Tampereen yliopisto

Vilma Lehto

YKSITYISYYDEN SUOJAAMINEN HAJAUTETUSSA OPPIMISESSA JA HAJAUTETUN OPPIMISEN SOVELTAMINEN DIGIPATOLOGIAAN

Kirjallisuuskatsaus

Kandidaatintyö
Lääketieteen ja terveysteknologian tiedekunta
Pekka Ruusuvuori
08 / 2023

TIIVISTELMÄ

Vilma Lehto: Yksityisyyden suojaaminen hajautetussa oppimisessa ja hajautetun oppimisen soveltaminen digipatologiaan
Kandidaatintyö
Tampereen yliopisto
Bioteknologian ja biolääketieteen tekniikan kandidaattiohjelma
Elokuu 2023

Keräämällä ja tilastoimalla potilaista kerättyä lääketieteellistä dataa on tehty suuria lääketieteellisiä läpimurtoja. Tällä hetkellä lääketieteellisen ja biologisen tiedon määrä kasvaa jatkuvasti. Samaan aikaan koneoppimisen ja syväoppimisen menetelmien yleistyessä on kerättävän datan potentiaali suuri. Kuitenkaan tällä hetkellä kerättyä tietoa ei päästä hyödyntämään kunnolla sen tarkan säätelyn takia. Euroopan tietosuoja-asetus ja Suomen tietosuojalaki määrittävät tarkat kriteerit henkilökohtaisen datan kuten terveysdatan käsittelylle. Hajautettu oppiminen tarjoaa mahdollisuuden terveysdatan käsittelyyn ilman potilaiden yksityisyydensuojan vaarantamista.

Hajautettu oppiminen on koneoppimisen menetelmä. Poiketen perinteisestä koneoppimisesta opetusdataa ei keskitetä tai jaeta muille osallistujille. Opettaminen tapahtuu jokaisen osallistujan omalla palvelimella ja omalla datalla. Tällöin osallistujan datan ei tarvitse poistua sen palvelimelta ollenkaan. Kun jokainen osallistuja on opettanut mallin, yhdistetään ne yhdeksi yleiseksi malliksi. Menetelmän ominaisuudet ovat herättäneet tutkijoiden mielenkiinnon ja sen soveltamista terveydenhuoltoon on alettu tutkia.

Vaikka hajautetussa oppimisessa ei tarvitse jakaa dataa on menetelmässä silti tietosuojariskejä. Opetettujen mallien siirtäminen palvelimien välillä voi vaarantaa potilaiden yksityisyyden. Opetetusta mallista on mahdollista saada tarkkaa tietoa organisaation potilaista ja jopa tunnistaa yksittäisiä potilaita. Tämän takia on hajautetun oppimisen menetelmiin aina sovellettava erilaisia tietosuojaa parantavia toimenpiteitä.

Tietosuojamenetelmistä työssä käydään läpi henkilötietojen poistamisen menetelmät, joihin kuuluu anonymisointi ja pseudonymisointi. Niissä datasta poistetaan tunnistetiedot, mutta ne eivät silti anna merkittäviä tietosuojatakuita. Häiriömenetelmät, joiden ydinideana on häiriön lisääminen dataan, parantavat tietosuojaa mutta heikentävät datan laatua. Näiden lisäksi käsitellään myös erilaisia kryptografisia menetelmiä, jotka perustuvat datan salaamiseen. Nämä suojaavat dataa hyvin ulkopuolisilta tietosuojariskeiltä, mutta kasvattavat ohjelman vaatimaa laskentatehoa. Kun jokaisella tietosuojamenetelmällä on hyvät ja huonot puolensa, käsitellään työssä myös menetelmien yhdistämistä.

Työssä perehdytään hajautetun oppimisen soveltamiseen digipatologiassa. Menetelmän soveltaminen alaan vaatii hieman datan esikäsittelyä, mutta sitä on onnistuttu soveltamaan moneen eri tehtävään. Vaikka tutkitut menetelmät eivät aivan yllä samalla tasolle kuin perinteiset keskitetyt koneoppimisen menetelmät ovat tulokset silti hyvin lupaavia. Verrattuna yhden laitoksen toteuttamaan koneoppimisen malliin hajautetun oppimisen mallit ovat usein tarkempia.

Avainsanat: hajautettu oppiminen, koneoppiminen, yksityisyydensuoja, digipatologia

ALKUSANAT

Tämä työ on tehty Tampereen yliopistolle osana Bioteknologian ja biolääketieteen tekniikan kandidaatinopintojani.

Haluaisin kiittää ohjaajani Pekka Ruusuvuorta, jolta sain mielenkiintoisen aiheen työhön. Haluaisin kiittää häntä myös saamastani tuesta ja ohjauksesta. Kiitokset ansaitsevat myös perhe ja läheiseni, jotka ovat tukeneet minua koko prosessin ajan. Erityisesti ystäviltä saatu vertaistuki ja kannustus olivat korvaamaton apu prosessissa.

Tampereella, 30.8.2023

Vilma Lehto

SISÄLLYSLUETTELO

1. JOHDANTO

Biologisen ja lääketieteellisen datan määrä kasvaa jatkuvasti [1]–[7]. Dataa kerätään esimerkiksi apteekkien, sairaaloiden ja potilaan toimesta [7]. Tällaisesta datamäärästä olisi mahdollista saada paljon uutta lääketieteellistä tietoa. Keräämällä ja yhdistämällä potilasdataa onkin tehty useita lääketieteellisisä läpimurtoja, kuten tupakoinnin yhdistäminen sydän- ja verisuonisairauksiin sekä syöpään. Suuri datamäärä ehkäisee myös epätasa-arvoa ja mahdollistaa tarkemman diagnosoinnin ja paremman hoidon suunnittelun ja arvioinnin. [8] Tällä hetkellä kuitenkin haasteena on, että data ei ole vapaasti saatavilla. Se on hajanaisesti eri laitosten palvelimilla ja sen jakamista säädellään tiukasti sen arkaluontoisuuden takia [1]–[7]. Yksityisyyden suojan lisäksi haasteita aiheuttaa myös datan keräämiseen, käsittelyyn ja varastointiin tarvittavat resurssit [5], [6], [9].

Viime vuosina lääketieteessä ja tutkimuksissa on alettu soveltamaan myös koneoppimisen ja syväoppimisen menetelmiä. Menetelmiä käytetään esimerkiksi päätöstukijärjestelmissä radiologian ja patologian osastoilla. Koneoppimisen menetelmät mahdollistavat tarkkojen ja luotettavien tilastollisten mallien tekemisen ja uudet lääketieteelliset sovellukset pohjautuvatkin nykyään pitkälti niihin. [5], [6] Toimiakseen kunnolla koneoppimisen menetelmät tarvitsevat suuren määrän dataa malliensa opettamiseen. Kuitenkaan tällä hetkellä olemassa olevaa lääketieteellistä dataa ei päästä kunnolla hyödyntämään, sen tiukan säätelyn takia. [1], [3], [6] Datan kerääminen yhteen paikkaan loisi etnisiä, laillisia, regulaattorisia ja teknisiä haasteita [5], [6]. Datan jakamisen rajoittaminen on aiheuttanut sen, että koneoppimisen menetelmistä ei saada kaikkea mahdollista hyötyä irti. Ratkaisuksi tilanteeseen on ehdotettu hajautetun oppimisen soveltamista lääketieteen aloille kuten kuvantamiseen. [1], [3]

Hajautettu oppiminen on menetelmä, jossa osallistujat kouluttavat yhteisen koneoppimisen mallin kuitenkaan jakamatta dataa toisilleen tai keskittämättä sitä mihinkään. Jokainen osallistuja kouluttaa mallia paikallisesti omalla palvelimellaan, jonka jälkeen mallit yhdistetään. [5], [6], [10]–[12] Näin saadaan käyttöön useamman laitoksen dataa ja mallista voidaan luoda kattavampi ilman osallistujien yksityisyydensuojan vaarantamista [7], [13]. Yksityisyyden suojaamisen lisäksi hajautettu oppiminen mahdollistaa paremman laskentatehon ja tehokkaamman mallien opetuksen [5].

Työn tavoitteena on perehtyä hajautettuun oppimiseen, sen toteuttamiseen ja sen riskeihin. Sillä vaikka dataa ei jaeta muilla osallistujille, on menetelmässä silti riskinsä [11]. Paikallisten mallien paljastuessa on niistä mahdollista saada selville arkaluontoista potilasdataa ja jopa tunnistaa potilaita [14]. Työ tarkoituksena on myös tutustua toimenpiteisiin, joilla potilaiden yksityisyys voidaan turvata ja tietosuojaa voidaan kasvattaa. Lopuksi työssä tutkitaan myös menetelmän soveltamista digipatologiaan.

Työn toisessa luvussa perehdytään hajautettuun oppimiseen ja sen erilaisiin malleihin tarkemmin. Tämän jälkeen käsitellään regulaatioita ja niiden vaikutusta lääketieteellisendatan käyttöön. Luvussa viisi käydään läpi erilaisia menetelmiä tietosuojan parantamiseen ja luvussa kuusi tarkastellaan, kuinka hajautettua oppimista voidaan soveltaa patologiaan. Viimeinen luku on työn yhteenveto.

2. HAJAUTETTU OPPIMINEN

Hajautettu oppiminen oli ensin Googlen suunnittelema menetelmä ja se oli tarkoitettu toimimaan eri puolilla maailmaa olevien puhelimien välillä [10], [15]. Menetelmässä osallistujat opettavat yhteisen koneoppimisen mallin jakamatta dataansa muille osallistujille tai keskittämättä sitä [1]–[6], [9]–[12], [14], [16]. Tällöin data pysyy koko ajan osallistujan omalla palvelimella. Sitä ei tarvitse siirtää järjestelmän palomuurien ulkopuolelle, mikä parantaa menetelmän tietoturvaa. [5], [6], [14] Menetelmän ydinideana on tehokkaan koneoppimisen mallin muodostaminen kuitenkin samalla suojaten käyttäjän yksityisyyttä [15].

Hajautettu oppiminen toteutetaan tyypillisesti muutamalla toistuvalla vaiheella. Ensin alustetaan opetettava malli. Tässä vaiheessa valitaan mikä koneoppimisen malli opetetaan. Yleensä käytetään neuroverkkoa. [1], [5], [9], [15] Mallin valinnan ja alustamisen lisäksi määritetään, montako osallistujaa tarvitaan yhdelle opetuskierrokselle, montako kierrosta tehdään ja milloin opetus on valmis. Kun alustus on valmis, malli jaetaan osallistujille. [4] Alustuksen jälkeen tapahtuu paikallinen opettaminen. Siinä osallistujat opettavat saamaansa mallia omalla datallaan ja päivittävät sen. Kun opetus on valmis, osallistujat lähettävät päivittämänsä mallit yhdistettäväksi. Paikalliset mallit yhdistetään yhdeksi yleiseksi malliksi ja päivitetty malli lähetetään osallistujille. Paikallisen opetuksen ja mallien yhdistämisen vaiheita toistetaan, kunnes alussa määritelty lopetusehto on saavutettu. Lopuksi kaikille osallistujille jaetaan viimeisin yleinen malli. [4], [10]

Matemaattisesti hajautettu oppiminen voidaan ilmaista funktiolla

$$\min_{\emptyset} L(\mathrm{X};\emptyset), \tag{1}$$

missä

$$L(X;\emptyset) = \sum_{i=1}^{N} w_i L_i(X_i;\emptyset). \tag{2}$$

Kaavoissa (1) ja (2) *L* kuvaa kokonaishäviöfunktiota eli tarkkuuden häviämistä huomioiden kaikki *N* osallistujaa ja koko opetusdatan *X*. Vastaavasti kaavassa (2) L_i kuvaa osallistujan *i* häviöfunktiota ja on laskettu käyttäen osallistujan *i* yksityistä dataa X_i. Kun

jokainen osallistuja on laskenut oman häviöfunktionsa, lasketaan nämä yhteen jokaisen osallistujan painokerroin w_i huomioiden. Tavoitteena on löytää yleinen malli, jonka tarkkuuden häviäminen on mahdollisimman vähäistä. [1]–[3], [6], [7], [13], [15]

