

Digitaaliset kaksoet yksilöidyssä syövän mallintamisessa

Sara Seppälä

Luonnontieteiden kandidaatin tutkielma

Turun yliopisto

Biolääketieteen koulutusohjelma

Lääketieteellinen tiedekunta

Biolääketieteen laitos

04.05.2026

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

Kandidaatintutkielma

Tutkinto-ohjelma, oppiaine: Biolääketieteen koulutusohjelma

Tekijä: Sara Seppälä

Otsikko: Digitaaliset kaksoiset yksilöidyssä syövän mallintamisessa

Ohjaaja: Pekka Ruusuvuori, apulaisprofessori

Sivumäärä: 29 sivua

Päivämäärä: 04.05.2026

Digitaalisella kaksosella tarkoitetaan järjestelmää, joka kytkee yhteen fyysisen kappaleen ja sen virtuaalisesti luodun vastinparin. Alun perin termi luotiin insinöritieteiden tarpeisiin peilaamaan tarkasteltavaa kohdetta ja ennustamaan siinä tapahtuvia reaaliaikaisia muutoksia, mutta konseptilla on nähty laajoja sovellusmahdollisuuksia myös lääketieteessä. Peilaamalla potilasta virtuaalisesti olisi mahdollista seurata tämän voinnin ja sairauksien kehittymistä sekä reagoida muutoksiin yksilöllistetyillä hoitopäätöksillä.

Syövän hoitotaakka kasvaa globaalisti. Sairauden patogeneesi on monimutkainen, ja kehittyvien tutkimusmenetelmien kuten omiikka- ja kuvantamisteknologioiden myötä datan määrä, joka pystytään mittaamaan potilaasta, on kasvanut. Digitaalista kaksosta on ehdotettu keinoksi hallitsemaan ja hyödyntämään tätä lisääntyneitä datamääriä yksilöllisen syöpähoidon suunnittelussa. Kaksosella voisi ennustaa taudin etenemistä, vastetta eri hoitovaihtoehtoihin ja mahdollisten haittavaikutusten ilmenemistä.

Digitaaliset syöpäkaksoiset koostuvat valituista mallinnettavista syövän ominaisuuksista, virtuaalisesta laskentamallista sekä näiden välisestä yhteydestä. Tutkielmassa tarkastellaan sekä kuvantamis- että molekulaarisen profiilin dataa hyödyntäviä sovellutuksia ja eritellään ne laskentamalliensa mukaan joko mekanistisiin tai koneoppimisperusteisiin. Digitaalisissa syöpäkaksosissa fyysisen ja virtuaalisen todellisuuden välinen yhteys syntyy päivittämällä laskentamalli uusilla kliinisillä tutkimustuloksilla, sillä taudista on hankala saada päivityksiä reaaliaikaisilla antureilla.

Validoitua digitaalista syöpäkaksosta ei ole vielä saavutettu. Ratkaistavia haasteita on jokaisessa kaksosen osassa: fyysisen todellisuuden luotettavassa abstrahoinnissa, riittävän datamäärän hankinnassa virtuaaliseen kopioon sekä dynaamisen yhteyden saavuttamisessa. Ennen kaksosen kliinistä käyttöönottoa olisi myös ratkaistava yhteiskunnallisia kysymyksiä liittyen tietoturvaan ja tasa-arvoon. Selkeä määrittely digitaalisen kaksosen ja tavanomaisen mallinnuksen välillä on tehtävä, jotta vältetään käsitteiden sekoittuminen ja mahdollistetaan digitaalisen kaksosen erityispiirteiden tarkoituksenmukainen soveltaminen. Toteutuessaan digitaalinen syöpäkaksos olisi edistysaskel kohti entistä yksilöidympää ja kohdennetumpaa syöpähoitoa.

Avainsanat: digitaalinen kaksos, mallintaminen, yksilöllinen lääketiede, kuvantaminen, omiikkateknologiat

Lyhenneluettelo

PET	Positroniemissiotomografia
TT	Tietokonetomografia
MRI	Magneettikuvaus
IoT	Esineiden internet
DQL	Syvä Q-oppiminen
T1Gd	Gadoliniumtehosteinen, T1-painotteinen magneettikuvaus
NSCLC	Ei-pienisolainen keuhkosyöpä
WT	Wilmsin tuumori

Sisällys

1	Johdanto	1
2	Digitaalisen kaksosen määritelmä	3
2.1	Digitaaliset kaksoset lääketieteessä	4
2.2	Digitaaliset kaksoset yksilöidyssä lääketieteessä	5
2.3	Digitaalisten kaksosten soveltaminen yksilöidyssä syövän mallintamisessa	7
3	Digitaalisen syöpäkaksosen osat	9
3.1	Fyysinen todellisuus.....	9
3.2	Virtuaalinen todellisuus.....	10
3.3	Fyysisen ja virtuaalisen todellisuuden yhteys	11
4	Molekulaariseen ja biologiseen profiiliin perustuvat mallit.....	13
4.1	Metabolisten biomarkkereiden käyttö koneoppimis pohjaisissa syöpäennusteissa 14	
4.2	Solupopulaatiotasojen käyttö mekanistisessa haittavaikutusennusteessa	15
5	Kuvantamisdataan perustuvat kaksoset ja niiden sovellukset.....	17
5.1	MRI:n käyttö neoadjuvanttihoitovasteen mekanistisessa ennustamisessa	17
5.2	Optimaalisen hoitopolun määrittäminen syvällä Q-oppimisella	18
5.3	Syövän levinneisyyden ennustaminen mekanistisen ja tekoälypohjaisen mallin yhdistelmällä.....	20
6	Molekulaarisen profiloinnin ja kuvantamisen hypermalli	22
7	Digitaalisten kaksosten tulevaisuus ja haasteet	24
7.1	Fyysisen todellisuuden haasteet: ulottuvuus ja olettamukset	25
7.2	Virtuaalisen kaksosen haasteet: datan määrä, datan hankinta ja laskentatapa ..	25
7.3	Yhteiskunnalliset haasteet: käyttöönotto, tietoturva ja tasa-arvo	27
8	Yhteenveto ja johtopäätökset	28

1 Johdanto

Vuonna 2022 diagnosoitiin maailmanlaajuisesti lähes 20 miljoonaa uutta syöpätapausta sekä melkein kymmenen miljoonaa syövästä johtuvaa kuolemaa¹. Nykyennustusten mukaan jopa joka viides ihminen sairastuu syöpään elämänsä aikana, ja sairauden ilmaantuvuuden odotetaan kasvavan^{1,2}. Syöpä siis kasvattaa osuuttaan kansanterveysongelmien taakassa, mistä kasvaa tarve tarkemmalle ja tehokkaammalle diagnostiikalle sekä hoidolle¹.

Syövän diagnostiikassa ja hoidossa otetaan jatkuvasti uusia harppauksia. Teknologisiin edistykseen voidaan lukea muun muassa uudet ei-invasiiviset kuvantamismenetelmät, kuten positroniemissiotomografia (PET) ja tietokonetomografia (TT), sekä syövän molekulaaristen ominaisuuksien tarkempi tutkimus omiikkateknologioilla, kuten genomiikalla, transkriptomiikalla ja proteomiikalla. Vaikuttava hoito vaatii keinoja, joilla voidaan määrittää tarkasti potilaan syövän erityispiirteet, kuten tarkka lokaatio, levinneisyys- ja erilaistumisaste, mikroympäristön piirteet ja metaboliset muutokset, voidaksemme selittää yksilöiden väliset erot taudin käyttäytymisessä ja hoitotuloksissa.³

Uusien mittausteknologioiden myötä käytössämme on alati kasvava määrä dataa niin kuvantamisesta, kliinisistä laboratoriotestauksista kuin omiikkapaneeleistakin. Erilaisten laskentamallien odotetaan tuovan hyötyä näiden monimutkaisten datasettien navigointiin sekä sitä kautta diagnostiikan ja hoidon kehittämiseen³. Yksi mahdollisista keinoista hyödyntää laskentamalleja on digitaalinen kaksosen.

Digitaalinen kaksosen ideana syöpähoidossa on potilaan syövän mallintaminen digitaalisesti niin, että simuloitu malli on yhteydessä oikeaan syöpään ja peilaa siinä tapahtuvia muutoksia päivittymällä, kun uutta dataa on saatavilla. Malli perustuu jatkuvaan datan keräykseen kohteestaan sekä ennustavaan matemaattiseen malliin. Digitaalisen kaksosen avulla olisi mahdollista testata virtuaalisesti erilaisia hoitovaihtoehtoja, jolloin tietoa niiden tehosta ja haitoista saataisiin ilman potilaaseen kajoamista. Kaksosella voitaisiin simuloida potilaan syövän uniikkeja biologisia piirteitä, joiden avulla voitaisiin valita paras mahdollinen hoito.⁴

Tämän opinnäytetyön tarkoituksena on luoda katsaus lääketieteellisen digitaalisen kaksosen määritelmään sekä kartoittaa kaksosen luomia mahdollisuuksia syövän tutkimuksessa ja hoidossa. Työssä havainnollistetaan potilaan ja virtuaalisen todellisuuden yhteyteen pyrkiviä mallinnuksia esimerkkitapausten kautta. Koska yhtä oikeaa tapaa luoda digitaalinen kaksosen ei ole, tutkielmassa

jaotellaan olemassa olevat sovellukset niiden hyödyntämän syötteen mukaan molekulaarisbiologisiin sekä kuvantamisperusteisiin malleihin. Lisäksi työssä eritellään sovellukset niiden laskentatavan mukaan joko mekanistisiin tai koneoppimispohjaisiin. Opinnäytetyön tarkoitus on myös tarkastella kriittisesti konseptin rajoitteita lääketieteellisessä käytössä sekä pohtia digitaalisen kaksosen ja perinteisen mallintamisen välistä suhdetta.

2 Digitaalisen kaksosen määritelmä

Digitaalisen kaksosen konsepti on saanut alkunsa insinööritieteistä, jossa sitä käytetään kuvaamaan järjestelmää, jonka avulla arvioidaan virtuaalisesti fyysisen kappaleen vaatimia ylläpitotoimia ja optimointia. Alkuperäiseen määritelmän mukaan digitaalisen kaksosen tulisi sisältää kappaleen koko elinkaari sekä sen aikana tapahtuneet muutokset, mutta termin laajennuttua myös muiden tieteenalojen käyttöön sen reunaehdot voivat vaihdella⁵. Termiä on sovellettu laajasti aina öljynjalostuksen riskien tunnistuksesta kaupunkisuunnitteluun, ja termin yleistymisen takia on tullut tarpeelliseksi erotella digitaalisen kaksosen määritelmä perinteisestä mallinnuksesta⁶.

Alun perin avaruusteknologiseen käyttöön kehitellyn, *National Aeronautics and Space Administrationin* digitaalisen kaksosen määritelmä kuuluu: integroitu, usean tason multifyysinen simulaatio välineestä, joka peilaa fyysisistä vastinkappaleeltaan määrittääkseen lentoaluksen elinkaaren⁷. Validoituja mallinnuksia fyysisistä kappaleista on tehty jo kauan, mutta ero mallin ja digitaalisen kaksosen välillä on kaksosen ja sen fyysisen vastinkappaleen välinen yhteys. Digitaalisen kaksosen erikoisominaisuudeksi ja mallista erottavaksi tekijäksi voidaan katsoa sen jatkuva päivittyminen uudella datalla, jolloin virtuaalinen kaksosen ikään kuin peilaa fyysisessä kappaleessa tapahtuvia muutoksia. Insinööritieteiden piirissä, esimerkiksi lentoalusten komponenteissa, yhteys on saatu reaaliaikaiseksi kappaleeseen kytkettyjen antureiden avulla, jotka lähettävät dataa virtuaaliseen kaksoseen jatkuvalla syötöllä⁶.

