laadusta. Hyvälaatuiset löydökset ovat usein säännöllisen muotoisia ja tarkkarajaisia. Epätarkasti rajautuvat löydökset voivat puolestaan viitata pahanlaatuisen löydöksen.

Rintasyövän tunnistaminen mammografiakuvista on erityinen tutkimusongelma, koska rintasyövän

Published under a CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

ilmenemismuotoja on monia. Osa ilmenemismuodoista on vaikeasti havaittavia jopa asiantuntijalle. Havainnoitavien kohteiden koot vaihtelevat suuresti, ja osa kohteista on alle millimetrin kokoisia. Tämä on yksi syy, miksi mammografiassa on korkeat vaatimukset kuvien erotuskyvylle.

Myös itse rintakudoksella on useita ilmenemismuotoja, ja rinta voi näyttäytyä kuvassa läpikuultavana, mutta myös lähes läpinäkymättömänä tiiviin kudoksen vuoksi. Jälkimmäisessä tapauksessa kasvainten havainnointi on vaikeaa. Edellä mainitut useat syövän ilmenemismuodot ja tiiviin kudoksen vaikutus kuvantamiselle tuovat haasteita sekä radiologin suorittamaan havainnointiin että tehtävän kouluttamiseen tekoälylle.

Lääketieteellinen kuva-analyysi

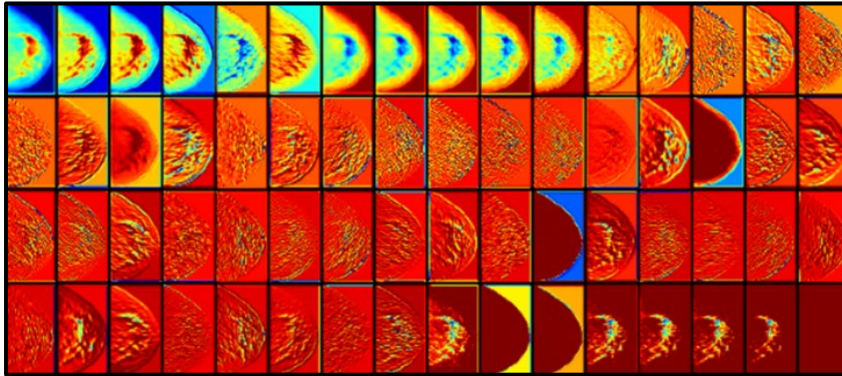
Miten rakentaa silta lääketieteen ja tietotekniikan välille? Väitöstyöni osalta vastaus löytyy nykyaikaisista lääketieteellisen kuva-analyysin menetelmistä. Lääketieteellinen kuva-analyysi ja laskenta yhdistävät eri tieteenaloja, kuten tietotekniikkaa, fysiikkaa ja matematiikkaa, sekä niiden teoreettisia ja soveltavia aloja ratkaistakseen lääketieteellisen kuvantamisen aihepiiristä ponnistavia tuotekehitys- ja tutkimusongelmia. Tähän väitöskirjaan liittyviin tietotekniikan soveltaviin aloihin lukeutuvat tietokonenäkö ja koneoppiminen, joiden menetelmiä käytetään oppimaan rintasyövän automaattiseen luokitteluun ja havaitsemiseen tarvittavat säännöt suoraan kuva-aineistosta.

Väitöstyössäni on keskitytty syväoppimiseen, koneoppimisen osa-alueeseen, ja siihen perustuvien mallien hyödyntämiseen mammografiakuvien sisältämän visuaalisen tiedon käsittelyssä. Syväoppimismallien avulla digitaalisesta kuva-aineistosta kyetään oppimaan merkityksellisiä piirteitä sekä niiden välisiä riippuvuussuhteita. Mallien opettamiseen tarvitaan yleensä laaja aineisto ja ennakkotietoa siitä, minkä tyyppiset mallit soveltuvat parhaiten sovelluskohteen ongelmiin.