Käytännössä osallistujat opettavat mallia muutaman kierroksen ennen kuin jakavat sen yhdistettäväksi [6], [13]. Yhdistämiseen ja häviöfunktioiden laskemiseen on monia erilaisia tapoja, joista McMahanin esittelemä FedAvg [13] on yleisin käytössä oleva malli, sen skaalautuvuuden ja vertailukelpoisen suorituskyvyn vuoksi [3], [5]. FedAvg-mallissa osallistujat päivittävät mallin parametrit gradienttilaskulla käyttäen omaa yksityistä dataansa. Paikallisen mallin päivitys voidaan tehdä useamman kerran ennen yhdistämistä. Yhdistäminen tehdään kaavan

$$L(X;\ \emptyset) = \frac{1}{N}\sum_{i=1}^{N} L_i(X_i;\emptyset) \tag{3}$$

mukaan. Siitä nähdään, että FedAvg-mallin muodostama yleinen malli on keskiarvo paikallisista malleista. [13] Useamman opetuskierroksen jälkeen näin saadaan kuitenkin aikaiseksi tarkka yleinen malli, jolla on hyvä yleistyskyky. Se mahdollistaa uusien osallistujien ottamisen mukaan jopa opettamisen aloittamisen jälkeen. [15]

Verrattuna perinteisiin koneoppimisen malleihin, joissa opetus tapahtuu yhdessä sijainnissa isosta määrästä keskitettyä dataa, hajautettu oppiminen saavuttaa hyvinkin vertailukelpoisia tuloksia [2], [3], [5], [6], [12]. Singh ym. tutkivat koronainfektion tunnistamista rintakehän röntgenkuvista. Heidän tutkimuksessansa perinteinen keskitetyn koneoppimisen malli saavutti 96 % tarkkuuden ja hajautetun oppimisen malli saavutti 93 % tarkkuuden. [10] Onnistuessaan hajautetun oppimisen mallit ovatkin tarkkoja ja kilpailukykyisiä [1], [2], [6], [9].

Kilpailukykyisen suoriutumisen ja paremman yksityisyydensuojan lisäksi hajautettu oppiminen säästää muistia [3], [6], [14] ja parantaa laskentatehoa [5]. Kun opettaminen tapahtuu sillä palvelimella, jolla data on, dataa ei tarvitse kopioida muualle. Tämä vähentää tarvittavan muistin määrää. [3], [14] Laskentatehon kasvaminen johtuu myös paikallisesta opetuksesta. Se mahdollistaa opetuksen tapahtumisen samanaikaisesti usealla eri palvelimella. Tällöin päästään hyödyntämään usean palvelimen laskentatehoa, eikä tarvitse kuormittaa vain yhtä palvelinta. [5]

Hajautetulla oppimisella on myös haasteensa. Esimerkiksi datan heterogeenisyys ja epätasainen jakautuminen osallistujien välillä aiheuttavat mallin tarkkuuden heikkenemistä. Ne voivat aiheuttaa jonkin organisaation liiallista painottamista tai luoda tilanteen, jossa yleinen malli ei ole kaikille osallistujille optimaalinen. Erityisesti terveydenhuollossa eri organisaatioilla käytössä olevat eri laitteet ja toimintatavat aiheuttavat

vaihtelua dataan. [6] Datan epätasaista jakautumista osallistujien välillä aiheuttaa erot datan määrässä osallistujien kesken [7]. Esimerkiksi suuren sairaalan ja paikallisen terveyskeskuksen potilas- ja datamäärät voivat erota paljonkin.

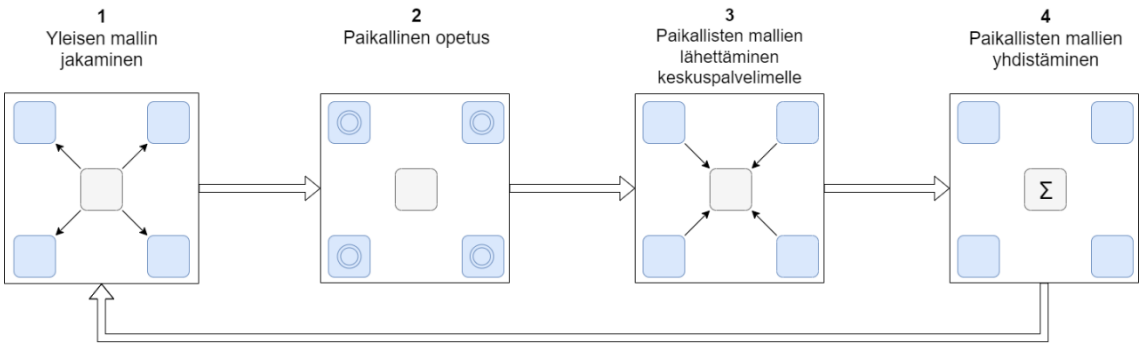
Toisena suurena haasteena hajautetussa oppimisessa on palvelimien väliseen kommunikaatioon kuluvat resurssit [6]. Osallistujien yhteyksien mahdollinen epävarmuus ja hitaus lisäävät kommunikaation haasteita [7]. Kommunikaation haasteet aiheuttavat sen hidastumista ja vievät resursseja muilta ohjelman toiminnoilta. Ne voivat johtaa koko ohjelman hidastumiseen. [17]

Kolmas menetelmän suurista haasteista on tietoturva. Vaikka hajautetun oppimisen tavoitteena on nimenomaan datan jakamisen tarpeen poistaminen ja siten yksityisyyden turvaaminen on siinä vielä haasteita. Tietosuojariskin aiheuttaa opetettavan mallin tai sen parametrien siirtäminen. Mallin parametreissa voi siirtyä dataa, joka vaarantaa potilaiden yksityisyyden. [6], [7] Tämä aiheuttaa ongelmia, sillä etenkin terveydenhuollossa potilaan yksityisyyden turvaaminen on tärkeää.

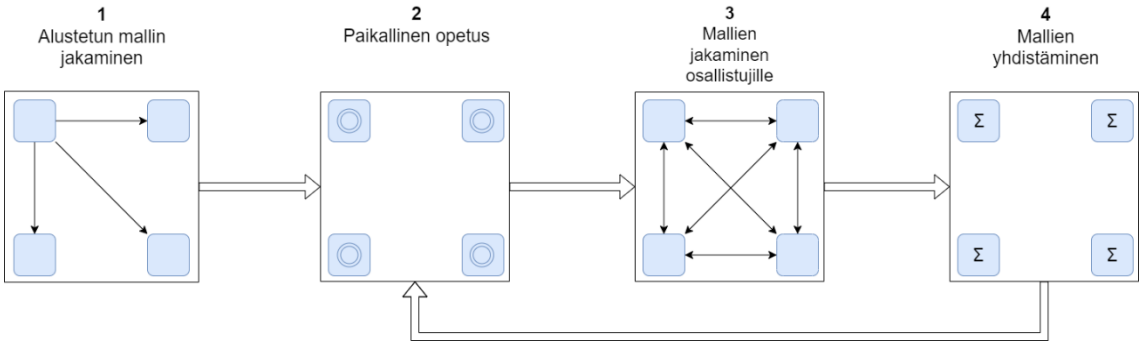
2.1 Erilaiset toteutustavat

Hajautetun oppimisen mallit voidaan jakaa erilaisiin ryhmiin niiden ominaisuuksien perusteella. Arkkitehtuurin perusteella mallit voidaan jakaa keskuspalvelimen avulla toimiviin keskitettyihin malleihin ja ilman keskuspalvelinta toimiviin ei-keskitettyihin malleihin [5], [7], [10]. Arkkitehtuurin lisäksi hajautetun oppimisen menetelmät voidaan luokitella niiden ajankäytön [1] ja opetusdatan perusteella [4], [15], [17], [18].

Keskitetyssä mallissa keskuspalvelin valitsee opetuskierroksen osallistujat, yhdistää paikalliset mallit ja välittää päivitetyn mallin kaikille osallistujille [5], [7]. Esimerkiksi FedAvg-malli on suunniteltu keskitetyksi malliksi. Keskistetyissä malleissa keskuspalvelin mahdollistaa sen, että osallistujat voivat pysyä toisilleen täysin tuntemattomina. [6] Ei-keskitetyssä mallissa osallistujat jakavat opettamansa mallit suoraa muille osallistujille, joko kaikille tai vaan muutamalle. Osallistujat itse yhdistävät saamansa mallit, minkä jälkeen ne jatkavat opetusta. Tällöin ulkopuolista keskuspalvelinta ei tarvita ollenkaan. [5], [7], [10] Kuva 1 havainnollistaa keskitetyn mallin toimintaa ja kuva 2 ei-keskitetyn mallin toimintaa. Tässä työssä käsitellään pääasiassa keskuspalvelimen avulla toimivia malleja.



Kuva 1**: Keskitetyn mallin toiminta, kuva tehty mukaillen lähdettä **[6]



Kuva 2**: Ei-keskitetyn mallin toiminta, kuva tehty mukaillen lähdettä **[6]

Toinen tapa luokitella, hajautetun oppimisen toteutukset, perustuu mallien ajan käyttöön. Hajautetun oppimisen mallit voidaan luokitella synkronoituihin, osittain synkronoituihin ja ei-synkronoituihin malleihin [1]. Synkronoidussa mallissa osallistujien mallit päivitetään yhdeksi malliksi vasta, kun kaikki ovat saaneet opetuksensa valmiiksi. Synkronoiduissa malleissa osa palvelimista saattaa siis vain odottaa muiden palvelimien opetuksen valmistumista. [1], [2]

Ei-synkronoidussa opetuksessa paikallisen mallin yhdistäminen yleiseen malliin tapahtuu saman tien opetuksen päätyttyä [1], [2]. Tällöin palvelimet ovat siis koko ajan käytössä eivätkä ”seiso tyhjänä”. Verrattuna synkronoituun malliin ei-synkronoitu malli käyttää resursseja paremmin, mutta sen tuottaman yleisen mallin tarkkuus ei ole yhtä hyvä kuin synkronoidussa mallissa [1].

Mallien välimaastoon on myös muodostunut osittain synkronoitu malli. Siinä osallistujat, jotka suorittavat opetuksen nopeammin, tekevät useamman opetuskierroksen kuin isommat ja hitaammat osallistujat. Tällöin opetus aika määräytyy hitaimman mukaan ja mikään palvelin ei vain odota muita. Mallien yhdistäminen tapahtuu synkronoidusti, kun kaikki ovat saaneet opetuksen valmiiksi. [1]

Hajautetun oppimisen menetelmät voidaan jaotella myös niiden käyttämän opetusdatan perusteella. Jos data on rakenteeltaan ja ominaisuuksiltaan samankaltaista, mutta

datalla on eri tunnisteet, on kyse pystysuoraan hajautetusta oppimisesta. [4], [15], [17], [18] Tällainen malli muodostuu yleensä, kun osallistujat toimivat samalla alalla [4], [10], [18]. Esimerkiksi lääketieteellisessä kuvantamisessa potilaista otetut kuvat ovat rakenteeltaan ja ominaisuuksiltaan samanlaisia, mutta kuvat ovat eri potilaista, jolloin niillä on eri tunnisteet.

Kun opetuksessa käytettävä data ei ole rakenteeltaan tai ominaisuuksiltaan samanlaista, mutta datan tunnisteet ovat identtiset, on malli vaakasuoraan hajautetun oppimisen mukainen [4], [10], [18]. Terveydenhuollosta tästä esimerkkinä toimii tilanne, jossa hajautettua oppimista sovelletaan terveyskeskuksen ja yksityisen terveysaseman välillä. Tällöin samasta potilaasta, samalla tunnisteella, voi löytyä useita merkintöjä, jotka ovat rakenteeltaan erilaisia, kuten kuvat, vapaamuotoinen teksti ja diagnoosikoodit.