Digitaalisen kaksosen voidaan siis katsoa koostuvan kolmesta komponentista: fyysisestä kappaleesta, sen virtuaalisesta kopiosta ja näiden ulottuvuuksien välisestä dynaamisesta yhteydestä. Digitaalisia kaksosia yhdistää myös niiden tarvitseman datan multimodaalisuus. Jotta fyysisestä kappaleesta ja sen tilasta saadaan kattavasti tietoa, kaksoseen syötettävät datamuodot ovat moninaisia ja tarvittavan datan määrä suurta⁶. Riittävä datamäärän saatavuus ja riittävä laskentateho ovatkin digitaalisen kaksosen yksiä toimintaa rajoittavia tekijöitä^{6,8}.

Yhtä oikeaa toteutusta digitaaliselle kaksoselle ei ole, vaan kaksosen laskentatapa, ulottuvuuksien välisen yhteyden nopeus, syötteen datamuodot ja fyysisen kappaleen mallintamisen yksityiskohtaisuus riippuvat kaksosen käyttötarkoituksesta. Jotta digitaalinen kaksosen olisi käyttökelpoinen, sen avulla mallinnettava ilmiö on tunnettava tarkasti, päivittyvien parametrien on kuvastettava kappaleen muuttuvaa tilaa ja ulottuvuuksien välisen yhteyden on oltava tarkoitukseen sopivan nopea, jotta kaksosta voitaisiin hyödyntää kappaletta koskevassa päätöksenteossa

muuttuvissa olosuhteissa⁶. Koska toteutustapoja on monia, digitaalista kaksosta ei pysty määrittelemään tietyllä teknisellä tavalla toimivaksi järjestelmäksi, vaan paradigman muutokseksi siinä, miten dataa fyysisestä maailmasta voidaan käyttää hyödyksi.

2.1 Digitaaliset kaksoset lääketieteessä

Digitaalisen kaksosen konseptissa on nähty monipuolisia mahdollisuuksia lääketieteelle. Metodologialla voitaisiin tehostaa lääkekehitystä^{9,10}, suunnitella terveydenhuollon resurssien jakoa¹¹, optimoida potilaan hoitopolku ja vähentää tehottomiksi osoittautuvien hoitojen määrää^{11,12} sekä yksilöidä hoidot niin, että potilas saisi mahdollisimman suuren hyödyn mahdollisimman vähäisillä haitoilla^{5,6}. Jopa koko potilaan elimistöä mallintavaa, kaiken mahdollisen eristettävän biologisen datan sisältävää digitaalista kaksosta on visioitu yksilöllisen lääketieteen tulevaisuutena⁵, mutta tämä on vielä teknologisesti kaukana toteutumisesta.

Insinöörityöiden tarpeisiin kehitetyn konseptin tuominen lääketieteeseen on monimutkaista. Mikäli digitaalinen kaksonen määritellään siihen tarvittavan kolmen komponentin – fyysisen objektin, virtuaalisen kopion ja niiden yhteyden – kautta, on yksi suurimmista ongelmista yhteyden ylläpito mallin ja potilaan välillä. Etenkin alkuperäisen insinöörityöperustaisen määritelmän mukaista reaaliaikaista yhteyttä on vaikea saavuttaa. Termiä ”digitaalinen kaksonen” ollaankin kirjallisuudessa ja tutkimuksessa käytetty mallinnuksista, joissa ihmistä tai elintä on mallinnettu virtuaalisesti mahdollisimman tarkasti, mutta ilman ulottuvuuksien dynaamista yhteyttä. Esimerkiksi lääkekehityksessä digitaalisten ”kaksosten” on ajateltu toimivan virtuaalisina potilaina tai eliminä esimerkiksi selvittäessä uuden lääkkeen menestymisen mahdollisuutta kliinisissä kokeissa, tutkittaessa lääkkeiden mahdollisia uusia off-label -käyttöaiheita¹⁰ ja keinotekoisien kontrollien luomisessa harvinaissairauksissa, joissa placebo-kontrollien käyttäminen olisi epäeettistä⁹. Näissä malleissa virtuaalisella kaksosella ei aina ole olemassa fyysistä vastinparia eli mallia vastaavaa potilasta.

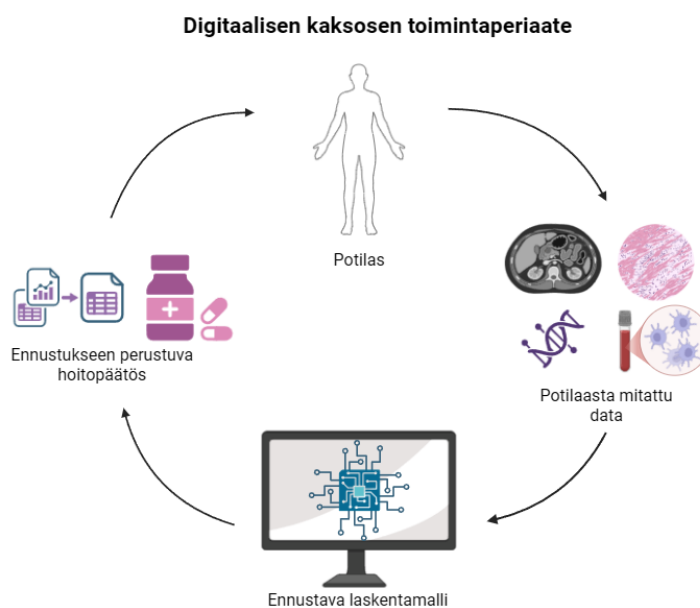
On kuitenkin perusteltua kyseenalaistaa lääketieteellisen digitaalisen kaksosen virtuaalisen ja fyysisen parin yhteydenpidon tiheyden tarvetta. Mikäli kaksosta käytetään potilaan sairauden mallintamiseen, reaaliaikainen yhteys patologisten muutosten ja mallin välillä voisi vaatia invasiivisen yhteyden potilaan kehoon. Poikkeuksen luovat tilanteet, joissa tietoa mallinnettavasta sairaudesta on saatavilla reaaliaikaisesti esimerkiksi päällä pidettävillä antureilla, kuten insuliinimittareilla¹³, kaatumisen havaitsevilla rannekeilla¹¹ tai verenpainemittareilla¹⁴. Kuitenkin harva patologisen tilan

muutos tapahtuu yhtäkkiä tai vaatii välitöntä reagoitua, jolloin potilaan virtuaalisen kaksosen harvempi päivittäminen, esimerkiksi lääkärin vastaanotolla, voisi riittää tuottamaan pätevän ja käyttötarkoitukseensa sopivan kaksosen.

Haasteista huolimatta digitaalisten kaksosten toteutuminen voisi tuoda monipuolista hyötyä lääketieteen alalle. Mallit voisivat kannustaa potilasta sitoutumaan ja osallistumaan omaan hoitoonsa, jonka lisäksi säännöllinen hoidon ja voinnin seuranta auttaisi hoitavaa lääkäriä riskien hallinnassa. Monipuolisen biologisen datan hyödyntäminen olisi askel kohti yksilöityä täsmähoitoa, jonka suuntaan esimerkiksi syövän hoidossa ollaan etenemässä.⁵

2.2 Digitaaliset kaksoset yksilöidyssä lääketieteessä

Yksi merkittävimmistä digitaalisen kaksosen sovellutuksista lääketieteessä on yksilöllisen hoidon mahdollistaminen. Vaikka kahdella potilaalla olisi sama tautidiagnoosi, voivat heidän patogeenisensä erota merkittävästi toisistaan. Yksilöllisen lääketieteen tarkoitus on kohdentaa potilaan hoito tarkasti juuri hänen tautinsa syntymekanismiin, esimerkiksi tutkimalla yksilöllisiä geneettisiä poikkeamia. Omiikka- ja kuvantamisteknologioiden kehittyttyä on mahdollista kerätä potilaasta monipuolisesti biologista dataa, jonka muodostama kokonaisuutta voisi hallita digitaalisen kaksosen avulla (kuva 1). Kaksoseen voisi liittää lisäksi potilaan kliinisen historian sekä saada jatkuvaa datasyötettä kannettavien antureiden kautta. Tuntemalla biologisen datan muodostama yksilöllinen kokonaisuus voisi olla mahdollista kohdentaa potilaan tarvitsema hoito juuri oikeaan paikkaan ja näin hallita haittavaikutusten ilmenemistä.⁵



Kuva 1. Digitaalisen syöpäkaksosen toimintaperiaate yksilöllisessä hoidossa. Potilaasta mitattava data syötetään ennustavaan laskentamalliin, jonka ennustetta voidaan hyödyntää hoitopäätöksiä tehdessä. Uudet mittaukset hoidon edetessä parantavat mallin ennustetta jatkohoitoa varten. Luotu BioRenderillä (<https://biorender.com>)

Yksilöllisiä digitaalisia kaksosia on hyödynnetty esimerkiksi suunnitellessa ortopedisiä leikkauksia ja postoperatiivista hoitoa^{15,16}. Tällöin potilaan luustosta on simuloitu 3D-malli, esimerkiksi kuvantamisdatan perusteella, ja mallinnettu vauriokohtaan kohdistuvaa mekaanista rasitetta ja määritetty paras stabilisointitapa postoperatiiviseen hoitoon. Yhteys potilaan ja mallin välillä voidaan saada aikaan esimerkiksi päälle puettavilla liikeantureilla¹⁵.

Digitaalisen kaksosen on lisäksi mahdollista mallintaa jotain potilaan tiettyä elintä ja ennustaa siinä tapahtuvia patologisia muutoksia ennen varsinaisen kliinisen oireyhtymän ilmentymistä. Etenkin sydämen toimintaa on mallinnettu laajasti ja malleja pyritty yksilöimään dynaamiseksi peilauksiksi potilaan tilasta. Esimerkiksi aortan pullistumaa on ennustettu valtimopainedatan perusteella¹⁴ ja aortan koarktaatiopotilaiden elektromekaanista aktiivisuutta on mallinnettu monitasomallinnuksella¹⁷. *The Living Heart Project*issa on luotu validoitu, koko sydämen 3D-malli elementtimenetelmällä (*finite element method*)¹⁸. Mallien tavoite on ollut sydäntautien patologian kokonaisvaltainen ymmärtäminen, sydäntapahtumien ennaltaehkäisy sekä mahdollisten lääkehoitojen vasteen ennustaminen potilaassa.

Potilaasta on myös mahdollista tehdä hänen terveyshistoriansa sisältävä avatar, virtuaalinen vastinpari, pilvipalveluun, joka sisältää kaiken kliinisen historian ja nykyisen terveydentilan kirjaukset eri sairaanhoidon osa-alueilta. Avatar voisi näin avustaa etenkin monisairaiden ihmisten hoidon koordinoinnissa ja hoidontarpeen ennakoinnissa niin, että se hälyttäisi, kun potilas tarvitsisi tiettyä terveydenhuollon palvelua. Virtuaalinen avatar voisi päivittyä lääkärin vastaanotolla tehtyjen tutkimusten pohjalta tai langattomien antureiden kautta. Digitaalisen kaksosen ei näin tarvitsisi mallintaa mitään tiettyä sairautta tai elimen tilaa yksityiskohtaisesti, vaan pitää hoitohenkilökunta ajan tasalla terveysdataan tehtävistä kirjauksista ja mahdollisista langattomasti mitattavista perustiedoista, kuten verenpaineesta tai loukkaantumisista, joiden perusteella hoitoa ohjattaisiin eteenpäin.¹¹

2.3 Digitaalisten kaksosten soveltaminen yksilöidyssä syövän mallintamisessa

Syöpä on yksi yleisimmistä kuolemaan johtavista syistä korkean elintason maissa. Kasvavan hoitotaakan alla on tarve entistä tehokkaammalle hoidolle ja aikaiselle diagnoosille, jotta tauti ei pääse etenemään¹.

