Ida Toivanen

Ennustavat mallit terveydenhuollossa

Tietotekniikan kandidaatintutkielma

23. lokakuuta 2018

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Ida Toivanen

Yhteystiedot: ida.m.toivanen@student.jyu.fi

Työn nimi: Ennustavat mallit terveydenhuollossa

Title in English: Predictive models in health care

Työ: Kandidaatintutkielma

Sivumäärä: 32+0

Tiivistelmä: Onnellisen elämän edellytyksenä voidaan pitää terveyden ylläpitoa. Ennustavat mallit ovat viime aikoina olleet yhä enemmän osallisena terveyteen liittyvässä tutkimuksessa ja terveydenhuollossa. Terveysriskien ehkäisemiseen pyrkivät menetelmät voivat olla esimerkiksi koneoppimista tai tilastotiedettä hyödyntäviä, ja käyttää tietoaineistonaan kliinisiä, geeni- ja/tai psykososiaalisia tekijöitä. Riskien ennaltaehkäisemiseen liittyy myös ongelmia, jotka tulee ottaa huomioon muutoksia terveydenhuoltoon suunnitellessa.

Avainsanat: data-analyysi, ennustava malli, logistinen regressio, syvät neuroverkot, kliiniset tekijät, geenitekijät, psykososiaaliset tekijät

Abstract: Leading a happy life requires one to take care of one's own health. Recently predictive models have contributed more and more to health care systems and the research of health. To prevent health risks methods such as machine learning or statistics and data can be based on clinical, genetic and/or psychosocial factors. Still there lays problems related to prevention that cannot be disregarded but must be taken into account.

Keywords: data analysis, predictive model, logistic regression, deep learning, clinical research, genomics, psychosocial factors

Sisältö

1 Johdanto

Parempi katsoa kuin katua – vanhan sanalaskun mukainen viisaus tuntuu saavan tukea myös 2010-luvun terveyttä kartoittavalta tutkimukselta, jonka analysoimisen apuna on käytetty perinteisempien menetelmien lisäksi myös yhä enemmän neuroverkkoja. Ihmisen odotettavissa olevan eliniän lisäksi terveydenhoidon tutkimuksessa on oltu kiinnostuneita esimerkiksi todennäköisyydestä sairastua tiettyyn sairauteen, terveysriskien vaikutuksesta terveydentilaan ja terveysongelmien ennaltaehkäisystä. Terveyden tutkimus ei jää pelkästään kliinisten tekijöiden vaikutuksen kuvaamiseen, vaan moninaisiin näkökulmiin kuuluvat myös varhaislapsuuden ja -aikuisuuden sekä sosioekonomisen aseman vaikutus terveys- tai kuolleisuusriskeihin (psykososiaaliset tekijät) sekä genomin tutkimus tautien syiden kartoittamisessa. Suuressa riskissä olevien ihmisten tunnistamista varten onkin kehitetty ennustavia malleja, joiden toimintaperiaatteet nojaavat tilastotieteellisiin menetelmiin, koneoppimiseen sekä nykyisin myös syviin neuroverkkoihin (eng. deep learning). Menetelmästä riippuen joudutaan kuitenkin tekemään kompromisseja tulosten tulkittavuuden sekä tarkkojen ja luotettavien tulosten saamisen välillä.

Tässä kirjallisuuskatsauksessa tarkastellaan ennustavien mallien käyttöä terveyden tutkimuksessa ottamalla huomioon myös terveyden moninaiset lähtökohdat. Kirjallisuuskatsauksen luvussa 2 kuvataan terveyden tutkimusta kliinisten, geneettisten ja psykososiaalisten tekijöiden kannalta. Luvussa 3 syvennytään lähinnä regressiomallia käyttäviin tutkimuksiin, jotka pyrkivät mittaamaan terveyttä riski-indekseillä kokonaisvaltaisesti, ja kuvataan vielä syvillä neuroverkoilla tehtyä data-analyysiä terveydenhuollossa. Luvussa 4 regressiomallista etenkin logistista ja syvistä neuroverkoista konvolutiivista käydään tarkemmin läpi, ja sivutaan myös niiden vahvuuksia ja mahdollisia ongelmiakohtia. Lopuksi yhteenvedossa kiteytetään vielä ilmiötä ja pohditaan menetelmien käyttöönoton riskejä ja mahdollisuuksia tulevaisuudessa.