Viimeisenä menetelmänä hajautettu oppiminen voidaan toteuttaa yhdistämällä se siirrosoppimiseen, jolloin kyse on hajautetusta siirrosoppimisesta. Siinä tavoitteena on soveltaa opittua tietoa eri lähteestä olevaan dataan [15]. Tällaisessa tilanteessa opetuksessa käytetty data eroaa sekä ominaisuuksiltaan, että tunnisteella. [4], [15], [17], [18] Näistä menetelmistä parhaiten kuva-analyysiin sopii pystysuoraan hajautetun oppimisen malli, ja tässä työssä keskitytäänkin jatkossa kyseiseen malliin.

3. REGULAATIOT

General Data Protection Regulation (GDPR) on EU:n tietosuoja-asetus, joka määrittää EU alueella asuvien ja matkustavien kansalaisten henkilötietojen käsittelyn ja siirtämisen säännöt. Asetus koskettaa kaikkea automaattista ja manuaalista henkilötietojen käsittelyä. Käsittelyksi lasketaan muun muassa tietojen kerääminen, tallentaminen, varastointi ja käyttäminen sekä poistaminen. Toisin sanoen asetus koskettaa kaikkea henkilötietoihin liittyvää toimintaa. [19] Tästä syystä tietosuoja-asetuksen tunteminen on tärkeää tutkiessa hajautetun oppimisen soveltamista terveydenhuoltoon.

EU-tason lisäksi myös Suomella on oma tietosuojalakinsa. Lakia sovelletaan, mikäli rekisterinpitäjän toimipiste sijaitsee Suomessa. Tietosuoja-asetus luo pohjan kaikelle henkilötietojen käsittelylle ja laki vain täydentää ja tarkentaa sitä tarvittaessa. [20] Tässä työssä käsitellään tarkemmin tietosuoja-asetusta ja tietosuojalaista nostetaan esille vain kohdat, jotka liittyvät hajautettuun oppimiseen tai terveystietojen käsittelyyn.

Tietosuojaa koskevien lakien lisäksi terveydenhuollon ohjelmia tutkiessa on tunnettava muitakin säädöksiä. Lääketieteen ohjelmisto saattaa täyttää lääketieteellisen laitteen määritelmän. Tällöin on huomioitava myös EU:n lääkinnällisiä laitteita koskeva asetus 2017/745 (MD-asetus). Jos ohjelmistoa käytetään diagnostiikkaan, saattaa se myös täyttää *in vitro* -diagnostiikkaan tarkoitettuja laitteiden määritelmän. Näistä laitteista säädetään EU:n asetuksessa 2017/746 (IVD-asetus). Ohjelmistolta, joka täyttää lääkinnällisen laitteen määritelmän, vaaditaan CE-merkintä. Tietosuoja-asetuksen ja -lain lisäksi luvussa käsitellään lääkinnällisten laitteiden säädökset ja hieman CE-merkintää. [21], [22]

3.1 Tietosuoja-asetus ja tietosuojalaki

Tietosuoja-asetus on annettu toukokuussa 2018 ja vaikka se tekee henkilötietojen käsittelystä ja jakamisesta hieman vaikeampaa [5] on sen tarkoituksena parantaa henkilötietojen suojaa [19]. Suomen tietosuojalaki on päivitetty tietosuoja-asetuksen perusteella ja se astui voimaan vuoden 2019 alussa [20]. Haasteita henkilötietojen suojeluun on tuonut teknologian kehitys ja globalisaatio, ja näihin päivitetyn tietosuoja-asetuksen onkin tarkoitus vastata. Asetuksen tavoitteena on turvata yksityishenkilöiden henkilötieto-

jen käsittelyyn liittyviä perusoikeuksia. Lisäksi tavoitteena on yhtenäistää EU-maiden tietosuojasäätelyä ja edistää EU-maiden digitaalista sekä taloudellista kehittymistä. [19]

Tietosuoja-asetus koskettaa kaikkia henkilötietoja. Henkilötiedoiksi lasketaan kaikki tiedot, jotka liittyvät tunnistettuun sekä suoraan tai välillisesti tunnistettavissa olevaan henkilöön. Perinteisten nimi- ja osoitetietojen lisäksi jopa pseudonymisoidut tiedot, jotka voidaan lisätietoja käyttämällä yhdistää oikeaan henkilöön, katsotaan henkilötiedoiksi. Määrittäessä, onko henkilö tunnistettavissa tiedoista vai ei, on otettava huomioon kaikki tunnistamiseen todennäköisesti käytettävät keinot. Myös tietoihin, jotka voidaan palauttaa tunnistettavaan muotoon, tulee soveltaa asetusta. Kuitenkaan täysin anonymisoituihin tietoihin, joista henkilöä ei ole mahdollista tunnistaa, ei tarvitse soveltaa tietosuoja-asetusta. [19]

Kerätäkseen tai käsitelläkseen henkilötietoja organisaation on saatava siihen suostumus rekisteröidyltä, eli henkilöltä kenen henkilötietoja käsitellään. Suostumus on selkeästi annettu hyväksyntä käsittelylle. Suostumuksen täytyy olla vapaasti annettu, tietoinen, yksiselitteinen ja kosketettava juuri kyseistä tilannetta. Antaakseen suostumuksen rekisteröidyn täytyy tietää mitä varten hänen tietojaan kerätään ja kuinka niitä käsitellään. [19]

Suostumuksen saamisen jälkeen organisaatio voi aloittaa henkilötietojen käsittelyn. Käsittelyn on oltava lain- ja asianmukaista sekä läpinäkyvää. Henkilötietojen on oltava riittäviä, oleellisia ja paikkansapitäviä. Käsittelyn on rajoituttava vain välttämättömään käsittelyyn ja tietoihin. Organisaatio ei saa kerätä tietoa, jota se ei tarvitse tavoitteidensa täyttämiseen. Käsittelylle on asetettava määräaika ja tietoja saa säilyttää vain niin kauan kuin käsittelyn tarkoitus niin edellyttää. Henkilötietojen käsittelyssä on aina otettava huomioon tietojen suojaaminen. Tietosuoja-asetuksen artiklan 25 mukaan käsittelijän on aina toteuttava sopivia tietosuojaperiaatteita ja huomioitava turvallisuus jo tuotetta tai toimintaa suunniteltaessa. Organisaation vastuulla on toteuttaa sopivat riskiä vastaavat tekniset ja organisatoriset ratkaisut suojan takaamiseksi. Etenkin tietojen minimoinnin periaate, eli käsiteltävän tiedon määrän rajoittaminen, on taattava. [19]

Asetus määrittelee tietyt henkilötietoryhmät erityisiksi esimerkiksi niiden arkaluontoisuuden takia. Terveystiedot lasketaan erityisiksi henkilötiedoiksi ja siten asetuksen huomioiminen kaikessa niiden käsittelyssä on tärkeää. Terveystiedoiksi lasketaan kaikki nykyisestä, entisestä ja tulevasta terveydestä kertovat tiedot. Terveystiedot pitää sisällään fyysisestä terveydestä ja mielenterveydestä kertovat tiedot. Valokuvien käsittely ei itsessään sisälly erityistilanteisiin. Mutta jos valokuviin sovelletaan erityisen teknisiä

menetelmiä, jotka mahdollistavat tunnistamisen, täytyy niitä käsitellä kuin muitakin erityisiä henkilötietoryhmiä. [19]

Tietosuoja-asetuksen artiklassa 9 kielletään mainittujen erityisten henkilötietojen käsittely tiettyjä poikkeuksia lukuun ottamatta. Eräs näistä asetuksen poikkeuksista on rekisteröidyltä saatu nimenomainen lupa. Ilman tätä on erityisten henkilötietojen käsittely vaikeaa. [19] Kuitenkin Suomen tietosuojalaissa artiklaa ja erityisesti poikkeuksia täydennetään. Esimerkiksi tarpeellisten terveystietojen käsittely terveydenhuollon palveluntarjoajan toimesta sallitaan, jotta terveyspalvelujentuottaja voi järjestää ja toteuttaa hoidon. [20]

Vaikka tietosuojalaki antaa enemmän mahdollisuuksia erityisten henkilötietojen käsittelyyn määrittää se kuitenkin lisätoimenpiteitä, jotka täytyy toteuttaa käsiteltäessä erityisiä henkilötietoja. Lisätoimenpiteiden tavoitteena on rekisteröidyn oikeuksien suojaaminen. Lain mukaan käsittelystä on pidettävä lokia, josta voidaan jälkeenpäin tarkistaa, kenen toimesta tietoja on käsitelty. Henkilötietoihin pääsy on estettävä tarpeen mukaan ja henkilötiedottiedot ovat pseudonymisoitava sekä salattava. [20]

Käsittelyn loputtua on organisaation poistettava tiedot välittämöstä. Heidän on myös välitettävä tieto poistamisesta eteenpäin, jotta tiedot voidaan poistaa kaikkialta, johon ne on jaettu. Tietojen poistamiseen liittyy myös rekisteröidyn oikeus ”tulla unohdetuksi”. Oikeutta voidaan soveltaa tilanteessa, jossa henkilötietoja ei enää tarvita tehtävään, johon ne oli kerätty, tietojen käsittely on lainvastaista tai jos rekisteröity vetää suostumuksensa pois. Kun sovelletaan rekisteröidyn oikeutta tulla unohdetuksi, on organisaation kunnioitettava sitä ja poistettava tiedot. Myös tällöin organisaation on välitettävä tieto tietojen poistamisesta eteenpäin kaikille, joille tietoja on jaettu. [19]

Muita rekisteröidyn oikeuksia, jotka tulee huomioida hajautettua oppimista tutkiessa, on oikeus olla joutumatta pelkästään automaattisen profiloinnin kohteeksi, mikäli se vaikuttaa häneen merkittävällä tavalla. Profiloinniksi määritetään kaikki henkilötietojen automaattinen käsittely ominaisuuksien kuten terveyden määrittämiseksi. Tätä ei sovelleta, jos asiasta on sovittu erikseen. Lisäksi rekisteröidyn oikeus tulla unohdetuksi, oikeus pyytää henkilötietojen siirtämistä ja oikeus vaatia tietojen oikaisua, saattavat aiheuttaa yksittäisten opetuksessa käytettävien potilastietojen poistumista tai muuttumista. Organisaation vastuulla on kaikkien rekisteröidyn oikeuksien toteuttaminen ja henkilötietojen suojaaminen, kuten asetuksessa määritetään. Organisaation on myös pyydettäessä kyettävä osoittamaan noudattavansa tietosuoja-asetusta. [19]

3.2 Lääkinnällisten laitteiden regulaatiot

EU:n alueella on voimassa kaksi lääkinnällisten laitteiden direktiiviä: lääkinnällisten laitteiden asetus 2017/745 (MD-asetus) ja *in vitro* -diagnostiikassa käytettävien lääkinnällisten laitteiden asetus 2017/746 (IVD-asetus). Nämä asetukset määrittävät säännöt lääkinnällisten laitteiden ja niiden lisälaitteiden markkinoille tuomisesta ja käyttöönotosta unionissa. Asetukset määrittävät myös valmistajan vastuun käyttöönoton jälkeen. [21], [22]

Lääkinnällinen laite määritellään MD-asetuksessa artiklassa 2. Tästä määritelmästä on tärkeää huomata, että myös ohjelmistot voidaan laskea lääkinnälliseksi laitteeksi. Ohjelmisto lasketaan lääkinnälliseksi laitteeksi, jos se on tarkoitettu tuottamaan tietoa diagnosointiin, sairauden ehkäisyyn tai ennusteen laatimiseen. Myös patologisen tilan tutkimiseen tarkoitetut ohjelmistot lasketaan lääkinnällisiksi laitteiksi. [21] Tutkittaessa hajautetun oppimisen soveltamista terveydenhuoltoon ovat nämä kaikki tehtäviä, joihin menetelmää voidaan soveltaa.