Syöpää on jo pitkään kohdeltu geneettisenä sairautena, johon vaikuttavat yksilön henkilökohtaiset mutaatiot pahanlaatuisiksi muuttuneissa soluissa. Tiedetään, että potilaan genomi voi vaikuttaa sekä farmakokineettisiin että farmakodynaamisiin geeneihin, jotka puolestaan määrittävät syövän alttiuden tietyille hoidoille ja antavat näin osaselityksen sille, miksi jotkut potilaat hyötyvät hoidosta ja toiset eivät¹⁹. Kenties tunnetuin esimerkki molekulaarisesta profiloinnista on HER2-mutaatiotyypitys ja rintasyöpähoidon ohjaus potilaan kyseisen geenin mutaatiostatuksen perusteella. Korkean suorituskyvyn omiikkateknologioiden kehityksen myötä potilaista on mahdollistaa eristää yhä monipuolisempaa molekulaarista dataa, kuten epigeneettisiä markkereita, solumetabolia-aktiivisuutta ja proteiiniekspressiotasoja, joiden mukaan syöpätapaukset voidaan luokitella yhä tarkemmin ja jakaa potilaat entistä yksilöidympiin hoito- ja ennusteryhmiin²⁰.

Syövän kompleksisuus ei rajoitu pelkästään potilaiden välisiin eroihin, vaan myös syövän itsensä sisällä voi olla eroja tuumorisoluissa ja solujen mikroympäristössä. Jollakin puolella tuumoria geeni-ilmentymät tai epigeneettiset muutokset saattavat olla erilaisia verrattuna tuumorin toiseen puoleen. Tämä antaa syöväälle mahdollisuuden sen selviytymistä edistävään evoluutioon, jos syöpähoito tehoaa vain osaan soluista ja jättää loput eloon ja jakautumaan²¹. Tästä syövän sisäisestä heterogeenisuudesta on myös mahdollista eristää dataa; korkean resoluution spatiaalisella yksisolutraskriptomiikalla on mahdollista määrittää *in situ*, mitä geenejä yksi syöpäsolu ilmentää ja visualisoida ilmentymät syöpäkudoksen alkuperäisessä orientaatioissa²².

Omiikkadataa ei sovelleta kliinisessä käytössä kovinkaan laajasti, mutta erilaiset kuvantamisen keinot ovat yleisiä. Yleisin syöpähoidossa käytetty kuvantamistapa on TT-kuvaus²³, jonka lisäksi muita mahdollisuuksia ovat PET- ja magneettikuvaus (MRI). Kuvantamisella pyritään yhä tarkempaan syövän levinneisyyden ja hoitovasteen tarkkailuun^{24,25}. Kvalitatiivisten menetelmien rinnalle on kehitetty kvantitatiivinen radiomiikka, jonka avulla kuvantamisdatasta voidaan eristää suuri määrä dataa sen visuaalisen tulkinnan lisäksi, esimerkiksi tuumorin sisäisestä heterogeenisyydestä²³.

Mahdollisen potilaasta eristettävän datan määrän kasvu tekee kliinisestä datasta multimodaalista, jonka kokonaisuutta ja yhteyksiä voisi mahdollisesti hallita digitaalisella kaksosella⁶. Käytössä olevat

syövän hoitokeinot, kuten sytostaatit, sädehoito ja leikkaukset, ovat rankkoja ja mahdollisesti invasiivisia toimenpiteitä potilaalle, joista voi koitua laajoja haittavaikutuksia, toksisuutta ja pitkiä sairaalajaksoja. Näin ollen olisi sekä potilaan että terveydenhuollon resurssien kannalta optimaalista määrittää potilaalle toimiva hoito mahdollisimman vähillä haittavaikutuksilla ja näin vähentää myös tehottomiksi osoittautuvia hoitoja²⁴. Potilaskohtainen digitaalinen syöpäkaksoson voisi tarjota vastauksen tähän; sen avulla voidaan ennustaa hoitovastetta^{5,6}, auttaa hoitavaa tahoja tekemään optimoituja hoitopäätöksiä^{11,12} sekä minimoida haittoja^{5,6} käyttämällä mahdollisimman laajasti potilaan yksilöllistä kliinistä tietoa hyväksi. Mahdollisina datamuotoina voidaan käyttää kuvantamista^{12,24}, digipatologiaa⁸ tai molekulaarista ja biologista profilointia, kuten esimerkiksi metaboliapaneeleita ja solupopulaatiotutkimuksia^{26,27}.

Jotta digitaalinen syöpäkaksoson peilaisi mahdollisimman tarkoituksenmukaisesti fyysistä vastinpariaan, sen eri osat on suunniteltava tarkasti. Esimerkiksi hyödynnettävän kliinisen datan muoto, virtuaalisen kaksoson laskentatapa sekä sen koulutukseen tarvittavan datan määrän arviointi vaihtelevat kaksoson eri toteutustapojen välillä.

3 Digitaalisen syöpäkaksosen osat

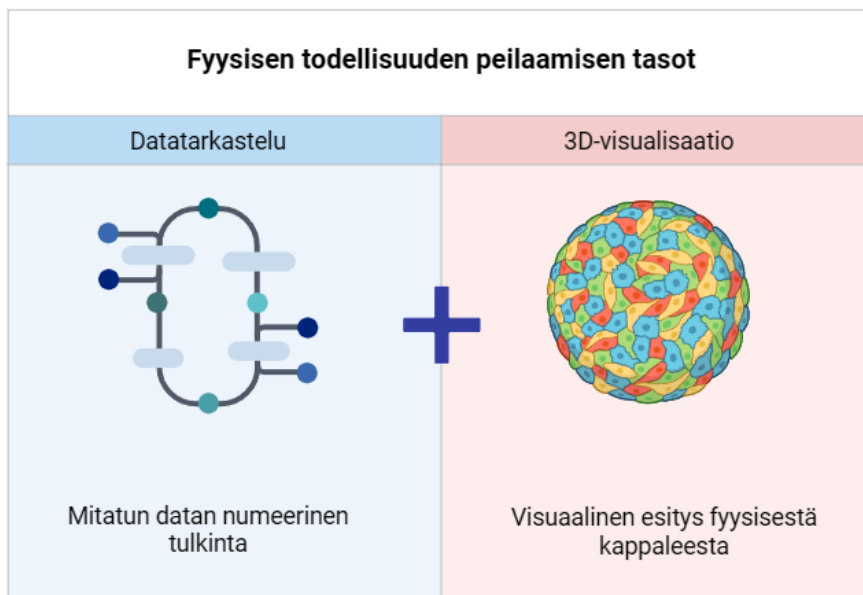
Digitaalisen syöpäkaksosen tulisi koostua samoista kolmesta osasta kuin muidenkin eri tarpeisiin luotujen digitaalisten kaksosten: fyysisen todellisuuden kappaleesta, sen virtuaalisesta kopiosta ja näiden välisestä dynaamisesta yhteydestä²⁸. Lääketieteen tarkoituksiin tulevan kaksosen tulisi rakenteensa ja toimintansa avulla saavuttaa viiden I:n periaatteet; sen tulisi olla yksilöity (*individualized*), yhteydellinen (*interconnected*), interaktiivinen (*interactive*), informatiivinen (*informative*) ja vaikutuksellinen (*impactful*)²⁹. Toisin sanoen digitaalisen syöpäkaksosen tulisi tuoda merkittävää, yksilöllistä hyötyä potilaalle antamalla tämän sairaudesta tietoa, jota ei saisi muuten kuin peilaamalla fyysistä todellisuutta virtuaalisella mallilla.

3.1 Fyysinen todellisuus

Digitaalisen kaksosen fyysinen vastinkappale voi olla kokonainen systeemi, sen ympäristö tai systeemissä tapahtuvat muutokset. Syöpäkaksosen osalta tämä voi tarkoittaa esimerkiksi kasvainta, sen mikroympäristöä tai muutosta leviämisoikeudessa. Myös potilaan koko keho voi toimia systeeminä tai tuumoria majoittavana ympäristönä. Jokainen digitaalisen syöpäkaksosen sovellus määrittelee erikseen, mikä on se fyysisen todellisuuden laajuus, jota peilataan.²⁸

Syövän digitaalisten kaksosten on visioitu kattavan mahdollisimman laajasti potilaan fyysistä tilaa. Kaksoseen voisi integroida elämänmittaista tietoa potilaan elintavoista, mikrobistosta, metaboliasta, immuunijärjestelmästä, genetiikasta, pituudesta, painosta tai mistä tahansa mitattavasta suureesta³⁰. Jos potilaan kehoa ajattelee syövän ympäristönä, kaikkien näiden muuttujien lisääminen siirtää tuumoriympäristön myös kaksosen virtuaaliselle osalle, mikä mahdollistaa erilaisten simulaatioiden pyörittämisen. Tällöin voidaan luoda ennustuksia fyysisen kappaleen tilasta vaihtoehtoisissa olosuhteissa²⁸.

Olemassa olevat syöpäkaksoset peilaavat usein syövässä tapahtuvia muutoksia; esimerkiksi tuumorin koon muutoksia hoitovasteen funktiona²⁴, mikroympäristön muutoksia³¹, solupopulaatiotasoa²⁷ ja pahanlaatuisten solujen leviämistä⁸. Toisaalta muutosten mallintamiseen vaaditaan usein mallinnus myös itse tuumorista systeeminä, esimerkiksi visuaalisessa 3D-muodossa²⁴. Kuvassa 2 on eritelty fyysisen todellisuuden mallintamisen vaihtoehtoja. Numeerinen data voi sisältää esimerkiksi edellä mainitut solupopulaatiotasot, 3D-visualisaationa taas voi toimia esimerkiksi spatiaalinen solukartta.



Kuva 2. Fyysisen todellisuuden peilaamisen tasot syövän mallintamisessa. Digitaalinen kaksonen voi laatia ennusteita joko pelkän numeerisen datan tai tuumorin visuaalisen tulkinnan pohjalta. Menetelmien yhdistelmät ovat myös mahdollisia. Luotu BioRenderillä (<https://biorender.com>).

3.2 Virtuaalinen todellisuus

Digitaalisten kaksosten virtuaalinen osan tarkoitus on mahdollistaa fyysisen vastinparin monitorointi ja reagointi sen muutoksiin. Tekniikka pohjaa esineiden internettiin (IoT, *Internet of Things*, jossa fyysiset kappaleet ovat langattomasti kytkettyjä internetverkostoon, missä niillä on virtuaalinen vastinkappale. Kytkentä tapahtuu älykkäillä antureilla, jolloin tieto fyysisestä todellisuudesta siirtyy virtuaaliseen. Yhteys on dynaaminen ja verkosto toimii ilman ihmisen interventiota; muutos virtuaalisessa maailmassa aiheuttaa toimintamuutoksen fyysisessä objektissa ja toisinpäin. Konseptin alkuperä on todennäköisesti Kevin Ashtonin vuonna 1999 pitämässä esitelmässä ja teknologiaa on sittemmin hyödynnetty esimerkiksi logistiikassa ja älykodeissa³². Lääketieteellisessä käytössä IoT mahdollistaa potilaan jatkuvan seurannan ja etänä tapahtuvan hoidon ohjaamisen, kun antureilla mitataan esimerkiksi verenpainetta tai glukoositasoja³³.

Fyysisen todellisuuden siirtäminen virtuaaliseen muotoon vaatii aina jonkin verran abstrahointia, kun todelliset fysikaalisbiologiset prosessit käännetään matemaattiseksi ja tilastolliseksi informaatioksi. Mallit ovat joko data- tai laskentamalleja. Datamallit kuvaavat fyysisen todellisuuden tilaa siitä saadun datan perusteella, kun taas laskentamallit havaitsevat yhteyksiä datan sisällä ja kuvaavat, miten fyysinen kappale toimii ja muuttuu²⁸. Laskentamallien perusteella on mahdollista jakaa kaksoiset kolmeen luokkaan: mekanistisiin, koneoppimisperusteisiin ja näiden yhdistelmiin. Mekanistisella mallilla esimerkiksi hoitovasteen tai tuumorimassan ennustaminen tapahtuu

matemaattisella kaavalla, joka yksilöidään potilaan parametreilla. Se ei itse luo yhteyksiä moniulotteisen datasetin sisällä, vaan noudattaa siihen koodattuja lainalaisuuksia. Koneoppimisperusteiset mallit taas pystyvät luomaan itsenäisesti yhteyksiä datasta, syväoppimismallit jopa generatiivisesti, antaen ulostulona tilastollisen todennäköisyyteen perustuvan arvion⁸. Eri mallien toisiaan täydentävät yhdistelmät ovat myös mahdollisia^{8,28}.