2 Terveyttä mittaavia riski-indeksejä

Terveyden tutkimus on keskittynyt lähinnä empiirisesti mitattaviin kliinisiin tekijöihin, mutta sitä on myös poikkitieteellisesti tarkasteltu niin geneettisestä kuin psykososiaalisesta näkökulmastakin. Terveyden monisyisen luonteen kuvaamista varten tässä luvussa käydään ensin läpi näitä yksittäisiä kliinisiä, geneettisiä ja psykososiaalisia tekijöitä ja sitten kuvaillaan terveysriskejä laskevia ennustavia malleja.

2.1 Kliiniset tekijät

Kliinisten tekijöiden tutkimus antaa arvokasta tietoa ihmisten terveydentilan hoidosta ja siitä, kuinka terveysriskejä voidaan välttää olemalla niistä tietoisempia ja toimimalla tiedon perusteella. Niiden käyttäminen on olennaisessa osassa kuolleisuusriskin laskemisessa ja terveysriskejä kartoittavissa tutkimuksissa. Esimerkiksi Lozanon johtamassa tutkimuksessa (2012) kartoitettiin vuosien 1990 ja 2010 yleisimpiä kuolinsyitä maailmanlaajuisesti. Kaiken kaikkiaan 235 kuolinsyytä kartoittavan tutkimuksen perusteella vuonna 2010 kohtalokkaimpia yksittäisiä kuolinsyitä olivat sepelvaltimotauti, aivoinfarkti, keuhkoahtaumatauti, alahengitystieinfektio, keuhkosyöpä ja HIV/AIDS. Yleisemmällä tasolla tartuntataudit, raskaudenaikaiset syyt, vastasyntyneisiin vaikuttaneet syyt ja ravitsemukselliset syyt käsittivät 24,9% kuolemista maailmanlaajuisesti vuonna 2010 ((Lozano yms. 2012); katso myös (Lim yms. 2012)).

Terveysriski-indeksistä puhuttaessa voidaan viitata myös tulevaisuudessa tapahtuviin terveysongelmiin, joiden mahdollisuutta kartoitetaan ennustavilla malleilla. Ennustavilla malleilla pyritään vaikuttamaan terveyskäyttäytymiseen lisäämällä interventioita perinteisesti käytetyn parantavana hoidon rinnalle. Näin saadaan mahdollisesti tehokkaampia tapoja terveyden ylläpidolle ja samalla tehtyä säästöjä terveydenhoitokustannuksissa (ennustavista malleista lisää luvussa 3). Terveysriskejä kartoittavalla tutkimuksella voidaan vaikuttaa myös preventiolla, jolloin on tarkoitus ennakoida riskejä ja tehdä muutoksia koko populaatiotasolla (Fineberg 2013).

Näin ihmisen hoitokeinoja tunnistamalla ja tutkimusresursseja keskittämällä voidaan edistää tehokkaammin ihmisten hyvinvointia ja terveyttä (Wendler 2017).

Terveydentilan kartoittamista varten käytössä ovat olleet esimerkiksi ihmisen terveysriskiarviointi (eng. human health risk assessment, HRA), jossa arvioidaan omaa tämänhetkistä terveyttä kartoittamalla kliinisiä terveysriskejä. Sieck ja Dembre (2014) arvioivat HRA:n käyttöönoton vaikutuksia hoitoon hakeutumisessa kolme vuotta kestävässä tarkkailussa, jossa HRA:ta hyödyntävä koehenkilöryhmä käytti kontrolliryhmää enemmän terveyspalveluita ja myös halvemmin kokonaiskustannuksin. Tutkimuksen HRA:ta varten kerättiin kahdeksan biometristä mittaa (paino (lb), systolinen verenpaine (mmHg), diastolinen verenpaine (mmHg), kokonaiskolesteroli (mg/dL), HDL-kolesteroli (mg/dL), LDL-kolesteroli (mg/dL), triglyseridit (in mg/dL) ja glukoosi (mg/dL)), joista koeryhmä onnistui parantamaan seitsemää (Sieck ja Dembe 2014). Oman terveyskäyttäytymisen tiedostaminen ja riskitekijöiden kontrollointi onkin tärkeässä osassa terveyden ylläpidossa.