IVD-laitteiden määritelmä pohjautuu lääkinnällisten laitteiden määritelmään. Sekin pitää sisällään ohjelmistot ja niiden osat. Jos ohjelmiston on tarkoitus tuottaa tietoa patologisesta tai fysiologisesta tilasta se lasketaan IVD-laitteeksi. Tällöin esimerkiksi digipatologiaan kuvien tulkintaan suunnitellut ohjelmistot kuuluvat asetuksen alaisuuteen. [22]

Lääkinnälliset laitteet luokitellaan riskiluokkiin I, II a, II b ja III. Näistä luokan III laitteiden käytöllä on kaikkein korkein riski ja luokan I laitteilla matalin. Luokkaa nostaa laitteen käyttötarkoitus ja käyttöaika. Suurin osa terveydenhuollon ohjelmistoista luokitellaan luokkaan I. Kuitenkin ohjelmistot, joita hyödynnetään diagnostiikassa tai terapeuttisten päätösten tekemisessä kuuluvan vähintään luokkaan II a. Ohjelmiston luokka voi kuitenkin olla myös II b tai III, jos sen tuottaman tiedon perusteella tehdyt päätökset voivat johtaa potilaan terveyden vakaavaa tai pysyvään heikkenemiseen, kirurgiseen toimenpiteeseen tai kuolemaan. Terveydenhuollossa hajautettuun oppimiseen perustuvien ohjelmistojen tavoitteena on usein tuottaa tietoa diagnostiikan tai hoidon suunnittelemisen tueksi. Tällöin ohjelmistot luokitellaan vähintään luokkaan II a. [21]

In vitro -diagnostiikan laitteet luokitellaan luokkiin A, B, C ja D. Luokittelu tapahtuu laitteen käyttötarkoituksen ja riskin perusteella. Se mihin luokkaan hajautetun oppimisen ohjelmistot kuuluvat riippuu siitä, mihin niitä sovelletaan. Esimerkiksi hajautetun oppimisen soveltaminen COVID-19 taudin diagnosointiin [10], pistäisi ohjelmiston luokkaan D. Sillä luokan D laitteet tunnistavat tartuntatauteja ja niiden aiheuttajia. Luokan D laitteilla on korkein riskiluokitus. Yleisin luokka on kuitenkin B, sillä siihen menevät kaikki

laitteet, joiden käyttötarkoitusta ei erikseen mainita asetuksessa. Mitä yleisemmässä käytössä laite on, sitä matalampi on sen luokitus. [22]

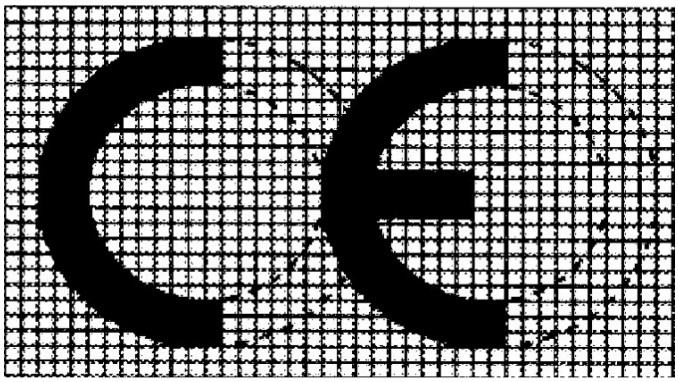
Lääkinnällisten laitteiden saaminen markkinoille edellyttää kliinisen tutkimuksen toteuttamista. Kliinisessä tutkimuksessa tulee osoittaa, että laite saavuttaa valmistajan ilmoittaman suorituskyvyn, sekä muut laitteen kliiniset hyödyt. Kliinisen tutkimuksen on perustuttava jatkuvaan suorituskyvyn arviointiin ja tuettava laitteen käyttötarkoitusta. Suorituskyvyn arvioinnin on osoitettava tieteellinen validiteetti ja analyyttinen sekä kliininen suorituskyky. Kliininen tutkimus on mahdollisuuksien mukaan toteutettava oikeanlaisessa käyttöympäristössä. Jos laite kuuluu korkeimpiin luokkiin, on sen suorituskyvyn arviointi uusittava vähintään kerran vuodessa. [21], [22]

Asetusten mukaan valmistaja on myös velvollinen tekemään laitteella riskienhallintajärjestelmän. Se on jatkuva prosessi, joka kestää koko laitteen elinkaaren ajan. Asetus edellyttää valmistajaa huomioimaan ja vähentämään riskejä jo suunnitteluvaiheessa. Valmistaja on myös velvollinen poistamaan tai vähentämään tunnettuja riskejä, koko laitteen elinkaaren ajan. Kun kyse on ohjelmistosta, joka täyttää lääkinnällisen laitteen kriteerit, on se riskien vähentämiseksi suunniteltava ja toteutettava uusimman kehityksen mukaan. [21], [22]

Käyttöön otettavista lääkinnällisillä laitteilla on oltava kaikki tekniset dokumentit kunnossa. Edellä käsitelty riskienhallintajärjestelmä on eräs näistä vaadituista teknisistä dokumenteista. Tämän lisäksi laitteelle on tehtävä laadunvalvontajärjestelmä, jolla varmistetaan asetuksen noudattaminen tehokkaasti ja laitteen riskiluokkaan sopivalla tavalla. Laadunvalvontajärjestelmän on pidettävä sisällään tuotteen suunnittelun, kehityksen ja ylläpidon toimenpiteiden tarkkailu. Laadunvalvontajärjestelmän lisäksi valmistajan on tehtävä tiivistelmä laitteen turvallisuudesta ja kliinisestä suorituskyvystä. Tiivistelmän on oltava sellaisessa muodossa, jota käyttäjä ja potilaskin ymmärtävät. [21], [22]

3.3 CE-merkintä

CE-merkintä kertoo tuotteen täyttävän sen vähimmäisvaatimukset. Merkintä myönnetään tuotteelle, kun se täyttää olennaiset vaatimukset, on direktiivien mukainen ja sen arviointi on tehty vaatimusten mukaisesti. CE-merkinnän käyttöä valvotaan viranomaisten toimesta ja merkinnän saa kiinnittää vain tuotteisiin, joista säädetään Euroopan lainsäädännössä. [23] CE-merkintä näkyy kuvassa 3.



Kuva 3: *CE-merkintä, lähteestä* ***[21]***

Lääkinnällisten laitteiden asetukset edellyttävät, että laitteissa on CE-merkintä. Merkintä on kiinnitettävä laitteeseen tai sen pakkaukseen ennen markkinoille saattamista. Jos merkinnän kiinnittämien ei ole mahdollista on se silti löydyttävä laitteen käyttöohjeista. Ohjelmistoissa CE-merkinnän on oltava näkyvillä käynnistyssivulla tai ohjelman yleisten tietojen ja versiotietojen yhteydessä. [21]–[23]

4. KIRJALLISUUSKATSAUKSEN MENETELMÄT

Tiedonhaku kirjallisuuskatsausta varten aloitettiin 13. helmikuuta 2023. Tiedonhaun alussa valittiin käytettäväksi tietokannaksi Tampereen yliopiston hakupalvelu Andor. Tiedonhaku suoritettiin lähes kokonaan englanninkielisillä termeillä, sillä jo ensimmäinen haku sanoilla ”hajautettu oppiminen” tuotti nolla tulosta.

Ennen tiedonhaun aloittamista tehtiin muutamia rajauksia. Ensimmäinen ja tärkein rajaus oli julkaisuvuosi. Työssä halutaan keskittyä uusimpaan ja ajankohtaisimpaan tietoon, joten kaikki ennen vuotta 2020 tehdyt julkaisut rajattiin pois. Toinen tärkeä ehto oli, että julkaisut ovat saatavilla verkossa. Myös aiheeseen liittyvät patentit rajattiin pois.

Alustavien rajausten ja tietokannan valinnan jälkeen siirryttiin tiedonhankintavaiheeseen. Pääasiallisena hakusanana käytettiin englanninkielistä termiä ”federated learning”. Aihe kohdennettiin oikeaan soveltamisalaan yhdistämällä hakuun Boolean operaattorin ”AND” avulla termi ”healthcare” tai ”medical imaging”. Haku toteutettiin myös käyttämällä hakusanan ”federated learning” sijasta termiä ”distributed learning”. Tämä ei kuitenkaan tuottanut haluttuja tuloksia. Sillä työssä käsitellään menetelmään liittyviä tietosuojamekanismeja, tehtiin haku myös hyödyntämällä termejä ”privacy” ja ”privacy-preserving”. Näin saatiin hakutulosten joukkoon myös tietosuojaa käsitteleviä julkaisuja.

Hakutuloksista rajaus tehtiin otsikon perusteella ja aineiston kooksi saatiin 62 julkaisua. Aineistoa rajattiin ennestään abstraktien perusteella. Jotkin julkaisut luettiin kokonaan ennen kuin niiden käytöstä päätettiin. Aineistoa rajatessa pohdittiin, käsitteleekö julkaisu hajautettua oppimista ja sen yhdistämistä lääketieteelliseen kuvantamiseen tai terveydenhuoltoon. Jos julkaisu ei käsitellyt mainittuja aiheita pohdittiin, käsitteleekö se hajautetun oppimisen toteuttamista tietoturvallisesti. Jos julkaisu ei käsitellyt kumpaakaan aihetta rajattiin se pois aineistosta. Aineistoa rajatessa pohdittiin myös tuoko julkaisu lisäarvoa tuomalla uutta tietoa aineistoon tai selkeyttääkö se jotakin aihetta. Tavoitteena oli vähentää aineiston sisältämää toistoa ja hyvin samankaltaisten artikkelien määrää. Näin aineisto saatiin rajattua hieman alle kahteenkymmeneen julkaisuun.

Seuraavaksi tiedon haussa siirryttiin täydentämään mahdollisesti aineistoon jääneitä aukkoja. Tämä toteutettiin valitsemalla valitun aineiston lähteistä artikkeleita, jotka toistuivat paljon tai jotka käsittelivät aihetta hieman uudella tavalla. Näiden artikkelien koh-

dalla saatettiin tehdä poikkeus alussa määritettyyn vuosirajaukseen 2020. Jo valitun aineiston lähteistä päädyttiin lisäämään aineistoon kolme julkaisua.

Tiedonhankinnan viimeinen vaihe oli regulaatioita koskeva tiedonhankinta. Regulaatiot ja lait eivät ole saatavilla Andorin kautta, joten niiden haku suoritettiin Googlen ja Finlexin tietokannan avulla. Hakusanoina käytettiin termejä ”GDPR”, ”tietosuojalaki” ja ”lääkinnällisten laitteiden regulaatiot”. Tulokset suodatettiin otsikon perusteella, sillä regulaatioiden yksiselitteiset otsikot kertoivat selkeästi, onko kyseessä oikea regulaatio.

5. YKSITYISYYDEN SUOJAAMINEN HAJAUTETUSSA OPPIMISESSA

Henkilötiedot ja rekisteröidyn yksityisyys suojataan lain toimesta sekä EU:n tietosuoja-asetuksessa että Suomen tietosuojalaissa. Tietosuoja-asetuksessa on mainittu, että datan turvallisuus ja henkilötietojen suojaaminen on huomioitava jo ohjelman suunnittelun aikana. [19] On siis tärkeää, että suunniteltaessa hajautetun oppimisen menetelmiä huomioidaan heti alussa potilaan yksityisyyden turvaaminen.