Digitaalisen kaksosen virtuaalinen osa on useimmiten laskentamalli. Toisin kuin pelkkä ennustava simulaatio, digitaalinen kaksonen pyrkii tarkastelemaan objektin mennyttä, nykyistä ja tulevaa tilaa eli sen koko elinkaaren aikana tapahtuneita ja tapahtuvia muutoksia²⁸. Laskentamalli muodostuu sen perusteella, mitä fyysistä piirrettä syövästä mallinnetaan; se voi olla esimerkiksi geometrinen rekonstruktio tuumorista²⁴, molekulaarinen simulaatio signaalintireittien muutoksista²⁶ tai biologinen ennuste solupopulaatiomuutoksista²⁷.

3.3 Fyysisen ja virtuaalisen todellisuuden yhteys

Yleinen metodologia fyysisen ja virtuaalisen todellisuuden tiedonvaihdon välillä on tila-avaruusmalli. Tilalla tarkoitetaan simuloitua fyysisen todellisuuden piirrettä, joka muuttuu mallin yhdistäessä aiemmat mittaustulokset uuteen syötteeseen. Toisin sanoen tila kuvaa systeemin käytöstilaa, joka syöväkohdalla voi olla esimerkiksi jakautumistila ja remission tila. Tila-avaruusmallin parametreilla tarkoitetaan tilojen välisiä yhteyksiä; ne kuvaavat tilojen välisiä riippuvuuksia sekä sitä, miten uusi syöte vaikuttaa tilaan. Parametrit voivat olla staattisia tai muuttua ajan saatossa²⁸. Esimerkiksi lääkekonsentraation vaikutus jakautumiseen tai remission saavuttamiseen on parametri.

Fyysisen ja virtuaalisen todellisuuden yhteyden tulisi olla kaksisuuntainen²⁹. Fyysisen todellisuuden yhteys virtuaaliseen saavutetaan keräämällä fyysisestä kappaleesta dataa ja päivittämällä sillä virtuaalisen vastinparin tilaa. Tällainen yhteys voidaan saavuttaa esimerkiksi antureilla, mutta myös manuaalisella keräyksellä, kuten kuvantamistalalla²⁴, verikokeilla tai nestebiopsiolla^{27,30,34–36}. Koska kerättäviä datamuotoja voi olla monia, kaksonen on multimodaalinen ja sen päivitys voi vaatia monen eri datamuodon fuusiota ja käsittelyä yhteensopivaksi²⁸. Jatkuva päivitys mahdollistaisi mallin ennusteiden parantumisen²⁷.

Päinvastainen yhteys, eli virtuaalisen todellisuuden yhteys fyysiseen, tarkoittaa tiedonsiirtoa takaisin fyysiseen kappaleeseen. Laskentamalli luo ennusteen fyysisen kappaleen tulevasta tilasta, jota käytetään päätöksenteon tukena²⁹. Käytännössä tämä tapahtuu arvioimalla fyysiseen kappaleeseen

kohdistuvan intervention tai uusien mittausten tarvetta, jotka toteutuessaan muuttaisivat kappaleen tilaa.²⁸

Digitaalisen kaksosen alkuperäisessä tarkoituksessa ja sen sovellusten ideaalitapauksissa yhteys fyysisen ja virtuaalisen todellisuuden välissä olisi reaaliaikainen²⁹. Määritelmä on kuitenkin laajentunut alkuperäisestä, ja yhteyden voidaan ajatella olevan riittävä, jos kaksosella on kyky päivittyä riippumatta uuden datan saant nopeudesta. Etenkin lääketieteellisestä kaksosesta puhuttaessa eri datamuotoja voi tulla eri tahtiin; verenpaine voidaan mitata antureilla reaaliaikaisesti, kun taas uutta kuvantamisdataa saadaan vasta lääkärin vastaanotolla²⁸. Syöpäkaksosen osalta on määritettävä erikseen hoitotulosten kannalta tarkoituksenmukainen päivitystiheys.

4 Molekulaariseen ja biologiseen profiiliin perustuvat mallit

Terveiden ja pahanlaatuisten solujen molekulaariset profiilit eroavat toisistaan muun muassa muuttuneen geeni-ilmentymän, metabolian ja mikroympäristön osalta. Lisäksi myös syövän itsensä sisällä solujen välillä voi vallita suuri heterogeenisuus³⁶. Näiden tarkkojen molekulaaristen muutosten tunteminen olisi askel kohti henkilökohtaisempaa ja paremmin kohdennettua syöpähoitoa. Yksittäisiä matemaattisia mallinuksia on luotu muun muassa syöpäsolujen genomiikasta ja transkriptomiikasta^{34,36}, potilaan metabolisista biomarkkereista^{30,35}, solumääristä²⁷, tuumorin mikroympäristön luonteesta³¹ ja jopa syöpäsolujen muuttuneesta sähköfysiologiasta³⁷.

Edellä mainitut mallinnukset voisivat itsenään toimia digitaalisen kaksosen osana selittäessään ja ennustaessaan potilaan taudin kehittymistä ja vastetta hoitoon. Verrattuna esimerkiksi kuvantamista hyödyntäviin malleihin molekulaariseen malliin tarvittava data on usein saatavilla potilaasta helpommin, esimerkiksi verikokeen kautta³⁵. Kuitenkaan kirjallisuudessa esitetyt, biologisiin näytteisiin perustuvat digitaaliset kaksoset eivät vielä ole yltäneet kliiniseen validaatioon tai dynaamiseen yhteyteen potilaaseen^{27,30,35}, joka erottaisi pelkän mallinnuksen määritelmän mukaisesta digitaalisesta kaksosesta. Syöpäsolulinjojen ja yksisoluumiikkojen pohjalta tehdyt mallit esimerkiksi tietyn syöpätyypin genomien yhteydestä lääkevasteeseen tai metabolisten reaktioiden poikkeavuuksista ovat edistyksellisiä askelia kattavan digitaalisen kaksosen luomiseksi, mutta sellaisenaan käytettynä vain tilastollisia ja mekanistisia malleja tietystä solusta tiettyä ajanhetkenä^{30,34,36}. Lisäksi molekulaarista dataa hyödyntävien digitaalisten kaksosten haasteena on usein sopivan datan vähyys ja sen laadun vaihtelu, joka vaikeuttaa ennustavan mallin luontia³⁵.

Tutkielmassa esitellään seuraavaksi digitaalisen kaksosen termiä käyttäviä, potilaan molekulaarista ja biologista dataa hyödyntäviä mallinuksia viime vuosilta. Koska yksiselitteistä luokittelutapaa digitaalisille kaksosille ei ole, tarkasteluun on valittu sekä koneoppimisperustaisia että mekanistinen malli, jotta eri toteutustavat tulevat edustetuksi. Mallien luojat ovat korostaneet, kuinka molekulaariset mallinnukset ottavat niiden rajoituksista huolimatta huomioon syövän heterogeenisen molekulaarisen profiilin ja voisivat näin ollen olla tarvittava askel kattavan digitaalisen kaksosen luomiseksi^{27,30,35}.

4.1 Metabolisten biomarkkereiden käyttö koneoppimis pohjaisissa syöpäennusteissa

Metaboliseen dataan perustuvat mallit lähtevät oletuksesta, että monimutkaisessa kehossa sen yhtä lailla monimutkainen metabolinen koneisto voidaan nähdä yhtenä toiminnallisena osa-alueena, ja jota voidaan näin ollen mallintaa digitaalisessa kaksosessa³⁵. Hou et al. loivat vuonna 2022 ja Moztaarzadeh et al. vuonna 2023 metabolisiin biomarkkereihin perustuvat syöpämallit syövän aiheuttaman muuttuneen metabolisen profiilin perusteella^{30,35}. Moztaarzadeh et al.:n malli keskittyi rintasyövän diagnosoimiseen, kun taas Hou et al.:n malli metabolisten poikkeamien ja potentiaalisten haittatapahtumien ennustamiseen.

Vaikka kumpikaan luoduista malleista ei sellaisenaan vastaa digitaalista kaksosta, voi kirjoittajien mukaan niiden toimintaperiaatetta käyttää kaksosen luomisen osana. Metaboliset mallit luovat edellytyksen päästä tarkastelemaan potilaan metabolisen tilan historiaa ja kehittymistä hoidon myötä, mikä puolestaan tarjoaa tilaisuuden ennustaa sairauden tulevaa kehityskaarta³⁵. Fyysisen kappaleen koko elinkaaren tarkastelu onkin yksi digitaalisen kaksosen piirteistä²⁸. Kumpikin ryhmä hyödynsi koneoppimista metabolisen datan tulkinnassa sekä yhteyksien löytämisessä biomarkkeritasojen ja annetun syöpähoidon välillä.

Moztaarzadeh et al. havaitsivat satunnaispuuregression ja *gradient boosting* -algoritmin olevan tehokkaimpia koneoppimismenetelmiä määrittämään potilaan rintasyöpästatus tunnettujen dysreguloitujen syöpämarkkereiden perustella³⁰. Vaikka luotu malli toimii tällaisenaan pelkästään diagnosointimenetelmänä, voisi sitä käyttää ennustavana työkaluna, jos algoritmi oppisi tunnistamaan varhain havaittavia merkkejä syövän kehityksestä tai häviämisestä.

Hou et al. hyödynsivät takaisinkytkettyä neuroverkkoa (*recurrent neural network*) analysoidessaan viiden sädehoitoa saaneen potilaan metabolisia paneeleita eri pisteissä hoidon aikana. Ryhmä hyödynsi siirto-oppimista (*transfer learning*) luodessaan jokaiselle potilaalle oman mallin ensin luodusta yleisestä mallista. Malli onnistui ennustamaan potilaan metabolisen profiilin muutoksia luotettavasti vain parin päivän aikaikkunassa, jonka ryhmä arvelee johtuvan ainakin osittain datan vähyydestä³⁵.

Kyseiset tutkimukset on otettava osoituksena, kuinka digitaaliseen kaksoseen sisällytettävän metabolisen mallin voi luoda, mutta toimivan ja validoidun mallin rakentamiseen tarvittaisiin näitä tutkimuksia huomattavasti enemmän dataa. Lisäksi digitaalisen kaksosen toteutuessa mallien tulisi päivittyä esimerkiksi uusien verikoetulosten myötä. Mikäli kuvailtu syövän biomarkkereihin

perustuva kaksonen toteutuisi, voisi sen avulla testata esimerkiksi erilaisten hoitovaihtoehtojen toimivuutta ja ennustaa taudin etenemistä^{30,35}.

4.2 Solupopulaatiotasojen käyttö mekanistisessa haittavaikutusennusteessa

Zhang *et al.* loivat vuonna 2025 mallin, joka ennusti sytostaattihoidon jälkeistä neutropeniaa akuuttia myeloonista leukemiaa sairastavilla potilailla. Neutropenia, eli neutrofiilien vähentynyt määrä on yleinen haittavaikutus akuutin myeloonisen leukemian hoidossa, sillä syöpää targetoivat sytostaatit saattavat aiheuttaa myös luuydinvaurioita. Vähentyneen neutrofiilimäärän johdosta potilaan infektioriski kasvaa ja pahimmassa tapauksessa hoito saatetaan keskeyttää siitä ilmenneiden haittojen vuoksi. Luodun mallin tarkoitus oli haittavaikutusten riskin arviointi sekä lopulta parempi potilaskohtainen hahmotus sytostaattihoidon hyöty-haittasuhteesta.²⁷

Ryhmä loi mekanistisen, differentiaaliyhtälöperusteisen mallin, joka käytti syötteenään neutrofiilien lukumäärää ja niiden esiasteiden prosenttiosuutta. Data solumääristä ja lääkkeiden antoajankohdista saatiin rekistereistä retrospektiivisesti jo toteutuneista pitkittäistutkimuksista ja se jaettiin kahteen osaan: mallin kouluttamisdataan ja erilliseen testausdataan²⁷. Retrospektiivisen luonteensa takia ryhmän luomasta mallista puuttuu digitaalisen kaksonen määritelmään kuuluva dynaaminen yhteys potilaaseen, mutta se olisi mahdollista saavuttaa päivittämällä mallia esimerkiksi lääkärin vastaanotolla sen jälkeen, kun malli on luotu potilaan aikaisempien tietojen perusteella. Tuloksena ryhmä totesi mallin ennustavan potilaan hoidon jälkeistä neutrofiilimäärää suhteellisen hyvin²⁷.