Syväoppimismallit muuntavat syötteenä olevan opetusaineiston monitasoisiksi kuvauksiksi eli representaatioiksi. Kuva-aineiston tapauksessa representaatiot muodostuvat liu'uttamalla kuvan yli tietyille kuvan piirteille herkkää matriisia eli suodatinta. Menetelmän oppima tieto on sisällytettyä mallin sisäisiin painokertoimiin, jotka on optimoitu niin, että mallin tekemä virhe on mahdollisimman pieni.

Opetetun mallin painokertoimet ovat myös siirrettävissä toiseen käyttötarkoitukseen. Mammografiakuvissa esiintyville rintakudoksen hahmoille ja tekstuurille herkistyneen mallin painot voivat tuottaa suotuisan lähtötilanteen, mikäli esimerkiksi kärkeän pahanlaatuinen-hyvänlaatuinen-luokittelun sijaan halutaankin muokata mallia biologisen prosessin indikaattoreiden tarkempaan luokitteluun.

Syväoppimismenetelmien kehittämiseen on olemassa avoimen lähdekoodin ohjelmakirjastoja, aliohjelmakokoelmia, modulaarisesti käytettäväksi, ja näin ollen kaikkea ei tarvitse luoda tyhjästä.



Kuva 1. Konvoluutioneuroverkon piirrekarttoja. Konvoluutioneuroverkon piirrekartat ovat visuaalisia esityksiä sille, mitä neuroverkon suodattimet ovat kuvassa havainneet. Jokainen piirrekartta liittyy tiettyyn suodattimeen ja korostaa alueita, joissa tietyt piirteet esiintyvät. Konvoluutioneuroverkon syvempien kerrosten piirrekartat tunnistavat yhä monimutkaisempia rakenteita, kuten esimerkiksi epänormaalin vaihtelun rintakudoksessa.

Tutkimustehtävät ja aineisto

Väitöstyöni tavoitteena oli tutkia tekoälyn soveltamista rintasyövän automaattiseen arviointiin mammografiakuva-aineistosta. Ottaen huomioon, että vuosittain tehdään suuri määrä rintojen kuvantamistutkimuksia, on perusteltua kehittää menetelmiä kuvien tulkitsemiseen liittyvän taakan keventämiseksi.

Aihetta voidaan lähestyä suunnittelemalla menetelmiä rintasyövän seulontatutkimusten luokitteluun ja löydösten havaitsemiseen, sekä arvioimalla näiden menetelmien rajoituksia ja soveltuvuutta käyttötarkoitukseensa. Sen lisäksi, että kyetään antamaan luokitus kuvan tasolla, olisi hyödyllistä saada tietoa epänormaalin rintakudoksen tarkemmasta ilmiästä eli fenotyypistä.

Tällä hetkellä on rajallisesti ymmärrystä syväoppimismallien valikoivuudesta niiden nimenomaisille koulutusaineistoille. Lisäksi emme tiedä miten kattavasti erilaisia syövän alalajeja tarvitaan syväoppimismallien menestyksekkääseen hienosäätökoulutukseen, kun populaatio vaihtuu.

Väitöstyössä keskityttiin seuraaviin tutkimustehtäviin:

1. Ensimmäiseksi tutkittiin rintasyövän seulontatutkimusten luokitteluun kehitettyjen syväoppimismenetelmien yhteensopivuutta suomalaisten mammografiaseulontatulosten kanssa.
2. Toiseksi arvioitiin siirto-oppimisella saavutettavia etuja kehitettäessä menetelmiä massojen segmentointiin.
3. Kolmanneksi tutkittiin syväoppimisen soveltuvuutta rintakudoksen kuvantamisen avulla havaittavissa olevien ilmiäsuojen luokitteluun.

Myös vaihtoehtoiset laskentaratkaisut ovat mielenkiintoinen tutkimuskohde radiologian työnkulun tehostamisen näkökulmasta. Aineiston luontisijainnin lähellä tapahtuva tiedon käsittely, reunalaskenta edustaa uusien teknologioiden mahdollistamana siirtymää kohti hajautettua laskentaa [6].