Vaikka hajautetun oppimisen perusideana on, että dataa ei jaeta muille osallistujille ei se yksistään riitä turvaamaan dataa [3]. Hajautetussa oppimisessa tietosuojariskit muodostuvat yleensä paikallisten mallien parametrien paljastumisesta. Parametrit voivat paljastaa hyvinkin arkaluontoisia asioita käytetystä opetusdatasta. [4]

Hajautetussa oppimisessa tietosuojauhat voivat tulla koko järjestelmän ulko- tai sisäpuolelta. Vaikka tutkitaan hajautetun oppimisen soveltamista terveydenhuoltoon, jossa osallistujat ovat yleensä hyvin luotettavia, mahdollisuus epäluotettavasta tai hyökkäyksen kohteena olevasta osallistujasta tai keskuspalvelimesta on silti olemassa. Ulkoisia turvallisuus uhkia voivat olla mallin loppukäyttäjät, jotka saavat käsiinsä valmiin mallin tai tekijät, jotka salakuuntelevat palvelimien välistä kommunikaatiota. Salakuuntelun avulla ne voivat saada käsiinsä paikallisen mallien tai valmiin yhdistetyn mallin. [4]

Soveltaessa tietosuojamenetelmiä hajautettuun oppimiseen isoimpana haasteena on datan käytettävyyden säilyttäminen. Hajautettua oppimista tutkiessa menetelmien käytettävyyttä tutkitaan yleensä vertailemalla tietosuojatun mallin tarkkuutta vertailumallin tarkkuuteen. Vertailumallina käytetään yleensä mallia, johon ei ole lisätty tietosuojaa parantavia mekanismeja. [4]

Tietosuojamenetelmät voidaan jakaa anonymisointimenetelmiin, häiriömenetelmiin ja kryptografisiin menetelmiin [4]. Tässä kappaleessa käsitellään jokainen kategoria ja käydään läpi niiden sisältämiä menetelmiä. Näiden lisäksi käsitellään myös menetelmiä, jotka eivät suoraan kuulu minkään kategorian alaisuuteen. Menetelmistä käydään läpi niiden toimintaperiaatteet, vahvuudet ja heikkoudet, sekä käsitellään mihin hajautetun oppimisen vaiheeseen menetelmiä voidaan soveltaa. Lopussa pohditaan menetelmien yhdistämistä tietosuojan takaamiseksi ja mallin tehokkuuden säilyttämiseksi.

5.1 Anonymisointimenetelmät

Anonymisointimenetelmät ovat tunnetuimpia ja eniten käytettyjä tietosuojan keinoja. Menetelmissä tietoja muokataan niin, että henkilön tunnistaminen ilman lisätietoja olisi mahdotonta. [11] Tavoitteena on silti säilyttää tietoaineiston arvo [10] ja käytettävyys [4]. Nimestään huolimatta myös pseudonymisointi kuuluu anonymisointimenetelmiin [11]. Pseudonymisointi onkin vaatimuksena sekä EU:n tietosuoja-asetuksessa että Suomen tietosuojalaissa erityisten henkilötietojen käsittelylle [19], [20]. Tästäkin syystä on tärkeää tietää mitä menetelmät käytännössä ovat.

Anonymisointi tarkoittaa tunnistetietojen poistamista datasta. Tunnistetiedoiksi lasketaan kaikki tiedot, joiden perusteella data voidaan yhdistää oikeaan henkilöön, esimerkiksi nimi, ikä ja sukupuoli. Teksti- tai numeromuotoisen datan anonymisointi on suhteellisen helppoa, mutta lääketieteellinen data sisältää muitakin muotoja, kuten kuvia. Tällöin kuvista on poistettava metadatan lisäksi myös muut tunnistetiedot, kuten kasvojen piirteet, joiden avulla potilaan henkilöllisyys voitaisiin saada selville. [11] Täysin anonyymin datan luomiseksi on siitä tunnistetietojen lisäksi poistettava kaikki yhteydet ja polut alkuperäiseen tunnistettavissa olevaan dataan [8].

Pseudonymisointi on anonymisoinnista hiukan eroava menetelmä. Siinä tietojen poistamisen lisäksi luodaan poistettujen tietojen tilalle uusia keinotekoisia tietoja. Pseudonymisoinnissa luodaan tunnisteavain, jonka avulla data voidaan yhdistää oikeaan henkilöön. Menetelmä siis mahdollistaa datan anonyymin käytön sekä tietojen yhdistämisen oikeaan henkilöön. [11] Tunnisteavainten taulukon turvallinen käsittely on tärkeää. Se täytyy säilyttää erillään datasta eikä sitä jaeta pseudonymisoidun datan mukana uudelleen. [8] Pseudonymisointi on hieman anonymisointia haasteellisempaa, sillä se vaatii tietojen muokkausta pelkän poistamisen lisäksi. Haasteita menetelmään tuo myös tunnistustaulukon turvassa pitäminen. [11]

Molemmat menetelmät ovat hyvin yksinkertaisia ja yleensä ne ovatkin kaikkein helppokäyttöisimpiä keinoja tietosuojan parantamiseen. Haasteina menetelmissä on niiden peruminen. [11] Esimerkiksi jos potilas vaatii tietojensa oikaisua, on organisaation vaikeaa paikantaa oikea potilasdata ja korjata se. Kuitenkaan tästä huolimatta pelkät anonymisointi menetelmät eivät tarjoa riittävää henkilötietojen suojaa, sillä esimerkiksi yhdistelemällä useita anonyymejä tietoaineistoja on henkilöllisyyden selvittäminen silti mahdollista [10].

Tutkiessa hajautetun oppimisen soveltamista terveydenhuoltoon kannattaa huomioida, että EU:n tietosuoja-asetusta tai Suomen tietosuojalakia ei tarvitse soveltaa, jos käytettävä data on täysin anonyymia. [19], [20] Kuitenkin täysin anonyymin datan luominen

on haasteellista. Mikäli datasta ei saada täysin anonyymia on se varsinkin terveystietoja käsitellessä pseudonymisoitava [19]. Hajautetussa oppimisessa anonymisointi tai pseudonymisointi toimivatkin yleensä tietosuojan ensimmäisenä askeleena. Niiden tavoitteena on opetusdatan yksityisyydensuojan parantaminen ja niiden päälle voidaan rakentaa kestävä tietosuoja kokonaisuus. [11]

5.2 Häiriömenetelmät

Häiriömenetelmien ydinidea on häiriön, esimerkiksi kohinan, lisääminen alkuperäiseen dataan. Tavoitteena on, että sen lisääminen dataan ei muuta siitä saatavaa tilastollista tietoa. [4] Häiriömenetelmissä oletuksena on, että aineiston tai algoritmin systemaattinen satunnaismuutos vähentää yksittäisen tietolähteen merkitystä mutta säilyttää kokonaisuuden tilastollisesti muuttumattomana [11].

Häiriömenetelmien avulla on mahdollista saavuttaa differentiaalinen yksityisyys. Se on matemaattinen malli, joka antaa numeerista tietoa yksityisyyden säilymisestä. Differentiaalisen yksityisyyden ehdot täyttyvät, jos yhden datalähteen muuttuminen ei muuta syntyvää lopputulosta merkittävästi. [24] Hajautetussa oppimisessa tämä tarkoittaa, että yksittäisen potilaan datan lisääminen tai poistaminen ei muuta mallia huomattavasti [14]. Tämä takaa sen, että yksittäisen henkilön tiedot pysyvät turvassa ja ehkäisee uudelleen identifiointi hyökkäyksiä [9].

Differentiaalista yksityisyyttä tavoiteltaessa lisättävän kohinan määrää säädetään parametrin epsilon (ε) avulla. Mitä suurempi ε sitä vähemmän häiriötä lisätään ja vastaavasti mitä pienempi ε sitä enemmän häiriötä lisätään. Häiriön määrä on suoraan verrannollinen luodun yksityisyydensuojan määrään. Mikä tarkoittaa, että mitä enemmän häiriötä sen parempi suoja. [9] Häiriön määrä vaikuttaa myös mallin tarkkuuteen. Mitä korkeampi yksityisyydensuoja häiriöllä luodaan, sitä matalampi on mallin tarkkuus [14]. Li ym. sovelsi menetelmää tutkiessaan Alzheimerin taudin havaitsemista. Tutkimuksessa osoitettiin, että ε:n laskeminen 2:sta 0.8:aan laski mallin tarkkuutta hieman yli prosentin. Differentiaalinen yksityisyys voidaan siis tuottaa ilman merkittävää suorituskyvyn laskua. [16]

Differentiaalinen yksityisyys ja häiriön lisääminen voidaan toteuttaa monella tapaa. Esimerkiksi tietoaineiston satunnaisella sekoittamisella, kohinan lisäämisellä dataan tai tiettyjen toimintojen lisäämisellä neuroverkkoon. [11] Valitulla menetelmällä on kuitenkin vaikutusta hajautetun oppimisen mallin suorituskykyyn. Esimerkiksi verratessa Laplace ja Gaussian kohinan lisäämistä dataan Li ym. huomasi Laplace kohinan tuottavan tarkemman mallin. [16]

Haasteita häiriömenetelmissä aiheuttaa itse häiriön lisääminen. Datan muokkaaminen heikentää datan laatua. Se saattaa aiheuttaa ongelmia erityisesti sovelluksissa, joissa opetusdataa on vain rajallisesti saatavilla. [11] Toinen haaste on häiriön poissuodattamisen mahdollisuus. Tarpeeksi isosta määrästä aineistoa on mahdollista suodattaa häiriö pois, jolloin menetelmät eivät takaa yksityisyydensuojaa. [8]

Hajautetussa oppimisessa häiriömenetelmiä ja differentiaalista yksityisyyttä sovelletaan yleensä, ennen tai jälkeen paikallisten mallien yhdistämistä. Muiden osallistujien ja keskuspalvelimen luotettavuuden perusteella valitaan tavoiteltava yksityisyydensuojan taso. [4] Luodusta yksityisyydensuojan tasosta riippuen puhutaan joko globaalista tai lokaalista differentiaalisesta yksityisyydestä. Globaalissa differentiaalisessa yksityisyydessä suojataan paikallisia malleja muilta osallistujilta, mutta ei keskuspalvelimelta. Jos suojaudutaan myös keskuspalvelimelta, puhutaan lokaalista differentiaalisesta yksityisyydestä. [5] Lokaalin yksityisyyden mallissa häiriötä lisätään paikallisiin malleihin ennen niiden yhdistämistä. Globaalissa yksityisyydessä häiriö lisätään vasta yhdistettyyn malliin ennen sen välittämistä osallistujille. [9] Lokaali differentiaalinen yksityisyys tarjoaa paremman yksityisyyden suojan, mutta siinä lisättävän häiriön määrä on suurempi [4].

5.3 Kryptografiset menetelmät

Kryptografiset menetelmät ovat erilaisia salausmekanismeja, joita käytetään laajasti koneoppimisen yhteydessä [4]. Menetelmät ovat idealtaan yksinkertaisia, mutta teknisesti vaikeita. Ne ovat luotettavia ja niitä voidaan soveltaa sekä algoritmeihin, että dataan itseensä. [11] Yleisiä kryptografisia menetelmiä ovat homomorfinen salaus, salainen jakaminen ja turvallinen yhteislaskenta (SMC, engl. secure multi-party computation). [4]

5.3.1 Homomorfinen salaus

Homomorfinen salaus on kryptografinen menetelmä, joka mahdollistaa salatun tiedon käyttämisen tietyissä toimenpiteissä ilman salauksen purkamista. Menetelmän ansioista salattua tietoa voidaan käyttää kuin sitä ei olisi salattukaan. [11] Menetelmä perustuu salatun tiedon homomorfisiin ominaisuuksiin. Niiden perusteella luodaan salausavaimet salauksen tekemiseen ja purkamiseen. [1] Hajautetussa oppimisessa salausavaimet tuottaa yleensä keskuspalvelin tai jokin osallistujista [4].