Mitä enemmän dataa on saatavilla, sitä tarkemmaksi mallit yleensä muotoutuvat. Kyseisessä tutkimuksessa vain seitsemästä potilaasta oli kerätty riittävästi tietoa sekä mallin koulutukseen, että testaukseen ja nämä potilaat edustivat joukkoa, jolla oli keskimäärin pidempi selviytymisaika diagnoosista ja paremmat vasteet sytostaattihoitoihin. Vastoin odotuksia mallin luomat ennustukset näiden potilaiden neutropeniasta olivat heikompia kuin muiden, mikä saattaa viitata mallin ylisovittumiseen niiden potilaiden suuntaan, joista dataa ei saatu kerättyä yhtä paljon²⁷.

Puhuttaessa digitaalisesta kaksosesta sen yhden odotetuimmista hyödyistä tulisi olla mahdollisimman reaaliaikainen arviointi potilaan tilasta; näin ei ole, jos potilaasta tarvitaan pitkittäistutkimusdataa mallin luomiseen. Uuden syöpäpotilaan osalta tämä tarkoittaisi pitkäaikaista datankeruuta ennen toimivaa ennustavaa mallia, jolloin mallin relevanssi hoidon osana pienenesi. Ryhmä havaitsi myös, että tietyt taustaoletukset neutrofiilien lukumäärän reagoinnista hoitoon eivät

päteneet kaikille potilaille, joten he loivat toisen mallin, jonka tarkoituksena oli potilaan genomia ja muita kliinisiä tietoja hyväksikäyttäen arvioida, minkälainen laskentamalli toimisi kyseisen potilaan haittavaikutusten ennustamiselle²⁷. Mikäli pitkittäisseurantadataa eri potilaista saataisiin runsaammin, voisi olla mahdollista kehittää useampi tämänkaltainen neutropeniaa ennustava laskentamalli, joihin uuden potilaan tapaus voitaisiin sovittaa. Malli voitaisiin päivittää henkilökohtaiseksi diagnoosihetkellä saatavilla olevasta datasta esimerkiksi siirto-oppimisen avulla, jota Hou et al. hyödynsivät metabolisessa mallinnuksessaan³⁵.

5 Kuvantamisdataan perustuvat kaksoset ja niiden sovellukset

Kuvantamisdataan perustuvien digitaalisten syöpäkaksosten ajatellaan molekulaaristen ja biologisten lähestymistapojen tapaan voivan tuoda hyötyä muun muassa hoitovasteen ennustamiseen ja hoidosta hyötyvän potilaspopulaation tunnistamiseen²⁴. Lisäksi malleja voidaan käyttää sopivan hoitoyhdistelmän määrittämisessä jo heti hoidon alussa, jolloin vältetään tehottomat ja haittavaikutuksia aiheuttavat interventiot¹². Kuvantamista hyödyntämällä voidaan pyrkiä esimerkiksi suunnittelemaan leikkauksia tehokkaammin ennustamalla tarkemmin, minne kaikkialle syöpäkudos on levinnyt elimistössä. Tällöin leikkaus voitaisiin kohdistaa tarkemmin ja poistaa syöpäkudos tehokkaasti vaurioittamatta tervettä kudosta⁸.

Kuten molekulaariseen profiiliin perustuvissa malleissa, kuvantamismallien hyödyntämät datamuodot ovat monipuolisia. Lääketieteellistä kuvantamisdataa saadaan esimerkiksi MRI-, TT-, ja PET-kuvauksista sekä histologisista kudospäätteistä digipatologian keinoin. Radiomiikalla kuvantamisdatasta pystytään eristämään suuria ja multimodaalisia datasettejä, joiden avulla on mahdollista kouluttaa tilastollisia laskentamalleja³⁸. Ideaalisti malleihin integroidaan kuvantamisdatan lisäksi myös muita relevantteja potilastietoja, kuten kliininen historia ja mahdolliset aiemmat hoitotulokset, jotta malli kattaa potilaan elinkaaren mahdollisimman monipuolisesti¹². Verrattuna molekulaarisiin malleihin kuvantamisdataan perustuvien mallien suurin etu on niiden käyttämän datan spatiaalinen sekä ei-invasiivinen luonne²³, mutta molekulaaristen mallien tavoin ne kärsivät usein patologisen datan vähyydestä⁸.

Koska kuvantamiseen perustuvia digitaalisia syöpäkaksosia ei ole molekulaaristen kaksosten tapaan vielä kliinisessä käytössä, ei niille ole myöskään yksiselitteistä luokittelutapaa. Jokainen malli kuitenkin luodaan sairauslähtöisesti kuvaamaan tiettyä syöpää, tiettyä hoitokeinoa ja, mikäli mahdollista, tiettyä potilasta^{8,12,24}. Tutkielmassa esitellään seuraavaksi edistyneimmät yritykset luoda kuvantamisdatapohjaisia, potilasta peilaavia malleja. Mallit on aiempien esimerkkitapausten tapaan jaoteltu laskennallisesti mekanistisiin, koneoppimispohjaisiin ja näiden yhdistelmämalliin.

5.1 MRI:n käyttö neoadjuvanttihoitovasteen mekanistisessa ennustamisessa

Vuonna 2022 Wu *et al.* yhdistivät triplanegatiivisten rintasyöpäpotilaiden MRI-kuvantamisdatan biologiseen matemaattiseen malliin²⁴. Aikaisemman tutkimustiedon mukaan vain kolmasosa

potilaista hyötyi standardikäytön hoidosta merkittävästi³⁹, mutta MRI-mittaustulosten ennen ja jälkeen hoidon tiedettiin tarjoavan ennustavia sovellutuksia tuumorin koon muutoksista⁴⁰. Luodun digitaalisen kaksosen tarkoitus oli hyödyntää MRI-kuvien ennustuspotentiaalia määrittämään yksilön todennäköisyyttä hyötyä standardineoadjuvanttihoidosta. Mallin yksilöllisyys saavutettiin käyttämällä sen koulutukseen pelkästään potilaan omaa kuvantamisdataa, jolloin hoitovaste-ennuste ei perustunut populaatiokeskiarvoihin²⁴.

Dataa potilaiden tuumoreista kerättiin sekä dynaamisella kontrastivahvistetulla (*dynamic contrast-enhanced*), että diffuusiopainotetulla (*diffusion-weighted*) MRI:llä ennen hoitoa, sen aikana ja sen jälkeen²⁴. Radiopatologit segmentoivat kuvantamisdatan hyvän- ja pahanlaatuiseen kudokseen, jonka avulla tuumoreista koottiin visuaaliset solukartat. Kartasta saatiin dataa tuumorisolujen diffuusiosta, jakautumisesta ja kuolemasta, jonka avulla laskentamalli tuotti arvion terapian toimivuudesta²⁵. Tätä yksilöityä laskentamallia käytettiin ennustamaan jatkohoidon mahdollista toimivuutta, jolloin syntyi toimintakehys digitaaliselle kaksoselle. Digitaalisen kaksosen luoma lopullinen hoitovaste-ennuste arvioitiin jatkohoidon jälkeisellä leikkauksella²⁴. Yhteyden potilaaseen voidaan katsoa syntyneen uusissa MRI-mittauspisteissä, kun laskentamallia kalibroitiin niiden tulosten perusteella.

Digitaalinen kaksonen todettiin ennustavan tarkasti yksittäisen potilaan tuumorin loppustatusta neoadjuvanttihoidon jälkeen. Kliinisessä käytössä malli voisi antaa tukea hoitopäätökseen ennustamalla neoadjuvanttihoidon toimivuutta ja auttaa näin vähentämään turhia haittavaikutuksia tunnistamalla potilaat, jotka eivät todennäköisesti hyötyisi kyseisestä hoidosta²⁴. Kuten molekulaarisissa digitaalisissa syöpäkaksosissa, MRI-mallin kouluttaminen vaatii kuitenkin longitudinaalista dataa jo toteutuneesta syöpähoidosta, jolloin näin luotu kaksonen ei tarjoaisi välitöntä ennustusta uusille syöpätapauksille, ellei hyödynnetä siirto-oppimista Hou et al:n tavoin.

5.2 Optimaalisen hoitopolun määrittäminen syvällä Q-oppimisella

Monissa syöpätapauksissa lopputulemaltaan tehokkaimman hoitopolun määrittäminen heti diagnosivaiheessa on mahdotonta. Hoito perustuu usein peräkkäisiin päätöksiin, jotka tehdään edellisten tulosten perusteella; esimerkiksi arvioimalla miten syöpä on vastannut hoitoon tai onko liian suuria haittavaikutuksia ilmennyt. Suunielusyövän osalta klinikko valitsee jatkuvan kemoterapian, sädehoidon sekä näiden hoitojen yhdistelmän väliltä potilaan hoitopolun, joka perustuu syövän paikallisiin muutoksiin ja hoidosta ilmenneisiin haittavaikutuksiin¹². Syövän hoidossa

voitaisiin ottaa askel tehokkaampaan ja yksilöllisempään hoitoon, jos jo diagnoosihetkellä osattaisiin ennustaa, mitkä hoitopäätökset tuottaisivat potilaalle parhaan lopputuleman.

Suunielusyövässä nielemisvaikeuksien esiintymiselle ja eloonjäämisen ennustamiseksi on luotu syvään Q-oppimiseen perustuva malli (*deep Q-learning*, DQL)¹². DQL on koneoppimisen keino, jossa syvä neuroverkko koulutetaan ennustamaan eri päätösten mahdollisia lopputulemia Q-arvon avulla. Q-arvo kuvaa mallin saamaa ”palkintoa”, eli sitä, kuinka suotuisaan lopputulemaan päätösyhdistelmällä päästään, ja josta malli oppii, mitkä valinnat johtavat todennäköisemmin parhaaseen tulokseen. Keinoa on hyödynnetty alustavien kliinisten päätöstentekomallien rakentamisessa^{12,41}.

Tardini et al. loivat vuonna 2022 DQL:ään sekä digitaalisen kaksosen dyadiin perustuvan hoitopäätösmallin. Dyadi, eli pari, koostui sekä potilaan tilan että lääkärin päätösten mallintamisesta. Potilasta kuvaava datamalli luotiin saatavilla olevien kliinisten tietojen perusteella, joihin lukeutui esimerkiksi syövän aste, tuumorin sijainti, sukupuoli ja elintavat. Osiin potilasmalleista integroitiin myös radiomiikkadataa tuumoreista. Dyadin toinen pari ennusti eri hoitopäätösten toimivuutta ja mallinsi näin ikään hoitavan lääkärin päätöksentekoa.¹²

Luotu digitaalinen syöpäkaksospari ennusti hoitopolun tuloksia, eli yleistä selviytymistä ja nielemisvaikeuksia, noin 87 % tarkkuudella. Paras malli teki 70 % ajasta saman päätöksen kuin hoitava lääkäri, mikä toimii jonkinasteisena vertailukohtana mallin kliiniselle pätevyydelle¹². Huomionarvoista oli, että parhaaseen malliin ei ollut sisällytetty kuvantamisdatasta saatua radiomiikkadataa. Tämä osoittaa, kuinka toimivan lääketieteellisen digitaalisten kaksosten luominen on täysin riippuvaista valittavista ennustettavista päätetulemista sekä siihen integroitavan datan yhteensopivuudesta taustalla oleviin laskentakeinoihin. Kaksosen luojat korostavat, että malli voisi mahdollisesti toimia avustavana tekijänä hoitopäätösten teossa, mutta ei vielä itse päättää prosessista¹².