2.2 Geenitekijät

Kliinisten tekijöiden lisäksi terveyteen vaikuttavista tekijöistä geenien osuutta on tutkittu laajasti etenkin sairauksien puhkeamisen mahdollisia syitä etsittäessä. Näihin kuuluvat genominlaajuiset assosiaatiotutkimukset (eng. genome-wide association study, GWAS), joissa keskitytään sairauksien diagnosointiin solutasolla geenimarkkereita tunnistamalla (Ganna yms. 2013). Geenimarkkereiksi (eng. genetic marker) voidaan luokitella joko geenien alleelit tai koodamattomien kromosomaalisten alueiden alleelit, jotka tunnistavat tietyn kromosomin alueen (Sanders ja Bowman 2014). Geenimarkkereita ovat esimerkiksi yksittäiset nukleotidipolymorfismit (eng. single nucleotide polymorphism, SNP) eli SNP-merkit, jotka ovat DNA-sekventointeja vertaamalla esille saatuja nukleotidiparien eroja (Sanders ja Bowman 2014). Esimerkiksi Gannan tutkimuksessa (2013) tarkasteltiin yhteensä 707:stä yleisestä SNP-merkistä luodun geneettisen arvon yhteyttä 125:een sairauteen tai kuolleisuuden riskitekijään. Coxin mallia (eng. Cox proportional model) käyttämällä tutkijat havaitsivat perimän vaikuttavan tietyissä sairauksissa (keuhko-, paksusuolen-, rinta-

ja eturauhassyövässä sekä diabeteksessa, sydänkohtauksessa, sepelvaltimotaudissa ja sydämen vajaatoimintassa) kuolleisuuteen, vaikkakin perimän ja kyseisten sairauksien esiintyvyyden välinen yhteys oli vieläkin vahvempi. Tutkimus yhtyykin näkemykseen siitä, että kuolleisuudella ja sairauksilla on monimutkainen geneettinen pohja (Ganna yms. 2013).

Perimän ymmärtäminen on myös yksi varteenotettava kanta terveyden tutkimuksessa, jotta sairauden käyttäytymistä voitaisiin ennakoida helpommin ennustavilla malleilla. Etenkin syvät neuroverkot osoittautuvat käteviksi genomiikassa perimän monimutkaisuuden ja korkeadimensioisuutensa vuoksi. Esimerkiksi IBM:n Clinical Genomics tai Watson Health on yhdistänyt sähköistä terveysrekisteridataa ja genomiikkaa (Eggebraaten yms. 2007) pääosin kaupallistetuksi tuotteeksi, jolloin terveydenhuollon edistämistä varten genomiikan viitekehykseen on saatu lisättyä myös kliininen tutkimuspuoli. Genomiikkaa käsittelevistä ennustavista malleista on lisää luvussa 3.2.

2.3 Psykososiaaliset tekijät

Kliinisten ja geenitekijöiden lisäksi myös psykososiaalisen taustan vaikutusten arvioiminen yksilön terveyteen ja elämänlaatuun kuuluu terveyden kokonaisvaltaisen tarkasteluun. Psykososiaalisiksi tekijöiksi voidaan luokitella tekijöitä, jotka vaikuttavat ihmisen psyyken kehittymiseen ja mielenterveyteen niin yksilöllisestä kuin yhteisöllisestäkin näkökulmasta.

Nuorten aikuisten (18–29-vuotiaiden) lapsuuden aikaisia tapahtumia kartoitettiin esimerkiksi Kestilän väitöskirjassa (2008), jossa niiden yhteyttä tutkittiin myöhemmän iän erilaisiin terveysongelmiin. Tutkimuksessa esimerkiksi havaittiin, että mitä alempi koulutustaso nuorilla aikuisilla oli, sitä todennäköisemmin he arvioivat oman terveytensä keskiverroksi tai heikoksi. Huomattavan osan koulutuksellisia eroja selittivät lapsuuden aikaiset sosiaaliset elinolosuhteet, ja lapsuudessa koetut vastoinkäymiset (kuten vanhempien avioero, säännöllinen työttömyys tai alkoholi- ja mielenterveysongelmat) olivat yhteydessä nuorten aikuisten koettuun henkiseen