Salausalgoritmia kutsutaan homomorfiseksi salaukseksi, jos toimenpiteeseen ◦ pätee $EN(m1)°EN(m2) = EN(m1 \circ m2)\ \forall m1, m2\ \epsilon M$. Määritelmässä *M* tarkoittaa salaamatonta alkuperäistä tietoa ja $EN()$ merkintä kuvaa tiedon salaamista. Määritelmä osoittaa,

että toimenpiteestä saadaan sama tulos, jos se suoritetaan salaamattomalla datalla ja vain lopputulos salataan kuin jos lähtöarvot salataan ja toimenpide suoritetaan salatulla tiedolla. [4]

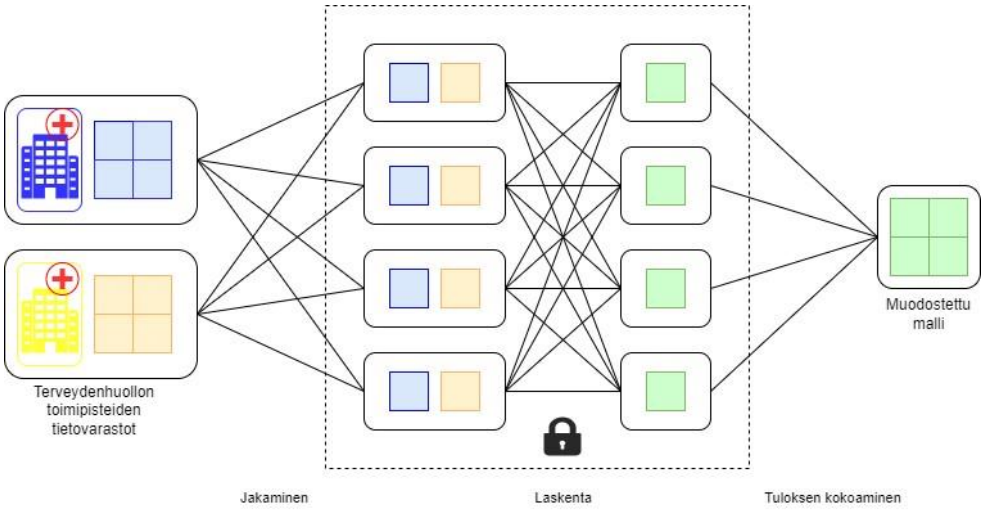
Homomorfista salausta voidaan soveltaa sekä algoritmiin että dataan itseensä [11]. Ja menetelmää onkin onnistuneesti sovellettu konvoluutioneuroverkkoihin, jotka toimivat yleensä hajautetun oppimisen opetettavana tekoälyn mallina. Menetelmää voidaan myös soveltaa kuvien salaamiseen. Kuvat käsitellään matriiseina ja salataan yksityisellä salausavaimella. Salatuilla matriiseilla suoritetaan halutut toimenpiteet. Lopuksi salaus on mahdollista purkaa kuvista salausavaimen avulla. [9]

Homomorfisen salauksen etuna on sen yksikertainen idea. Tämä helpottaa menetelmän selittämistä potilaille, jolloin he kykenevät tekemään tietoisen suostumuksen tietojensa käyttämisestä. [11] Menetelmän heikkoutena on sen vaatima laskentateho ja kasvavat kommunikaatio kuluvat resurssit [4].

Hajautetussa oppimisessakin homomorfisesta salauksesta hyödynnetään paikallisten mallien yhdistämisessä. Salaamalla paikalliset mallit ennen niiden lähetystä ja yhdistämistä keskuspalvelimella parannetaan mallien tietoturvaa. Tällöin mallit kulkevat palvelimien välillä aina salatussa muodossa ja niiden käsittely keskuspalvelimella tapahtuu salauksen alaisena. [11], [15] Kun koko yhdistäminen tapahtuu salauksen alaisena, voidaan yleinenkin malli lähettää salattuna takaisin paikallisille palvelimille, jotka purkavat salauksen [4]. Näin sekä paikallisia malleja että yleistä mallia saadaan suojattua niiden siirtämisen ja käsittelyn aikana. Menetelmä tarjoaa suojaa kommunikaatiokanavaan tai keskuspalvelimeen mahdollisesti kohdistuvia hyökkäyksiä ja salakuuntelua vastaan. Menetelmää voidaan myös hyödyntää tilanteessa, jossa keskuspalvelinta ei pidetä täysin luotettavana. [1]

5.3.2 Turvallinen yhteislaskenta

Turvallinen yhteislaskenta (engl. secure multi-party computation), SMC, on eri tekniikoilla monen osallistujan välille muodostettava tietoturvamalli. Se mahdollistaa yhteisen laskennan ilman, että yksittäinen osallistuja saa tietoa käytettävästä datasta. [11] Mallissa käytettävä data salataan, pilkotaan osiin ja jaetaan osallistujien kesken niin, että yksikään osallistuja ei voi paljastaa kaikkea salattua tietoa [9]. Tämän jälkeen laskenta suoritetaan salattualla datalla, kunnes tulos on saatu valmiiksi. Vasta valmiista tuloksesta puretaan salaus. [7] Menetelmää sanotaan SMC-menetelmäksi, jos tuotettava tulos on oikein ja osallistujat eivät tiedä toistensa dataa [4]. Menetelmää havainnollistetaan kuvassa 4.



Kuva 4: *Turvallinen yhteislaskenta, kuva tehty mukaillen lähdettä* ***[7]***

SMC:n avulla voidaan suojata dataa keskuspalvelimelta ja siihen mahdollisesti kohdistuvalta hyökkäykseltä. Menetelmän toteuttaminen on usein teknisesti vaikeaa, vaatii ylimääräistä laskentatehoa ja yleensä hidastaa ohjelman toimintaa [10]. SMC vaatii myös datan jakamista osallistujien välillä, jolloin siitä aiheutuvat kommunikaatio kustannukset ovat korkeat [7], [11]. Kuitenkin käytettynä yhdessä muiden menetelmien kanssa parantaa SMC tietosuojaa ja mahdollistaa esimerkiksi pienemmän häiriön lisäämisen [24].

SMC sopii luonnostaan yhteen hajautetun oppimisen kanssa [7], jossa on tavoitteena yhteisen mallin muodostaminen ilman osallistujien datan jakamista. Hajautetun oppimisen yhteydessä menetelmää on sovellettu paikallisten mallien jakamisessa ja yhdistämisessä keskuspalvelimella [4]. Tällöin keskuspalvelin ei saa yksittäisiä kokonaisia paikallisia malleja tietoonsa, vaan niiden salatut ja erotellut parametrit. Näistä salatuista parametreista muodostetaan yleinen malli.

5.3.3 Salainen jakaminen

Salainen jakaminen on eräs käytössä olevista salausmenetelmistä. Menetelmä perustuu salattavan asian pilkkomiseen osiin osallistujien kesken. Menetelmässä kolmas luotettu osapuoli, jakaja, jakaa salaisuuden *n* osaan osallistujille. Tämän lisäksi jakaja määrittää montako osaa pitää olla koottuna, jotta avain saadaan muodostettua uudelleen. Tätä minimi osien määrää merkitään kirjaimella *t*. Jotta salattu asia saadaan rakennettua uudelleen, tarvitsee vähintään *t* määrän osallistujia lähettää osansa salaisuudesta. [4]

Salainen jakaminen mahdollistaa paikallisten mallien välittämisen keskuspalvelimelle salatussa muodossa. Näin malleja saadaan suojattua kommunikaatiokanavaan kohdis-

tuvalta tietosuojahyökkäyksiltä. Menetelmän heikkoutena on sen tarve luotettavalle jakajalle. Ja se on altis pahantahtoisen osallistujan aiheuttamalle tietosuojariskille. [4]

Li ym. soveltaa salaisen jakamisen menetelmää hajautettuun oppimiseen paikallisten mallien yhdistämisessä. He jakavat keskuspalvelimen salausavaimen, joka tarvitaan mallien salauksen purkamiseen osallistujien kesken. Tällöin paikallisten mallien yhdistäminen voidaan tehdä vain, kun tarpeeksi moni paikallinen palvelin on lähettänyt parametrinsa ja osan salausavaintaan. [16]

5.4 Muut menetelmät

Hajautetun oppimisen menetelmien kokonaisrakennetta miettiessä on hyvä huomioida epäluotettavien osallistujien mahdollisuus. Helppo tapa vähentää epäluotettavan osallistujan aiheuttamaa tietosuojariskiä on valita vain osa osallistujista jokaiselle kierrokselle. Kun osallistuja ei saa jokaisen kierroksen jälkeen päivitettyä yleistä mallia, on sen vaikeampi käyttää mallia muiden osallistujien opetusdatan selvittämiseen. [4]

Epäluotettava keskuspalvelin aiheuttaa korkeamman tietosuoja riskin kuin epäluotettava osallistuja, sillä sen vaikutus hajautetun oppimisen menetelmän toimintaan on suurempi. Tämän riskin alentamiseksi on alennettava myös palvelimen mahdollisuutta vaikuttaa prosessiin ja mahdollisuutta selvittää osallistujien opetusdatan sisältöä. Esimerkiksi muodostettavan mallin parametrien määrän vähentäminen vähentää osallistujilta keskuspalvelimelle siirtyvän tiedon määrää ja parantaa mallin tietoturvaa. [4]

Gong ym. loivat hajautetun siirrosoppimisen avulla menetelmän, joka hyödyntää julkisesti saatavilla olevaa dataa tietosuojan takaamiseen. Ensin opetettava koneoppimisen malli opetetaan paikallisilla palvelimilla niiden omalla yksityisellä datalla aivan kuten normaalistikin. Toisessa vaiheessa jokainen paikallinen malli pysäytetään. Kaikki paikallisten mallien käyttämät yksityiset opetusdatat jätetään täysin huomiotta. Keskuspalvelimen ja paikallisten mallien välillä sovelletaan siirrosoppimista jokaiselle palvelimelle jaetun julkisen datan avulla. Tällöin tieto siirretään hyödyntämällä pelkkää julkista dataa, eikä edes paikallisen mallin parametrejä tarvitse välittää keskuspalvelimelle. Näin menetelmässä saadaan luotua parempi yksityisyydensuoja. [12]

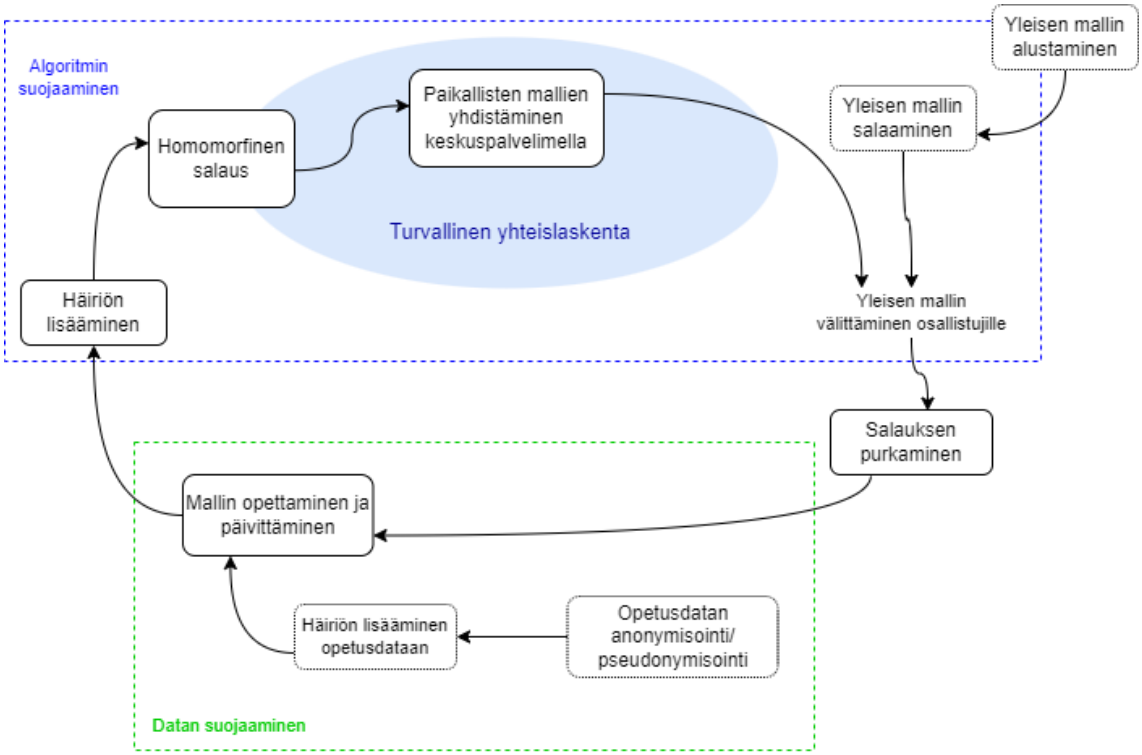
Edellä mainitussa Gong ym. luomassa mallissa on toinenkin erikoinen piirre, joka parantaa sen tietosuojaa huomattavasti. Paikalliset palvelimet eivät ota keskuspalvelimelta vastaan keskeneräistä yleistä mallia. Kommunikaatio palvelimien välillä on yksisuuntaista, kunnes yleinen malli on valmiina. Opetusta suoritetaan edelleen kierroksissa hyödyntämällä eri osia palvelimien omasta ja julkisesta datasta, mutta osallistuvat pal-

velimet eivät saa tietoa muiden vaikutuksesta yleiseen malliin. Paikalliset palvelimet siis keskittyvät vain oman mallinsa opettamiseen. [12]

5.5 Menetelmien yhdistäminen

Jokaisella yksityisyydensuojaamisen menetelmällä on omat heikkoutensa. Anonymisaatiomenetelmät eivät takaa täyttä yksityisyydensuojaa. Salaukseen pohjautuvat menetelmät tyypillisesti lisäävät laskentatehon ja kommunikaation tarvetta. Ja häiriön lisäämiseen perustuvat menetelmät yleensä heikentävät datan laatua ja sen käytettävyyttä. [4] Kun jokaisella menetelmällä on omat huonot puolensa, ei tietosuojaa kannata jättää vain yhden menetelmän varaan. Heikkouksien lisäksi on jokaisella menetelmällä myös omat vahvuutensa ja tarkoituksensa. Menetelmien yhdistäminen mahdollistaakin tietosuojan takaamisen ja datan käytettävyyden sekä mallin tarkkuuden säilyttämisen [4].