Kyseisessä mallissa on tärkeää huomioida, että lääkäriä simuloiva kaksonen perustui dataan potilaspopulaation hoitopoluista ja niiden tuloksista, eikä näin ole Wu et al:n kaksosen tavoin täysin yksilöity yhden potilaan sairaudelle. Muiden syöpäkaksosten tavoin yhteys potilaaseen olisi läsnä vain lääkärin vastaanotolla uusien mittausten ja niiden perusteella tehtyjen hoitopäätösten toimivuuden arvioinnin kautta.

5.3 Syövän levinneisyyden ennustaminen mekanistisen ja tekoälypohjaisen mallin yhdistelmällä

Mekanistiseen malliin saa hyvin integroituja biologisia lainalaisuuksia, mutta siltä puuttuu kyky havaita multimodaalisen datan sisäisiä yhteyksiä. Koneoppimisen keinot taas havaitsevat nämä yhteydet, mutta saattavat ylisovittaa ja johtaa harhaan, jos datan määrä on liian vähäinen. Gaw et al. rakensivat vuonna 2019 hybridimallin, joka yhdistää sekä mekanistisen että koneoppispohjaisen mallin ennustaessaan glioblastooman tuumorisolutiheyttä tarkemmin kuin mitä tiheyttä pystytään havaitsemaan perinteisillä kuvantamismenetelmillä⁸. Kyseisen julkaisun tekijät eivät käytä mallistaan termiä digitaalinen kaksonen, mutta edellä esiteltyjen kuvantamissovellutusten tapaan malli pyrkii yksilöimään ja optimoimaan glioblastoomapotilaiden osalta tehtäviä hoitopäätöksiä potilaista saatavan datan mukaan.

Yleinen hoito glioblastoomassa on kuvantamisella kohdennettu leikkaushoito. Valitettavasti monet rutiinikäytön kuvantamismenetelmät eivät pysty täysin tarkasti erottelmaan invasoitunutta tuumorikudosta terveestä aivokudoksesta, jolloin on riski, että kuvantamisen perusteella suunnitellussa leikkauksessa ei saada poistettua maligniteettia kokonaan. Esimerkiksi gadoliniumtehosteista, T1-painotteista MRI:tä (*gadolinium contrast-enhanced T1 weighed MRI*, T1Gd) ja PET-kuvantamista käytetään havaitsemaan vaurioitunut veriaivoeste ja proliferoituva osa kudoksesta, mutta menetelmillä ei välttämättä yksinään saada määritettyä absoluuttisen tarkkaa rajaa hyvän- ja pahanlaatuisen kudoksen välillä.⁸

Gaw et al. kouluttivat hybridimallin glioblastoomapotilaiden multiparametrisilla MRI-kuvilla sekä histologisilla biopsioilla. Jokaiselta potilaalta kerättiin kuuden tyyppiset kuvat: T1Gd, T2-painotettu (*T2-weighed*), dynaaminen kontrastivahvistettu (*dynamic contrast-enhanced*), keskimääräinen diffuusio (*mean diffusivity*), FA-kartta (*fractional anisotropy*) ja suhteellinen aivoveren tilavuus (*relative cerebral blood volume*). Patologit arvioivat biopsioiden tuumoritaakan Wu et al.:in tapaan histologisesti.⁸

Segmentoidun tuumorikudoksen solutiheyden muutokset arvioitiin käyttämällä Neal et al.:in kehittämää mekanistista proliferaatioinvaasiokaavaa⁴². Koneoppimismalli yhdisti mekanistisen kaavan analysoituihin biopsiatietoihin ja multimodaaliseen kuvantamisdataan, jolloin tuloksena algoritmi loi tarkennetun ennusteen maligniteetin levinneisyydestä. Näin luodun yhdistelmämallin tarkkuus testattiin kouluttamalla se vuoroin vain potilaskohtaisilla parametreillä ja vuoroin koko potilaspopulaation datalla. Potilaskohtaisen mallin havaittiin tuottavan tarkempia ennusteita, mutta

sisältävän huomattavan riskin ylisovittumiseen. Yhdistelmämallin havaittiin tuottavan yksittäisiä mallinnuskeinoja tarkempia ennusteita.⁸

Gaw et *al.*:in hybridimalli osoittaa datan määrän tärkeyden onkologisissa malleissa, jotka käyttävät laskentakeinonaan tekoälyä. Glioblastooman kaltaisissa harvinaisissa sairauksissa tervettä kontrollidataa on usein saatavilla paljon, kun taas patologisia muutoksia ilmentäviä biopsioita ja kuvia on huomattavasti vähemmän⁸. Tämä vaikeuttaa luotettavan mallin luontia, jos mallin koulutusdata vinoutuu pienen tai homogeenisen potilasjoukon perusteella. Työ osoitti kuitenkin, kuinka mekanististen mallien yhdistämisellä koneoppimismalleihin pystyy osin lieventämään tätä datan vähyydestä johtuvaa ennusteiden epätarkkuutta. Kuitenkin, mikäli luodun mallin tarkoituksena on vähentää epätarkasta leikkaussuunnittelusta johtuvia uusintaleikkauksia, kumoaa mallin koulutukseen vaaditut invasiivisten biopsiat tämän tarkoituksen.

6 Molekulaarisen profiloinnin ja kuvantamisen hypermalli

Ihmiselimistö, kuten syöpäkin, on loppujen lopuksi kompleksi ja heterogeeninen koneisto, jonka täydelliseen peilaamiseen voisi upottaa lähestulkoon loputtomasti parametrejä aina solutyypeistä geeni-ilmentymiin ja immuunijärjestelmään. Kuten edellisissä kappaleissa on esitetty, ei digitaalisen syöpäkaksosen kuitenkaan ole aina tarkoituksenmukaista, eikä toisaalta mahdollistakaan, mallintaa jokaista yksityiskohtaa sairaudesta tai potilaasta, jotta se toimii valitun lopputuleman ennusteena^{8,12,24,27,30,35}.

Biologisen datan saatavuuden kasvun myötä on kuitenkin avautunut mahdollisuuksia tutkia ihmistä ja tämän sairautta monella eri tasolla. Erilaisten analyysien, kuten kuvantamisen, geeni-ilmentymien ja metaboliapaneelien, tuottaman datan integroiminen yhteen malliin voisi yhdistää tietoa uudella tavalla ja johtaa entistä tarkempaan yksilölliseen hoitoon. Kolokotroni et al. ovat esittäneet vuonna 2024 useasta erillisestä mallista koostuvan hypermallin ei-pienisoluisen keuhkokarsinoman (NSCLC) ja Wilmsin tuumorin (WT) kemo- ja radioterapian tulosten ennustamiseen²⁶. Hypermallin keskuksena toimii *discrete event-discrete state* -simulaatio⁴³, joka ennustaa syöpäsolujen repopulaatiota riippuen syöpähoidon jälkeen muuttuneista tuumoriosuhteista.

Hypermalliin on mahdollista integroida hypomalleja, joiden parametrien muuttuminen esimerkiksi hoidon seurauksena muuttaa hypermallin parametrejä ja näin ennustetta. NSCLC:n ja WT:n mallintamisessa malliin kytkettiin verisuoni-, metabolinen, solusignaalintireitti- ja biomekaaninen malli perustuen näiden aiempiin yksittäisiin mekanistisiin mallinnuksiin⁴⁴⁻⁴⁷. Yhteys hypomallien kesken ja hypermallin keskukseen luotiin eriasteisilla topologioilla. Esimerkiksi verisuoni- ja metabolinen malli olivat tiukasti kytketty toisiinsa, jolloin niiden välinen kommunikaatio oli mahdollista. Kytkenällä kuvattiin tuumorin angiogeneesin yhteyttä tuumorisolujen ravintoaineiden saantiin ja sitä kautta muuttuneeseen glukoosimetaboliaan²⁶. Näin luodulla monipuolisella mallinnuksella päästään lähemmäs visioitua, koko ihmiskehon kattavaa digitaalista kaksosta⁵.

Hypomallit voivat perustua keskenään erilaisiin biologisiin datoihin. Kolokotroni et al:n hypomallit hyödynsivät laajasti potilaista saatavaa molekulaarista ja kuvantamisperäistä dataa. Potilaiden geneettinen ja proteomeettinen profiili eristettiin, jotta voitiin simuloida syövän muuttuneita metaboliareittejä. Tuumorit kuvannettiin ja kuvat segmentoitiin terveeseen ja pahanlaatuiseen kudokseen antaen samalla kudoksille niiden joustavuutta ja muuta mekaniikkaa kuvaavat parametrit, joilla ennustettiin kasvaimen mahdollista kasvusuuntaa. Mikro-RNA-profiili määrityksellä tutkittiin

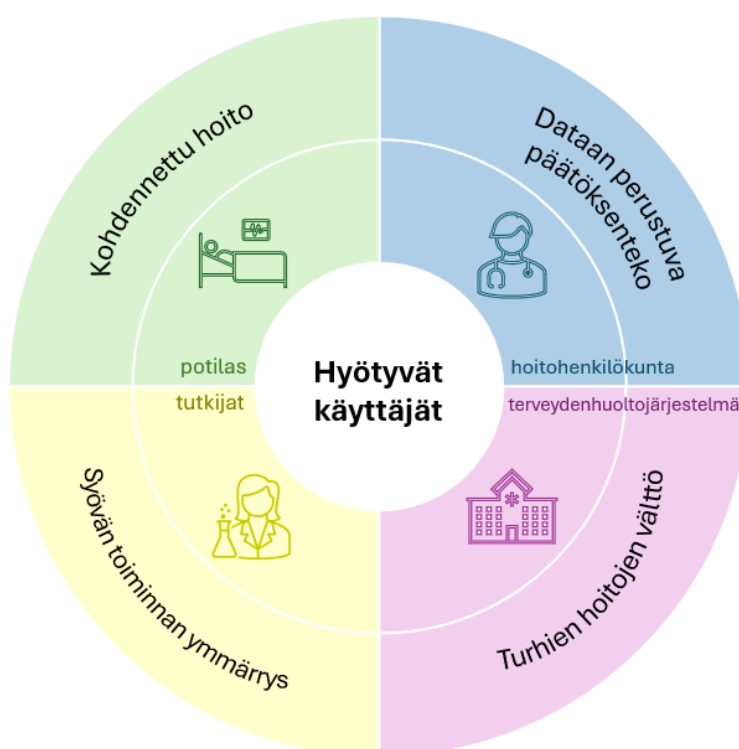
yliaktiiviset signaalointireitit, joita käytettiin määrittämään mahdolliset herkkydet eri hoitomuodoille.²⁶

Kaikki hypomallit eivät hyödyntäneet yksilöllistä ja muuttuvaa potilasdataa, vaan osa parametreistä määritettiin kirjallisuudesta tai oletettiin staattisiksi. Näin meneteltiin esimerkiksi verisuonituksen perusteella ennustettavaan tuumorin glukoosikonsentraatioon.²⁶

Hypermallin havaittiin tuottavan realistisia ennustuksia sekä NSCLC:n, että WT:n lokalisaatiosta hoidon jälkeen. Keuhkosyövän osalta hypermalli tuotti realistisia kasvuskenaarioita, kun taas WT:n tapauksessa ennustettu tuumorijäännöksi oli todellisuutta diffuusimpi. Esiitetty hypermallin konsepti olisi mahdollista ottaa käyttöön myös muiden syöpien mallinnuksessa, mutta kliiniseen käyttöön se ei ole validoitu, eikä kaikilta osin riittävän tarkka. Esimerkiksi kudoksen biomekaanisen mallin parametrit ovat vain arvioita, eivätkä potilaskohtaisia. Henkilökohtaisesta tuumorisolujen repopulaatiotaipumuksesta on myös hankala saada yksilöllistä tietoa, jonka vuoksi Kolokotroni et al. simuloivat useita mahdollisia parametriyhdistelmiä ei-potilaskohtaisilla virtuaalisilla tuumorimalleilla.²⁶

7 Digitaalisten syöpäkaksosten tulevaisuus ja haasteet

Digitaalisella syöpäkaksosella on nähty monia mahdollisuuksia uudistaa terveydenhuoltoa. Kaksonen antaisi laajat mahdollisuudet yksilöidyn lääketieteen edistämiseen sekä hoidon riskien ennustukseen ja hallintaan⁵. Kuvaan 3 on koottu konseptista mahdollisesti hyötyviä käyttäjiä. Toteutuessaan digitaalinen kaksonen hyödyttäisi potilasta mahdollistamalla tälle kohdennetun hoidon⁵, lääkäriä ja muuta hoitohenkilökuntaa antamalla klinisen päätöksenteon tukea¹², tutkijoita auttamalla ymmärtämään syövän käytöstä sekä terveydenhuoltojärjestelmää vähentämällä turhiksi osoittautuvien hoitojen määrää ja kustannuksia¹¹.