rasitteeseen ja itsearvioituun heikkoon terveyteen (Kestilä 2008). Teini-ikäisten terveysriskitekijöitä on tutkittu myös Sussmanin johtamassa tutkimuksessa (1995), jossa havaittiin 7. luokan riskitekijöiden vaikuttavan vuotta myöhemmin 8. luokalla. Tutkimuksessa muodostettiin hyvinvoinnin indikaattori, joka kuvasi oman maailman ymmärtämistä tai hyväksymistä, ja subjektiivisen distressin indikaattori, jolla viitattiin esimerkiksi suhteellisen alhaiseen perheen konfliktitasoon ja alhaiseen vertaisryhmän sosiaaliseen vaikutukseen. Tutkimuksessa huomattiin, että 7. luokalla hyvinvoinnin indikaattorilla ja subjektiivisen distressin indikaattorilla oli positiivinen yhteys, ja ongelmakäyttäytymisellä, alemmalla sosioekonomisella asemalla ja fyysisen terveyden laiminlyönnillä oli negatiivinen yhteys terveyskäyttäytymiseen vuotta myöhemmin (Sussman yms. 1995).

Yksilön koulutus- ja tulotasoa kuvaavan sosioekonomisen aseman vaikutusta on tutkittu myös niin terveyden tilaan kuin kuolleisuuteenkin (Adler yms. 1994), mutta jätetty pois esim. WHO:n ei-tartuntatautien prevention ja kontrolloinnin maailmanlaajuisesta toimintasuunnitelmasta (eng. the WHO Global Action Plan for the Prevention and Control of Non-Communicable Diseases) (Stringhini yms. 2017). Stringhini tutkimusryhmineen (2017) laajassa multikohorttitutkimuksessa ja meta-analyysissä selvittikin 25 x 25 riskitekijän ja etenkin sosioekonomisen aseman vaikutusta elinikään miehillä ja naisilla eri ikäluokissa. Tutkimuksessa 40–85–vuotiailla elinikää lyhensi alempi sosioekonominen asema 2.1 vuodella, tämänhetkinen tupakointi 4.8 vuodella, diabetes 3.9 vuodella, fyysinen aktiivittomuus 2.4 vuodella, hypertensio 1.6 vuodella, ylipaino 0.7 vuodella ja korkea alkoholin nauttiminen 0.5 vuodella. Alempi sosioekonominen asema oli myös ylempää sosioekonomista asemaa enemmän yhteydessä suurempaan kuolleisuuteen.

Osa terveysriskiarvioinneista yhdistää kliinisten riskitekijöiden lisäksi myös psykososiaalisia tekijöitä terveydentilan itsearviointilomakkeisiin. Phillipsin (2014) luomassa strukturoidussa terveysriskiarvioinnissa (tai MOHR, My Own Health Report) psykososiaalisina tekijöinä olivat ahdistus tai huoli, stressi ja masennus kliinisten BMI:n, pikaruoan, hedelmien ja vihannesten saannin, sokerijuomien nauttimisen, fyysisen aktiivisuuden, uneliaisuuden määrän, tupakan ja alkoholin käytön,

laittomien huumeiden ja reseptilääkkeiden väärinkäytön sekä perusterveyden arvion lisäksi. Tutkimuksessa kova stressi oli yleisin (noin 20 %:lla tutkimushenkilöistä) psykososiaalinen tekijä (Phillips yms. 2014).

Näiden lisäksi kannattaisi tarkastella myös mahdollisten kulttuurisidonnaisten aspektien vaikutusta terveyteen ja diagnosointiprosessiin. Monissa ennustavissa malleissa psykososiaaliset tekijät oli jätetty täysin pois muuttujista niiden muihin muuttujiin verrattuna heikon selittävyyden vuoksi. Koulutuksen kaltaisten tekijöiden tarkastelu antaa kuitenkin arvokasta tietoa siitä, kuinka terveysongelmien huomioiminen kannattaisi nuoresta iästä lähtien. Lapsuuden aikaisten vastoinkäymisten ja toksisen stressin tutkiminen (katso (Shonkoff yms. 2011) tai Adverse Childhood Experiences eli ACE-study) voisi ajatella olevan avainasemassa tässäkin, niin turvallisen lapsuuden, aivojen tuetun kehityksen kuin terveyteen liittyvien asenteiden vaikuttamisen ja terveyskasvatuksenkin näkökulmasta.

3 Kokonaisvaltaiset riski-indeksit

Tutkimusta yksittäisten terveyttä vaarantavien tekijöiden vaikutuksesta yksilön elämänkaareen löytyy runsaasti, mutta menetelmien parantuessa terveysriskien tarkempi kuvaaminen usean riskitekijän pohjalta on yleistynyt ja luo aiempaa kokonaisvaltaisemman kuvan terveydestä. Tällaiset riskifaktorit ovat usein kvantitatiivisesti mitattavissa olevia kliinisiä tekijöitä. Tässä alaluvussa käydään läpi kaksi kuolleisuusriskiä ja yksi useamman vuosikymmenen ajan käytetty terveyshyötyindeksi, ja lopuksi mainitaan ohimennen muita mahdollisia terveyden mittoja.