Kaissis ym. esittelee artikkelissaan eri tietosuojamenetelmien yhdistämistä ja suhdetta toisiinsa. Siinä menetelmät jaetaan karkeasti datan suojaamisen ja algoritmin suojaamiseen tarkoitettuihin menetelmiin. [11] Kuva 5 havainnollistaa tietosuojamenetelmien luokittelua ja visualisoi mihin eri vaiheisiin mitäkin menetelmää voidaan käyttää.



Kuva 5: *Tietosuojaa parantavien menetelmien yhdistäminen*

Kuvasta on havaittavissa, kuinka montaa eri menetelmää on mahdollista yhdistää yhteen hajautetun oppimisen sovellukseen. Prosessi voidaan aloittaa suojaamalla poti-

lasdata. Tämä saavutetaan hyödyntämällä datan anonymisointia tai pseudonymisointia. Myös häiriön lisäämistä yksittäisiin potilastietoihin voidaan hyödyntää, jolloin luodaan laaja differentiaalinen yksityisyys koko opetusdatalla. Potilastietojen luonteen mukaan voidaan näistä menetelmistä valita yksi tai useampi. [11]

Datan suojaamisen lisäksi täytyy miettiä algoritmin suojaamista. Tämä tapahtuu luontevasti suojaamalla palvelimien välinen kommunikaatio, sillä siinä tapahtuvat tietovuodot aiheuttavat vakavan tietosuojariskin. Ja vaikka potilasdata olisi suojattu voi paljastuneesta mallista silti tunnistaa potilaita. Kommunikaation ja siten algoritmin suojaaminen tapahtuu erilaisten kryptografisten toimenpiteiden avulla. Salaamalla palvelimien välinen viestintä esimerkiksi homomorfisilla menetelmillä ja soveltamalla SMC:tä paikallisten mallien yhdistämiseen, luodaan merkittävästi parempi tietosuoja paikallisille algoritmeille. [11]

Truex ym. tutkivat differentiaalisen yksityisyyden ja SMC:n yhdistämistä tuottaakseen tarkan ja turvallisen hajautetun oppimisen mallin. Menetelmässä osallistujat lisäävät Gaussian kohinaa omiin paikallisiin malleihinsa luodakseen lokaalin differentiaalisen yksityisyyden. Paikalliset mallit salataan ja lähetetään keskuspalvelimelle. Keskuspalvelin yhdistää salatut mallit niiden homomorfisten ominaisuuksien perusteella. Keskuspalvelin lähettää yhdistetyn salatun mallin osalle paikallisista palvelimista ja ne purkavat salauksen osittain. Kun palvelimella on tarpeeksi osittain purettuja yhdistettyjä malleja, pystyy se tuottamaan ja päivittämään yleisen mallin, joka ei ole salattu. [24]

Truex ym. luomaa mallia verrataan artikkelissa kahteen hajautetun oppimisen malliin. Toinen on toteutettu ilman minkäänlaisia lisätoimia tietosuojan parantamiseksi ja toisessa lokaali differentiaalinen yksityisyys saavutetaan pelkän häiriön lisäämisen avulla. Truex ym. menetelmä saavuttaa lähes yhtä hyvän tuloksen kuin hajautettu oppiminen ilman tietosuojaa parantavia toimenpiteitä. Verrattuna pelkästään häiriöön luottavaan menetelmään, malli saavuttaa huomattavasti paremman tuloksen. Truex ym. luoman mallin parempi suorituskyky selittyy todennäköisesti lisättävän häiriön määrällä. Sillä menetelmään sovelletun salauksen ansioista differentiaalinen yksityisyys saavutettiin pienemmällä häiriön määrällä. [24]

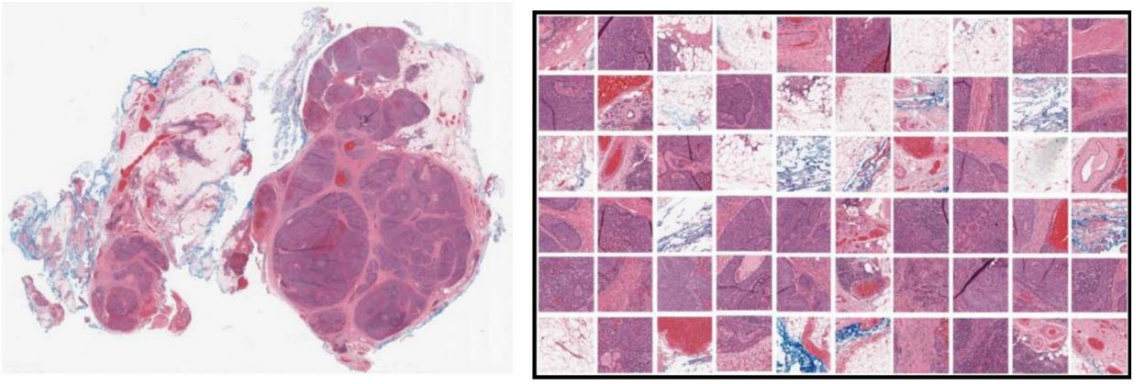
6. HAJAUTETUN OPPIMISEN SOVELTAMINEN DIGIPATOLOGIAAN

Digipatologiassa kuvien kerääminen yhteen paikkaan koneoppimisen mallia varten on poikkeuksellisen haasteellista. Regulatoristen haasteiden lisäksi digipatologiassa tilannetta vaikeuttaa kuvien tarkkuus ja sen vaatima kuvatiedoston koko. Kuvat ovat korkearesoluutioisia ja voivat olla jopa yli gigapikselin kokoisia. [3] Tällöin kuvien siirtäminen ja varastoiminen yhteen paikkaan aiheuttaa huomattavia teknisiä haasteita. Hajautetun oppimisen soveltaminen tilanteeseen helpottaa tilannetta huomattavasti, sillä se vähentää tiedon siirtämistarvetta. [14]

Kudos- ja digipatologiassa kuvien variaatio on laajaa. Variaatiota aiheuttaa organisaatioiden väliset erot värjäys- ja kuvantamismenetelmissä sekä erot käytössä olevissa laitteissa ja ohjelmistoissa. Tällöin yhden organisaation datalla opetettu koneoppimisen malli ei välttämättä suoriudu tarvittavalla tasolla toisessa organisaatiossa. Hajautettu oppiminen pienentää kuvien variaation aiheuttamia haasteita, sillä opettamisessa päästään hyödyntämään useamman organisaation dataa. Näin muodostettavalla mallilla on parempi yleistämiskyky ja se sopii useammalle organisaatiolle. [3]

Vaikka kudospatologian kuvat eivät vaadi yhtä huolellista kuvan käsittelyä ja anonymisointia, kuin esimerkiksi pään tietokone tomografiakuvat, joista on poistettava potilaan kasvonpiirteitä [11]. Täytyy kudospatologiankin kuvista anonymisoimiseksi poistaa kaikki tunnistamisen mahdollistava metadata. Tämän lisäksi kuvat vaativat muutakin esikäsittelyä, erityisesti niiden koon takia. Esikäsittelyllä voidaan pienentää käsiteltävien kuvien kokoa ja vähentää ohjelman tarvitsemaa laskennallista tehoa. [14]

Adnan ym. soveltaa menetelmässään monivaiheista oppimista ja valmistelee digipatologian kuvat pilkkomalla ne pienemmiksi ja ryhmittelemällä osat. Opetusdata ryhmitellään niin, että jokaisella ryhmällä on otsikko niiden sisällöstä. Ryhmät toteutetaan poistamalla kuvista kohdat, jotka eivät sisällä kudosta. Jäljelle jäävät osat pilkotaan ja kuvasta muodostuu ”kuvamosaiikki”. Kuvan osat ryhmitellään kategorian perusteella. Jokaisesta kategoriasta valittiin sattumanvaraisesti tietty määrä osia monivaiheisen oppimisen ryhmään. Kuvassa 6 näkyy vasemmalla alkuperäinen kudospatologian kuva ja oikealla siitä muodostettu kuvamosaiikki. [3]



Kuva 6: *Kuvamosaiikin muodostaminen* ***[3]***

Hajautettua oppimista voidaan soveltaa moneen kudospatologian tehtävään. Esimerkiksi sairauksien luokitteluun, hoidon toimimisen arviointiin ja ennusteen tekemiseen. Hajautetun oppimisen avulla saadaan sovellettua koneoppimisen malleja kudospatologiaan, joka mahdollistaa sellaisten ominaisuuksien huomaamisen, jota ihminen ei huomaisi. [14]

Lu ym. tutki hajautetun oppimisen soveltamista kudospatologian näytteiden luokitteluun. Kahden luokan lajittelutehtävässä hajautettua oppimista sovellettiin rintasyövän näytteiden luokitteluun. Useamman luokan lajittelutehtävässä tutkittiin munuaissyövän luokittelua kolmeen eri alatyyppiin. [14]

Datan valmistelu suoritetiin hyvin samaan tapaan kuin Adnan ym. [3] mallissa. Kuvista poistetaan tausta, ne pilkotaan ja ryhmiteltiin. Data jaettiin kolmeen osaan havainnollistamaan hajautetun oppimisen toimimista usean organisaation välillä. Data jaettiin suunnilleen yhtä suuriin osiin, jolloin datan epätasaista jakautumista osallistujien välillä ei päästä tutkimaan. [14]

Tutkimuksissa todettiin, että monesta lähteestä tehtyjen mallien tarkkuus oli huomattavasti parempi kuin vain yhden organisaation kuvilla tehty malli. He huomasivat myös hajautetun oppimisen tarkkuuden olevan erittäin lähellä ja vertailukelpoinen keskitetyn mallin kanssa. Samoja tuloksia saatiin myös, kun dataan lisättiin satunnaista häiriötä yksityisyyden suojan parantamiseksi. [14]

Kong ym. tutkivat hajautetun oppimisen soveltamista eturauhassyövän diagnostiikkaan. Eturauhassyöpä on miesten toiseksi yleisin syöpä. Tällä hetkellä diagnostiikka toteutetaan tutkimalla kudospatologian kuvia manuaalisesti. Tämä on kuitenkin hidasta ja kallista. Hajautetun oppimisen soveltaminen prosessiin voisi nopeuttaa sen automatisaatiota. Poiketen Lu ym. [14] tutkimuksesta Kong ym. jakoi datansa useampaan osaan tutkien seitsemän osallistujan hajautetutta oppimista. Kong ym. huomioi myös datan

epätasaisen jakautumisen ja kokeili menetelmän toimintaa erilaisilla datajakaumilla. [25]