Kuva 3. Digitaalisesta syöpäkaksosesta hyötyvät käyttäjät. Mahdollisia hyötyjä syntyisi niin potilaalle, hoitohenkilökunnalle, tutkijoille ja koko terveydenhuoltojärjestelmällekin. Luotu BioRenderillä (<https://biorender.com>).

Ennen toteutumistaan konseptilla on kuitenkin haasteensa, jotka vaativat niin terveydenhuollon, teknologian kuin politiikan asiantuntijoiden yhteistyötä²⁹. Lisäksi digitaalisen kaksosen määritelmän tarkennus estäisi konseptin laajentumista ja kulumista. Mikäli määritelmä ja odotukset sen mukana laajenisivat, olisi vaarana, ettei konsepti pystyisi täyttämään niitä kaikkia ja että sen jatkokehitys pysähtyisi⁶.

Kaikki kolme digitaalisen kaksosen osaa – fyysisen todellisuuden moninaisuus, sen virtuaalinen kopio ja näiden yhteys – luovat haasteita onnistuneen syöpämallin toteuttamiseen. Haasteiden takia mitään esitellyistä digitaalisen syöpäkaksosen yrityksistä ei ole validoitu, eikä näin ollen otettu laajaan käyttöön yksilöllisen lääketieteen tarpeisiin. Etenkin terveydenhuollossa, jossa on kyse ihmisten elämänlaadun ja -ajan parantamisesta ja pidentämisestä, kaksosen simuloimiin tuloksiin olisi voitava luottaa hoitoa ohjaavana tekijänä⁵.

7.1 Fyysisen todellisuuden haasteet: ulottuvuus ja olettamukset

Syöpä on monimutkainen sairaus, jonka kokonaiseen mallinnukseen voisi liittää dataa aina molekyylitason muutoksista sen taudinkehitykseen vaikuttaviin elintapoihin asti. Mallinnukset ovat kuitenkin yksinkertaistuksia tautitilan valituista ominaisuuksista, kuten esimerkiksi lääkehoidon vaikutuksesta syöpäsolujen jakautumiseen. Näin ollen mallien ulottuvuus jää usein rajalliseksi; miten mallin voisi saada ennustamaan esimerkiksi primaarituumorialueen ulkopuolisia etäpesäkkeiden muodostumista kudoksissa, joissa solunjakautumiseen vaikuttavat toisenlaiset tekijät kuin primaarialueella?²⁴

Mallinnukset nojaavat usein olettamuksiin, jotka eivät päde jokaisen potilaan kohdalla²⁷. Luodut mallit saattavat olettaa hankalasti mitattavia parametrejä staattisiksi tai käyttää kirjallisuudessa esiintyviä arvioita, jolloin mallien kyky peilata potilaskohtaista syöpää realistisesti kärsii^{26,27}. Omiikkadataa hyödyntävissä malleissa parametrit sovittuvat solujen keskiarvoiksi, jos uudempia yksisoluomiikkateknologioita ei käytetä³⁶.

Digitaalinen kaksonen simuloisi vaihtoehtoisia tuloksia syöpähoidossa. Matemaattiseen laskentaan perustuvat mallit luovat simulaationsa oletettujen arvojen ja todennäköisyyksien perusteella, jolloin mallilla on mahdollisuus tehdä taudin käyttäytymiselle todellisuudesta poikkeava, virheellinen ennuste. On eettinen kysymys ratkaista, minkälaiset poikkeamat ja kuinka suuri virhemarginaali kliiniseen käyttöön tulevalle mallille voidaan sallia³⁵.

7.2 Virtuaalisen todellisuuden haasteet: datan määrä, datan hankinta ja laskentatapa

Yleinen ongelma, johon lääketieteellinen digitaalinen kaksonen törmää, on laskentamallin koulutukseen tarvittavan datan vähyyys. Syöpää sairastavia potilaita on tarjolla tutkimukseen rajattu määrä, osasta potilaista ei ole saatavilla pitkäaikaista seuranta dataa, eikä useassa mallissa tarvittavaa

laajaa omiikkadataa molekulaarisista fenotyypeistä kerätä terveydenhuollossa rutiinisti^{8,24,34,35}. Tämä saattaa johtaa mallin ylisovittumiseen, eli ennusteiden vinoutumiseen ja tulosten vääristymiseen, kun mallin simuloimat tulokset sovittuvat vastaamaan vain pienen potilasjoukon tilanteita. Datan vähäinen määrä on este etenkin koneoppimista hyödyntäville kaksosille; siinä missä mekanistisen mallin voi kalibroida yhden potilaan datalla, koneoppimismalli tarvitsee koulutukseensa huomattavasti isomman potilasjoukon²⁴. Tämä tekee koneoppimismalleista erityisen herkkiä ylisovittumiseen, mutta ongelmaa voisi yrittää lieventää yhdistämällä malliin mekanistisen komponentin⁸. Populaatiodataan perustuvien mallien taustalla on enemmän tilastollista voimaa, mutta ne voivat vääristää tuloksia käyttäessään laskemiseen populaatiokeskiarvoja yksittäisten potilaiden tietojen sijasta²⁶.

Datan vähyyttä voidaan pyrkiä helpottamaan augmentaatiolla. Molekulaarisbiologisissa malleissa tämä voi tarkoittaa esimerkiksi puuttuvien aikapisteiden arviointia saatavilla olevan datan perusteella³⁵ ja kuvantamismalleissa kuvien käännoiksi, rajausta tai pääkomponenttianalyysiä⁴⁸. Myös synteettistä dataa voidaan hyödyntää datan lisäämiseksi, mutta sen luomisessa on varmistettava, että luotu data vastaa uskottavasti todellista, eikä suurena alkuperäisen datan vääristymiä⁴⁹. Datan augmentaatio voi näin myös olla virhelähde mallille, jos se johtaa artefaktiin tai mallin pahempaan ylisovittumiseen^{48,49}.

Kaksosiin tarvittavat datamuodot ovat moninaisia, eikä datan hankinnalle, muodolle tai käsittelylle ole standardoituja tapoja^{8,24}. Lääketieteellinen mittausdata sisältää usein taustakohinaa ja epävarmuutta, ja kerätyn datan laatu voi vaihdella alueittain johtuen esimerkiksi sairaaloiden erilaisista mittauslaitteistoista^{8,35}. Tämä tekee yleiseen käyttöön tähtäävän digitaalisen kaksosen kehityksestä vaikeaa; datan muodon ja laadun tulisi olla yhteneväistä, jotta se toimisi samalla tavalla eri käyttöympäristöissä²⁶.

Tietokoneiden laskentateho ja laskentaan kuluva aika eivät ole enää nykyaikana kaksosten luomista rajoittava tekijä¹². Kuitenkin koska syöpä, sen käytös ja hoitovaste ovat monen biologisen tekijän summia, taudin tarkka mallinnus vaatisi mitä todennäköisimmin useamman erilaisen laskentamallin yhteen kytkemisen, mikä monimutkaistaa laskentaa²⁶. Relevantti kysymys on, onko terveydenhuollossa tarpeeksi resursseja hankkimaan ja pyörittämään kaksosta, joka vaatii usean datamuodon yhteensovittamista usealla eri laskentatavalla ja tarpeeksi nopealla tahdilla, jotta kaksosen olisi realistinen apu kliiniseen päätöksentekoon^{5,26,30}?

Erityisen haastava tekijä on fyysisen todellisuuden yhteys virtuaaliseen. Kuvantamiseen perustuvissa kaksosissa yhteyden ylläpitäminen on haastavaa kerättävän datan ollessa sellaista, joka vaatii kuvantamisoperaation tai biopsian²⁴. Molekulaarisbiologisia profiileja hyödyntäviin malleihin voidaan syöttää uutta dataa esimerkiksi verikokeiden tuloksista³⁵, mutta tällöinkin kaksosen ja potilaan välinen yhteys syntyy ainoastaan, kun mallia päivitetään lääkärin vastaanoton jälkeen. Toisaalta kuten tutkielmassa on esitetty, reaaliaikainen yhteys ei välttämättä ole tarkoituksenmukainen syöpäkaksosen osalta.

7.3 Yhteiskunnalliset haasteet: käyttöönnotto, tietoturva ja tasa-arvo

Henkilökohtaisia terveystietoja käytettäessä potilaan yksityisyyden ja tietoturvan varmistaminen on ensiarvoisen tärkeää. Tietojenkäsittelyn turvallisuus on varmistettava esimerkiksi palomuureilla, tietojen salauksella ja potilastietojen anonymisaatiolla sekä tietoja käsittelevien henkilöiden autentikaatiolla. Näin varmistetaan myös potilaan luottamuksen säilyminen ja halukkuus osallistua oman hoitonsa kulkuun²⁶. Tietojen suojaus on tärkeää etenkin sairaaloiden tehdessä yhteistyötä mallien kehittämisessä ja ylläpidossa. Tietojen käsittelyssä on noudatettava tiukkoja standardeja, kuten EU:n yleistä tietosuojasetusta, jotta potilaan tietoturva varmistetaan³⁰.

Digitaalisen kaksosen käyttöönoton tulisi tehdä terveydenhuollolle mahdollisimman helpoksi. Kaksosen käyttäminen osana kliinistä työtä ei saisi kuormittaa hoitohenkilökuntaa, vaan auttaa ja sujuvoittaa tehtävää potilastyötä. Tämä vaatii aktiivista palvelumuotoilua kaksosen käyttöliittymän osalta.²⁶

Tällä hetkellä digitaalisen kaksosen mukaanotto yksilölliseen syöpähoitoon olisi kallista terveydenhuollon toimijoille. Ihmisten tasa-arvoisen kohtelun varmistamiseksi olisi tärkeää, että teknologia ei rajautuisi vain rikkaille alueille, sairaaloille ja ihmisille, vaan että digitaalinen kaksosen olisi jokaisen siitä hyötyvän saatavilla⁵. Tasa-arvo on otettava huomioon myös kehityksen aikana; kaksosissa käytettävän datan tulee olla monipuolisesti eri potilaspopulaatiosta ja ihmisryhmistä, jotta vältetään ennusteiden vinoutumat tietyn demografisen ryhmän suuntaan⁸.

8 Yhteenveto ja johtopäätökset

Digitaalinen kaksonen on alun perin insinööritieteiden tarpeisiin kehitetty konsepti, jonka käyttö on laajentunut myös muille tieteenaloille, mukaan lukien lääketieteeseen. Digitaalinen kaksonen on virtuaalinen kopio fyysisen todellisuuden kappaleesta, jonka avulla voidaan ennustaa kappaleessa tapahtuvia muutoksia sekä tutkia kappaleeseen kohdistuvien interventioiden mahdollisia tuloksia.