4. Neljänneksi arvioitiin reunalaskennan soveltuvuutta lääketieteellisen kuvantamisen työnkulun muokkaamiseen.

Näihin tutkimuskysymyksiin vastaamiseksi oli käytettävissä kaksi kotimaista aineistoa, jotka kerättiin Oulun kaupungilta (35/2019) ja Pohjois-Pohjanmaan sairaanhoitopiiriltä (179/2019) saatujen rekisteritutkimuslupien mukaisesti.

Kansainvälisestäkin nähtynä laaja yli 200 000 suora-digitaalisten mammografiakuvan aineisto saatiin käyttöön keväällä 2020 ja sen annotointi saatiin valmiiksi tammikuussa 2021. Tämän lisäksi käytössä oli tunnettu portugalilainen seulontamammografiatutkimusaineisto [7], jota käytettiin menetelmien validointiin. Vuoden 2020 lopulla kerättiin vielä noin 700 tutkimuksen laajuinen tomosynteesimammografia-aineisto ja sen annotointi saatiin valmiiksi kevättalvella 2021.

”Tehtäväsi ei ole ennustaa tulevaisuutta, vaan mahdollistaa se”¹

Lähteet

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021 May;71(3):209-249. doi: 10.3322/caac.21660.

[2] Larønningen ja muut. NORDCAN: Cancer incidence, mortality, prevalence and survival in the Nordic countries. Version 9.2 (23.06.2022). Association of the Nordic Cancer Registries. Cancer Registry of Norway; 2022 [viitattu 3.8.2023]. Saatavilla: <https://nordcan.iarc.fr>

Vaikka väitöstyössäni on ensisijaisesti tutkittu ja kehitetty tekoälymenetelmiä, vähintään yhtä tärkeää on tuottaa tietoa, joka auttaa osaltaan varmistamaan, että terveydenhuolto ja sen järjestelmät tuottavat tekoälyajan vaatimuksiin sopivaa aineistoa. Riippumatta siitä, että nykyinen toimintamalli on perinteisen työkulun kannalta täysin toimiva, tämänhetkinen tekoälyteknologia hyötyisi siitä, että aineisto olisi nykyistä eheämpää eikä sisältäisi tulkinallisuuksia. Tämä helpottaisi myös tekoälyn siirtämistä osaksi sairaalan työnkulkua.

Toiveenani on, että tehty työ vaikuttaa rintasyövän havainnointimenetelmien kehityksen lisäksi myös siihen, että vielä nykyistä tekoälyvalmiimpaa aineistoa syntyisi sairaalan normaalien toimintojen ohessa ja työnkuvaa voitaisiin kehittää uusien teknologioiden avulla.

[3] Suomen syöpärekisteri. Suomen syöpärekisteri, datan päiväys 05.05.2019, sovelluksen versio 2019-06-11-002. Suomen syöpärekisteri; 2019 [viitattu 4.8.2023]. Saatavilla: <https://tilastot.syoparekisteri.fi/syovat>

[4] Tartter PI, Weiss S, Ahmed S, Kamath S, Hermann G, Drossman S. Mammographically Occult Breast Cancers. *Breast J.* 1999 Jan;5(1):22-25. doi: 10.1046/j.1524-4741.1999.005001022.x.

[5] Jesinger RA. Breast anatomy for the interventionalist. *Tech Vasc Interv Radiol.* 2014 Mar;17(1):3-9. doi: 10.1053/j.tvir.2013.12.002.

¹ Antoine de Saint-Exupéry

[6] Satyanarayanan M. The emergence of edge computing. *Computer*. 2017;50(1):30–39. doi: 10.1109/MC.2017.9

[7] Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012 Feb;19(2):236-48. doi: 10.1016/j.acra.2011.09.014.

Antti Isosalo. Medical image analysis and computing in breast cancer evaluation using mammography data. *Acta Universitatis Ouluensis D Medica* 1811. Oulu; Oulun yliopisto, Lääketieteellinen tiedekunta; 2024.

<https://urn.fi/URN:NBN:fi:oulu-202410106255>