Yksityisyydensuoja mallissa luodaan häiriön lisäämisen avulla. Mallissa lisätään Gaussian kohinaa paikallisten mallien parametreihin ennen niiden yhdistämistä. Näin malliin saadaan luotua lokaali differentiaalinen yksityisyys ja potilaiden yksityisyys voidaan turvata. [25]

Kong ym. aloittaa tutkimuksen valmistelemalla ja keräämällä tietokannat. Opetustietokantana käytetään sairaaloiden ja julkisten tietokantojen kuvia. Menetelmän testaaminen toteutetaan sairaalan ja julkisen tietokannan avulla. Opetusdata valmistellaan poistamalla kuvista taustat ja pilkkomalla kuvat. Kuvista erotellaan ja ryhmitellään niiden sisältämät ominaisuudet. Näin kuvista saadaan käyttöön niiden oleelliset osat ja niistä saadaan tarpeeksi pieniä opettamista varten. Datan valmistelu on hyvin samankaltainen kuin aineimmin mainituissa Adnan ym. [3] ja Lu ym. [14] menetelmissä. [25]

Tutkimuksen tulokset ovat samankaltaisia kuin Lu ym. [14] tutkimuksessa. Hajautetun oppimisen avulla saavutetaan huomattavasti parempi tulos, kuin vain yhden organisaation toteuttamalla paikallisella koneoppimisen mallilla. Malli ei saavuttanut yhtä hyvää tulosta kuin perinteisen keskitetyn koneoppimisen malli. Tulos oli kuitenkin kilpailukykyinen perinteiseen menetelmään verrattuna. Testidatan jakautumisesta ja variaatiosta riippuen tuloksen tarkkuus vaihteli hieman. Heikoimmat tulokset saatiin, kun testaamiseen käytettiin todellista kliinistä tietokantaa, jolloin datan variaatio oli suurinta. Tämä oli kuitenkin odotettavissa ja tällä tietokannalla tarkkuus oli vain noin 3.5 % huonompi kuin perinteisessä koneoppimisen mallissa. [25]

7. YHTEENVETO

Työssä tarkasteltiin yksityisyydensuojan takaamista hajautetussa oppimisessa ja hajautetun oppimisen soveltamista digipatologiaan. Tietosuojan huomioiminen jo ohjelman suunnitteluvaiheessa on tärkeää, sillä terveydenhuollossa käsiteltävä data on hyvin henkilökohtaista. Aihe on haastava kokonaisuus, sillä yhden ominaisuuden parantuessa jokin toinen ominaisuus yleensä heikkenee. Tämän takia on tärkeää valita käytettävät tietosuojamenetelmät tarkasti käytettävän hajautetun oppimisen mallin ja opetusdatana ominaisuuksien perusteella.

Yhdistämällä erilaisia tietosuoja menetelmiä saadaan parempia tuloksia. Suojaamalla data anonymisointimenetelmillä ehkäistään potilaiden tunnistamista. Kun datan lisäksi suojataan muodostettava malli ja palvelimien välinen kommunikaatio kasvatetaan mallin tietosuojaa. Yhdistämällä salausmenetelmiä häiriömenetelmien kanssa mahdollistetaan pienemmän häiriön määrän lisääminen ja saadaan siten tarkempia malleja. Kuitenkin haasteita aiheuttaa edelleen muuttuvat käsitetyksemme tietosuojasta ja uusien tietosuojariskien ilmaantuminen. Menetelmä, jonka nyt ajatellaan olevan turvallinen, ei välttämättä ole sitä tulevaisuudessa.

Työssä tutkittujen sovellusten saamat tulokset tukevat hajautetun oppimisen hyödyntämistä. Tutkimuksissa osoitettiin menetelmän pystyvän hyvin kilpailukykyiseen suoriutumiseen verratessa perinteiseen keskitettyyn koneoppimisen malleihin. Datan keskittäminen ei kuitenkaan aina ole haluttu toimenpide, joten hajautettua oppimista verrattiin myös yhden organisaation toteuttamaan koneoppimisen malliin. Tässä hajautetun oppimisen menetelmällä saadut tulokset ja tarkkuus olivat huomattavasti parempia.

Hyvistä tuloksista huolimatta on hajautetun oppimisen todellisessa soveltamisessa terveydenhuoltoon vielä haasteensa. Datan laatu, jakautuminen osallistujien kesken ja eri organisaatioiden eri standardien aiheuttama variaatio aiheuttaa edelleen haasteita menetelmälle ja sen tarkkuudella. Kuitenkin kirjallisuuskatsauksen perusteella voidaan todeta menetelmän olevan hyvin lupaava ja lisätutkimuksen arvoinen.

LÄHTEET

[1] D. Stripelis *ym.*, ”Secure Federated Learning for Neuroimaging”, *arXiv.org*, 2022.

[2] B. Camajori Tedeschini *ym.*, ”Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation”, *IEEE access*, vsk. 10, ss. 8693–8708, 2022, doi: 10.1109/ACCESS.2022.3141913.

[3] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, ja H. R. Tizhoosh, ”Federated learning and differential privacy for medical image analysis”, *Sci Rep*, vsk. 12, nro 1, Art. nro 1, helmi 2022, doi: 10.1038/s41598-022-05539-7.

[4] X. Yin, Y. Zhu, ja J. Hu, ”A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions”, *ACM computing surveys*, vsk. 54, nro 6, ss. 1–36, 2022, doi: 10.1145/3460427.

[5] O. Aouedi, A. Sacco, K. Piamrat, ja G. Marchetto, ”Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions”, *IEEE journal of biomedical and health informatics*, vsk. PP, ss. 1–14, 2022, doi: 10.1109/JBHI.2022.3185673.

[6] N. Rieke *ym.*, ”The future of digital health with federated learning”, *npj Digit. Med.*, vsk. 3, nro 1, s. 119, syys 2020, doi: 10.1038/s41746-020-00323-1.

[7] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, ja F. Wang, ”Federated Learning for Healthcare Informatics”, *Journal of healthcare informatics research*, vsk. 5, nro 1, ss. 1–19, 2021, doi: 10.1007/s41666-020-00082-4.

[8] T. White, E. Blok, ja V. D. Calhoun, ”Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed”, *Human Brain Mapping*, vsk. 43, nro 1, ss. 278–291, 2022, doi: 10.1002/hbm.25120.

[9] U. Shah, I. Dave, J. Malde, J. Mehta, ja S. Kodeboyina, ”Maintaining Privacy in Medical Imaging with Federated Learning, Deep Learning, Differential Privacy, and Encrypted Computation”, Piscataway: IEEE, 2021, ss. 1–6. doi: 10.1109/I2CT51068.2021.9417997.

[10] G. Singh, V. Violi, ja M. Fisichella, ”Federated Learning to Safeguard Patients Data: A Medical Image Retrieval Case”, *Big data and cognitive computing*, vsk. 7, nro 1, ss. 18-, 2023, doi: 10.3390/bdcc7010018.

[11] G. A. Kaissis, M. R. Makowski, D. Rückert, ja R. F. Braren, ”Secure, privacy-preserving and federated machine learning in medical imaging”, *Nature machine intelligence*, vsk. 2, nro 6, ss. 305–311, 2020, doi: 10.1038/s42256-020-0186-1.

[12] X. Gong *ym.*, ”Federated Learning with Privacy-Preserving Ensemble Attention Distillation”, *IEEE transactions on medical imaging*, vsk. PP, ss. 1–1, 2022, doi: 10.1109/TMI.2022.3213244.

[13] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, ja B. A. y Arcas, ”Communication-Efficient Learning of Deep Networks from Decentralized Data”. arXiv, 26. tammikuuta 2023. Viitattu: 19. helmikuuta 2023. [Verkossa]. Saatavissa: http://arxiv.org/abs/1602.05629

[14] M. Y. Lu *ym.*, ”Federated learning for computational pathology on gigapixel whole slide images”, *Medical image analysis*, vsk. 76, ss. 102298–102298, 2022, doi: 10.1016/j.media.2021.102298.

[15] Y. Chen, X. Qin, J. Wang, C. Yu, ja W. Gao, ”FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare”, *IEEE Intelligent Systems*, vsk. 35, nro 4, ss. 83–93, heinä 2020, doi: 10.1109/MIS.2020.2988604.

[16] J. Li *ym.*, ”A Federated Learning Based Privacy-Preserving Smart Healthcare System”, *IEEE transactions on industrial informatics*, vsk. 18, nro 3, ss. 2021–2031, 2022, doi: 10.1109/TII.2021.3098010.

[17] R. Wang, J. Lai, Z. Zhang, X. Li, P. Vijayakumar, ja M. Karuppiah, ”Privacy-Preserving Federated Learning for Internet of Medical Things Under Edge Computing”, *IEEE Journal of Biomedical and Health Informatics*, vsk. 27, nro 2, ss. 854–865, helmi 2023, doi: 10.1109/JBHI.2022.3157725.

[18] Q. Yang, Y. Liu, T. Chen, ja Y. Tong, ”Federated Machine Learning: Concept and Applications”, *ACM transactions on intelligent systems and technology*, vsk. 10, nro 2, ss. 1–19, 2019, doi: 10.1145/3298981.

[19] *Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuoja-asetus) (ETA:n kannalta merkityksellinen teksti)*, vsk. 119. 2016. Viitattu: 2. huhtikuuta 2023. [Verkossa]. Saatavissa: http://data.europa.eu/eli/reg/2016/679/oj/fin

[20] E. P. Oy, ”FINLEX ® - Ajantasainen lainsäädäntö: Tietosuojalaki 1050/2018”. https://www.finlex.fi/fi/laki/ajantasa/2018/20181050 (viitattu 14. maaliskuuta 2023).

[21] *Euroopan parlamentin ja neuvoston asetus (EU) 2017/745, annettu 5 päivänä huhtikuuta 2017, lääkinnällisistä laitteista, direktiivin 2001/83/EY, asetuksen (EY) N:o 178/2002 ja asetuksen (EY) N:o 1223/2009 muuttamisesta sekä neuvoston direktiivien 90/385/ETY ja 93/42/ETY kumoamisesta (ETA:n kannalta merkityksellinen teksti)ETA:n kannalta merkityksellinen teksti*. 2017. Viitattu: 29. maaliskuuta 2023. [Verkossa]. Saatavissa: http://data.europa.eu/eli/reg/2017/745/2017-05-05/fin

[22] *Euroopan parlamentin ja neuvoston asetus (EU) 2017/746, annettu 5 päivänä huhtikuuta 2017, in vitro -diagnostiikkaan tarkoitetuista lääkinnällisistä laitteista sekä direktiivin 98/79/EY ja komission päätöksen 2010/227/EU kumoamisesta (ETA:n kannalta merkityksellinen teksti)ETA:n kannalta merkityksellinen teksti*. 2017. Viitattu: 30. maaliskuuta 2023. [Verkossa]. Saatavissa: http://data.europa.eu/eli/reg/2017/746/2017-05-05/fin

[23] ”FINLEX ® - Viranomaisten määräyskokoelmat: Lääkealan turvallisuus- ja kehittämiskeskus - 11.03.2022 FIMEA/2022/000832”. https://www.finlex.fi/fi/viranomaiset/normi/558001/47981?search%5Btype%5D=pika&search%5Bpika%5D= (viitattu 2. huhtikuuta 2023).

[24] S. Truex *ym.*, ”A Hybrid Approach to Privacy-Preserving Federated Learning”, teoksessa *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, teoksessa AISec’19. New York, NY, USA: Association for Computing Machinery, marras 2019, ss. 1–11. doi: 10.1145/3338501.3357370.

[25] F. Kong *ym.*, ”Federated contrastive learning models for prostate cancer diagnosis and Gleason grading”. arXiv, 16. helmikuuta 2023. Viitattu: 23. huhtikuuta 2023. [Verkossa]. Saatavissa: http://arxiv.org/abs/2302.06089