3.1 Perinteisiä menetelmiä hyödyntävät

Lim työryhmineen (2015) on muodostanut tieteellistä kirjallisuutta ja NHANES -ja BRFSS -datakantoja hyödyntäneen riskimallin, jossa lasketaan kokonaiskuolleisuusriski (eng. overall mortality risk) kuvaamaan 12 erilaisen riskitekijän vaikutusta yksilön kuolleisuuteen. Myös vältettävissä olevalle kuolleisuudelle laskettiin riskiarvo (eng. avoidable mortality risk), joka tässä sivuutetaan. Tutkimuksen tarkoituksena oli tunnistaa ja seurata suurimmassa terveysriskissä olevia ihmisiä ja näin suunnata preventiopalveluita niitä tarvitseville. Tutkimuksessa käytettiin riski-indeksien laskemisessa vuosien 2003–2010 NHANES-dataa ja vuosien 2006–2008 BRFSS-dataa, ja validoinnissa vuosien 1988–1994 ja 1999–2004 NHANES-datoja. Edellistä tutkimustietoa käyttäen valittiin yhteensä 12 riskitekijää (tupakointi, alkoholin käyttö, fyysinen aktiivisuus, ylipaino, korkea verenpaine, korkea kolesterolipitoisuus, korkea veren glukoosipitoisuus, turvavyön käyttö sekä hedelmien, kasvisten, pähkinöiden/siementen ja omega-3 rasvojen määrä ruokavaliossa), joista kerättiin dataa kyselylomakkein kokonaiskuolleisuusriskin laskemista varten. Seuraavissa kaavoissa (2.1–2.3) esitellään muodostettu kuolleisuutta ennustava riskiarvo seuraavalle (1) vuodelle ja seuraavalle kymmenelle vuodelle (2.4) (Lim yms. 2015).

Kokonaiskuolleisuusriski eli yksilön todennäköisyys kuolla yhden vuoden aikana

on

$$TQ_{iast} = 1 - \prod_{c}^{C}(1 - TQ_{iasct}), \tag{3.1}$$

jossa

$$TQ_{iasct} = RR_{ic}CQ_{a+t,sc} \tag{3.2}$$

on yksilön i todennäköisyys kuolla syystä c 1-vuotisen ajanjakson t aikana, kun hän on a:n vuoden ikäinen ja ja sukupuolta s. $CQ_{a+t,sc}$ on todennäköisyys kuolla syystä c, joka ei johdu mistään 12 riskitekijästä, kun ryhmä on ikääntynyt $t + a$ verran ja on sukupuolta s. Kaavassa oleva suhteellinen kokonaiskuolleisuusriski RR_{ic} on

$$RR_{ic} = e^{\sum_{r}^{R}(\beta_{re}(L_p - T_p))}, \tag{3.3}$$

joka edustaa yhdistettyä 12 riskitekijän altistuksen vaikutusta verrattuna ei-altistukseen yksilölle i syystä c. Yhdistämällä eri vuosien riskit saatiin kokonaiskuolleisuusriski laskettua seuraavalle kymmenelle vuodelle kaavasta

$$TQ10_{ias} = TQ_{iasc0} + \sum_{j=1}^{10} TQ_{iascj} \prod_{j=0}^{j-1}(1 - TQ_{iascj}) \tag{3.4}$$

Riski-indeksi validointiin käyttämällä apuna erottelua (eng. discrimination), jolloin laskettu ROC-käyrän alla olevan alue (eng. alue under ROC-curve, AUC) kuvasi riskimallin kykyä erottaa seurantajakson aikana kuolleet ja selvinneet ihmiset, sekä kalibrointia (eng. calibration, χ^2), jolla arvioitiin riski-indeksin kykyä ennustaa havaittu riskitaso eri populaation desiileissä. AUC-alue on mitta arvojen 0.5 ja 1 välissä, jossa 0.5 viittaa hyvin epätarkkaan luokittelijaan (ei tulisi pitää ilmiötä kuvaavana) ja 1 hyvin tarkkaan luokittelijaan. Erikseen sekä miehille että naisille lasketut AUC-arvot osoittivat riski-indeksin olevan melko tarkka (AUC = 0.84, keskivirhe 0.01), ja χ^2 viittaisi miehillä hyvin ennustettuun (χ^2 = 12.3, p = .196) ja naisilla suurimmassa desiilissä hieman yliarvioituun riskiarvoon (χ^2 = 22.8, p = .002).