Syöpä on monimutkainen sairaus, ja tautitapauksissa on suurta yksilöllistä vaihtelua. Vaihtelusta johtuen standardihoidot eivät tehoa jokaiselle potilaalle. Kehittyneiden kuvantamis- ja omiikkateknologioiden myötä on kuitenkin mahdollista mitata enenevässä määrin dataa potilaan taudin yksilöllisistä ominaisuuksista. Digitaaliset syöpäkaksoset tarjoavat mahdollisuuden entistä kokonaisvaltaisempaan yksilölliseen syöpähoitoon luomalla metodologisen alustan potilaasta mitattavan monimuotoisen tiedon hyödyntämiseen; potilaan tautia voidaan mallintaa virtuaalisesti ja ennustaa tietyn hoidon tehoa tai sen mahdollisia haittavaikutuksia.

Yhtä toteutustapaa digitaaliselle syöpäkaksoselle ei ole, vaan kaksonen luodaan sen arvion perusteella, minkä syövän ominaisuuden ennustaminen tuottaisi potilaalle hyötyä. Tässä työssä digitaaliset kaksoset luokiteltiin niiden hyödyntämisen datasyötteen sekä laskentamallin perusteella. Mekanististen mallien havaittiin olevan vähemmän alttiita ylisovittumiseen, kun taas tekoälyä hyödyntävien mallien edun havaittiin olevan yhteyksien havaitseminen suurissa dataseteissä. Molemmat laskentamalliluokat kärsivät etenkin datan vähyydestä.

Lupaavia syöpäkaksosia on kehitetty, mutta toistaiseksi yhtäkään ei ole validoitu. Alkuperäisen digitaalisen kaksosen määritelmän integrointi lääketieteeseen tuottaa haasteita etenkin fyysisen ja virtuaalisen todellisuuden yhteyden luomisessa, joka on merkittävin erottava tekijä digitaalisen kaksosen ja virtuaalisen mallinnuksen välillä. Yksilöidyssä digitaalisessa syöpäkaksosessa on kuitenkin perusteltua kyseenalaistaa yhteyden nopeuden vaatimukset; tautitilassa ei tapahdu reaaliaikaista reagoitua vaativia, kliinisesti merkittäviä muutoksia. Onko tällöin kyse vain virtuaalisesta mallista vai digitaalisesta kaksosesta? Kyse on rajanvedosta, joka on tehtävä termien erottamiseksi.

Digitaalisen syöpäkaksosen kliininen käyttöönotto vaatii vielä teknologista kehittymistä, jotta se tarjoaisi lisähyötyä nykyisten hoito- ja diagnoosikeinojen rinnalle. Ennen kaikkea konseptin vakiintuminen lääketieteeseen edellyttää standardoituja tapoja kerätä ja käsitellä dataa, jotta kaksosten validiteetti voidaan varmistaa. Toteutuessaan digitaaliset syöpäkaksoset edustaisivat merkittävää edistysaskelta dataan perustuvan yksilöllisen hoidon saralla. Ylipäätään digitaalisten

kaksosten kehitystyö edistää dataan ja laskentamalleihin perustuvaa lääketiedettä, tuli teknologia käyttöön millä termillä tahansa.

Lähdeluettelo

1. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **74**, 229–263 (2024).
2. International Agency for Research on Cancer. Estimated numbers from 2022 to 2050, Males and Females, age [0-85+].
https://gco.iarc.who.int/tomorrow/en/dataviz/trends?multiple_populations=1. Viitattu 30.12.2025
3. Pulumati, A., Pulumati, A., Dwarakanath, B. S., Verma, A. & Papineni, R. V. L. Technological advancements in cancer diagnostics: Improvements and limitations. *Cancer Rep.* **6**, e1764 (2023).
4. Wu, C. *et al.* Integrating mechanism-based modeling with biomedical imaging to build practical digital twins for clinical oncology. *Biophys. Rev.* **3**, 021304 (2022).
5. Sun, T., He, X., Song, X., Shu, L. & Li, Z. The Digital Twin in Medicine: A Key to the Future of Healthcare? *Front. Med.* **9**, 907066 (2022).
6. Wright, L. & Davidson, S. How to tell the difference between a model and a digital twin. *Adv. Model. Simul. Eng. Sci.* **7**, 13 (2020).
7. Glaessgen, E. & Stargel, D. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference* (American Institute of Aeronautics and Astronautics). doi:10.2514/6.2012-1818.
8. Gaw, N. *et al.* Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Sci. Rep.* **9**, 10063 (2019).
9. Subramanian, K. Digital Twin for Drug Discovery and Development—The Virtual Liver. *J. Indian Inst. Sci.* **100**, 653–662 (2020).
10. An, G. & Cockrell, C. Drug Development Digital Twins for Drug Discovery, Testing and Repurposing: A Schema for Requirements and Development. *Front. Syst. Biol.* **2**, 928387 (2022).
11. Liu, Y. *et al.* A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin. *IEEE Journals & Magazine.* **7**, 49088-49101 (2019)

12. Tardini, E. *et al.* Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad. *J. Med. Internet Res.* **24**, e29455 (2022).
13. Brown, S. A. *et al.* Six-Month Randomized, Multicenter Trial of Closed-Loop Control in Type 1 Diabetes. *N. Engl. J. Med.* **381**, 1707–1717 (2019).
14. Chakshu, N. K., Sazonov, I. & Nithiarasu, P. Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis. *Biomech. Model. Mechanobiol.* **20**, 449–465 (2021).
15. He, X. *et al.* Towards a shape-performance integrated digital twin for lumbar spine analysis. *Digital Twin.* **2** (2025).
16. Aubert, K. *et al.* Development of Digital Twins to Optimize Trauma Surgery and Postoperative Management. A Case Study Focusing on Tibial Plateau Fracture. *Front. Bioeng. Biotechnol.* **9**, 722275 (2021).
17. Jung, A., Gsell, M. A. F., Augustin, C. M. & Plank, G. An Integrated Workflow for Building Digital Twins of Cardiac Electromechanics—A Multi-Fidelity Approach for Personalising Active Mechanics. *Math. Basel Switz.* **10**, 823 (2022).
18. Baillargeon, B., Rebelo, N., Fox, D. D., Taylor, R. L. & Kuhl, E. The Living Heart Project: A robust and integrative simulator for human heart function. *Eur. J. Mech. Solids* **48**, 38–47 (2014).
19. Schärfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O. & Marks, D. S. Genetic variation in human drug-related genes. *Genome Med.* **9**, 117 (2017).
20. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32**, 644–652 (2014).
21. Hao, J.-J. *et al.* Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. *Nat. Genet.* **48**, 1500–1507 (2016).
22. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

23. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
24. Wu, C. *et al.* MRI-based digital models forecast patient-specific treatment responses to neoadjuvant chemotherapy in triple-negative breast cancer. *Cancer Res.* CAN-22-1329 (2022) doi:10.1158/0008-5472.CAN-22-1329.
25. Jarrett, A. M. *et al.* Evaluating patient-specific neoadjuvant regimens for breast cancer via a mathematical model constrained by quantitative magnetic resonance imaging data. *Neoplasia N. Y. N* **22**, 820–830 (2020).
26. Kolokotroni, E. *et al.* A Multidisciplinary Hyper-Modeling Scheme in Personalized In Silico Oncology: Coupling Cell Kinetics with Metabolism, Signaling Networks, and Biomechanics as Plug-In Component Models of a Cancer Digital Twin. *J. Pers. Med.* **14**, (2024).
27. Zhang, Y. *et al.* Digital twin models for predicting venetoclax and azacitidine-induced neutropenia in patients with acute myeloid leukemia. *Npj Digit. Med.* **8**, 596 (2025).
28. VanDerHorn, E. & Mahadevan, S. Digital Twin: Generalization, characterization and implementation. *Decis. Support Syst.* **145**, 113524 (2021).
29. Katsoulakis, E. *et al.* Digital twins for health: a scoping review. *Npj Digit. Med.* **7**, 77 (2024).
30. Moztarzadeh, O. *et al.* Metaverse and Healthcare: Machine Learning-Enabled Digital Twins of Cancer. *Bioengineering* **10**, (2023).
31. Kleczka, A., Dzik, R. & Kabała-Dzik, A. Digital Twin Meets the Bench: Natural Compounds Reshape the Ovarian Cancer Microenvironment. *Biomedicines* **13**, (2025).
32. Li, S., Xu, L. D. & Zhao, S. The internet of things: a survey. *Inf. Syst. Front.* **17**, 243–259 (2015).
33. Jara, A. J., Zamora-Izquierdo, M. A. & Skarmeta, A. F. Interconnection Framework for mHealth and Remote Monitoring Based on the Internet of Things. *IEEE J. Sel. Areas Commun.* **31**, 47–65 (2013).

34. Ding, M. Q., Chen, L., Cooper, G. F., Young, J. D. & Lu, X. Precision Oncology Beyond Targeted Therapy: Combining Omics Data with Machine Learning Matches the Majority of Cancer Cells to Effective Therapeutics. *Mol. Cancer Res. MCR* **16**, 269–278 (2018).
35. Hou, J., Deng, J., Li, C. & Wang, Q. Tracing and Forecasting Metabolic Indices of Cancer Patients Using Patient-Specific Deep Learning Models. *J. Pers. Med.* **12**, (2022).
36. Filippo, M. D. *et al.* Single-cell Digital Twins for Cancer Preclinical Investigation. in *Metabolic Flux Analysis in Eukaryotic Cells: Methods and Protocols* (ed. Nagrath, D.) 331–343 (Springer US, New York, NY, 2020). doi:10.1007/978-1-0716-0159-4_15.
37. Langthaler, S. *et al.* A549 in-silico 1.0: A first computational model to simulate cell cycle dependent ion current modulation in the human lung adenocarcinoma. *PLoS Comput. Biol.* **17**, e1009091 (2021).
38. Lao, J. *et al.* A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci. Rep.* **7**, 10353 (2017).
39. Spring, L. M. *et al.* Pathological complete response after neoadjuvant chemotherapy and impact on breast cancer recurrence and survival: a comprehensive meta-analysis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **26**, 2838–2848 (2020).
40. Musall, B. C. *et al.* Functional Tumor Volume by Fast Dynamic Contrast-Enhanced MRI for Predicting Neoadjuvant Systemic Therapy Response in Triple-Negative Breast Cancer. *J. Magn. Reson. Imaging JMRI* **54**, 251–260 (2021).
41. Schulte, P. J., Tsiatis, A. A., Laber, E. B. & Davidian, M. Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Stat. Sci. Rev. J. Inst. Math. Stat.* **29**, 640–661 (2014).
42. Neal, M. L. *et al.* Discriminating Survival Outcomes in Patients with Glioblastoma Using a Simulation-Based, Patient-Specific Response Metric. *PLoS ONE* **8**, e51951 (2013).
43. Stamatakos, G. S., Kolokotroni, E. A., Dionysiou, D. D., Georgiadi, E. Ch. & Desmedt, C. An advanced discrete state–discrete event multiscale simulation model of the response of a solid tumor to chemotherapy: Mimicking a clinical study. *J. Theor. Biol.* **266**, 124–139 (2010).

44. Hahnfeldt, P., Panigrahy, D., Folkman, J. & Hlatky, L. Tumor Development under Angiogenic Signaling: A Dynamical Theory of Tumor Growth, Treatment Response, and Postvascular Dormancy.
45. Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R. & Ruppin, E. Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect. *PLOS Comput. Biol.* **7**, e1002018 (2011).
46. Chen, W. W. *et al.* Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular Systems Biology.* **5**, MSB200874 (2009)
47. Choi, M. *et al.* Attractor Landscape Analysis Reveals Feedback Loops in the p53 Network That Control the Cellular Response to DNA Damage. *Science Signaling.* **5**, ra83 (2012).
48. Hussain, Z., Gimenez, F., Yi, D. & Rubin, D. Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. *AMIA. Annu. Symp. Proc.* **2017**, 979–984 (2018).
49. Giuffrè, M. & Shung, D. L. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *Npj Digit. Med.* **6**, 186 (2023).