Limin yms. (2015) tutkimuksen riski-indeksin laskentatapaa ja validointimenetelmiä sovellettiin myös toisessa tutkimuksessa (Massaro yms. 2017), jossa arvioitiin kattavan terveysriskin käyttöä polikliinisessä hoidossa käyttäen kahta eri datalähdettä, Framingham ja NHANES-tietokantoja. Massaron yms. (2017) tutkimuksessa

riskitekijöiksi valikoitu yhteensä 12, joista 4 kliinistä (verenpaine, painoindeksi, veren glukoosipitoisuus, kolesteroli), 4 ravitsemuksellista (kasvikset, hedelmät, pähkinät, omega-3) ja 3 behavioraalista (tupakan poltto, liikunta, turvavyön käyttö) tekijää. Riski-indeksi validoitiin NHANES-datalla (miehillä AUC = 0.84, χ^2 = 12.3; naisilla AUC = 0.84, χ^2 = 22.8), joka tuotti samanlaisia tuloksia kun Lim yms. (2015) tutkimuksessa, ja Framingham-datalla (miehillä AUC 0.74, χ^2 = 12.5; naisilla AUC 0.73, χ^2 = 21.6), tuottaen alkuperäistutkimusta heikompia tuloksia (AUC-erotukset 0.10-0.11) Massaro yms. (2017).

Italiassa sijaitsevasta 3.7 miljoonan asukkaan Emilia-Romagnan alueesta kerättyä terveyspalveluiden käytöstä koostuvaa tietokantaa käytettiin Louis'n johtamassa (2014) tutkimuksessa, joissa mahdollisten potilaiden tunnistamista varten kehiteltiin ennustava logistinen regressiomalli. Monimuuttujaisella logistisella regressiomallilla laskettiin mittavan datan pohjalta sairaalaan joutumisen tai kuolemisen riski (eng. risk of admission), joka kuvaa niiden potilaiden sairaalaan ottoa, joiden terveysongelmat olisi mahdollisesti hoidettavissa paremmalla potilaan hoidolla. Tutkimuksen tarkoituksena on käyttää mallia ehkäisemään mahdolliset kalliit terveydenhuoltokulut interventioilla ja hoidon paremmalla suunnittelulla (Louis yms. 2014).

Tutkimuksessa riippuvana muuttujana (eng. dependent variable) käytettiin sairaalaan otettujen lukumäärää tapauksissa, jossa oikealla potilaan hoidolla ongelman pahentuminen olisi voitu lykätä tai pysäyttää, tai ongelma olisi voinut olla kokonaan vältettävissä, sekä tapauksissa, joissa potilas on kuollut mistä tahansa syystä sairaalassa tai sen ulkopuolella vuonna 2012. Riippumattomat muuttujat (eng. independent variable) saatiin vuosien 2004–2011 väestötieteellisestä datasta (ikä, sukupuoli, asuinpaikan maantieteellinen sijainti) sekä kotihoidon datasta, apteekin käytön datasta ja sairaalasta pääsyä kuvaavasta abstraktista datasta, joista kartoitettiin tautikategoriat vahingoittuneen kehon systeemin tai osan perusteella. Tautikategorian laajuuden takia regressiomallia ei tässä erikseen esitellä, mutta regressiomallin perusmallin voi nähdä luvusta 4.1.

Mallia varten asetettiin riskirajat hyvin suuri riski (HSR, yli 25%) ja suuri riski (SR, 15–24%). Suurimmassa riskissä olevassa desiiliryhmässä (viimeinen kymmenesosa)

ennustettiin olevan keskimäärin 23.9%:n terveysriski oikean terveysriskin ollessa 24.2% (eli vain 0.3% erotus). Muissa desiileissä erotus oli 0.0–0.1%. Tietyt sairaudet olivat hyvin suuren riskin ja suuren riskin ryhmässä huomattavasti yleisempiä kuin aikuispopulaatiossa. Näihin sairauksiin kuuluivat sydän- ja verisuonitaudit (HSR = 84.2%, SR = 77.1% ja aikuispopulaatio = 26%), ruoansulatustaudit (HSR = 65.4%, SR = 49.3% ja aikuispopulaatio = 15.6%), syövät (HSR = 20.9%, SR = 10.6 % ja aikuispopulaatio = 2.7%) ja mielenterveyshäiriöt (HSR = 34.2%, SR = 25.0% ja aikuispopulaatio = 7.8%).

Tutkimuksessa käytettiin vuosien 2004–2012 dataa niin, että vuoden 2012 tulokset ennustettiin vuosien 2004–2010 terveyshistoriadataa (yhteensä 83 diagnoosikategoria/tasomuuttujaa ja 11 apteekinkäyttöä kuvaavaa muuttujaa) ja vuoden 2011 terveydenhuollon dataa käyttämällä. Näin pystyttiin vertaamaan varsinaista vuonna 2012 kerättyä dataa vuotta 2012 kuvaavaan ennustettuun dataan ja laskea mallin ennustamisen tarkkuus. Mallin evaluointia varten laskettiin herkkyys (HSR = 0.298 ja HSR+SR = 0.471; eng. sensitivity), tarkkuus (HSR = 0.981 ja HSR+SR = 0.951; eng. specificity) ja positiivinen ennustearvo (HSR = 0.411 ja HSR+SR = 0.298; eng. positive predictive value) ja AUC-alue (vuodelle 2012 laskettu AUC = 0.856 ja vuodelle 2011 laskettu AUC = 0.853) hyväksi. Tutkimuksessa arvot määriteltiin seuraavasti: sensitiivisyys oli osuus sairaalaan otetuista, joiden ennustettiin joutuvan sairaalaan (eng. true positive rate), tarkkuus oli osuus sairaalaan ottamattomista, joita ei ennustettu otettavan sairaalaan (eng. true negative rate), ja positiivinen ennustearvo oli osuus sairaalan otetuista, jotka ennustettiin ja todellisuudessa otettiin sairaalaan.

Terveyshyötyindeksi (eng. health utilities index, HUI) on etenkin USA:ssa ollut paljon käytetty mitta terveydelle, joka lasketaan hyötyfunktiota käyttäen kyselylomakedatan pohjalle. Terveyshaittojen numeeristen arvojen mittaamisen sijaan HUI:lla arvioidaan yleisiä arkipäiväiseen elämään vaikuttavia tekijöitä (esimerkiksi puhekykyä ja kognitiivisia prosesseja). Hyötyindeksin versiosta riippuen terveydentilaa kuvaavat tekijät hieman vaihtelevat, joten keskitymme tässä uusimpaan HUI3 -indeksiin.

HUI3:ssa käytettyä kyselomakkeilla kartoitettiin 8 terveysattribuuttia: puhe ($b3$),

emootio/tunnetila ($b6$) ja kipu ($b8$) asteikoilla 1–5 ja näkökyky ($b1$), kuuloaisti ($b2$), kävelykyky ($b4$), näppäryys ($b5$), kognitio ($b7$) asteikolla 1–6 eli yhteensä 8 attribuuttia ($b1$–$b8$). Terveyshyötyindeksi 3 lasketaan tällöin kaavasta

$$Utility = 1.37 \prod_{i=1}^{8} b_i \tag{3.5}$$

jossa b_i ovat terveysattribuutit kun i = 1, 2, ..., 8, ja monimuuttujaisten utiliteettien (eng. utility) oletetaan viittaavan koko elämän läpikestäviin kroonisiin tiloihin ja jatkuvan muuttujan skaalan ollessa välillä 0.0 (kuollut) – 1.0 (terve) (Feeny yms. 2002). Hyötyindeksejä voidaan käyttää kustannushyötyanalyyseissä laatupainotettujen elinvuosien (eng. quality-adjusted life years, QALY) ja populaatiotutkimuksessa laatupainotetun elinodotteen (eng. quality-adjusted life expectancy) laskemiseen (Furlong yms. 2001). Niitä on laajasti käytetty myös kliinisissä tutkimuksissa, väestönlaajuisissa terveysgallupeissa ja ekonomisissa arvioinneissa (Feeny yms. 2002).

Limin (2015), Louis'n (2014) ja HUI:n kaltaisten terveyttä ja kuolleisuutta mittaavien indeksien lisäksi paljon olemassaolevaa tutkimusta on myös muista kuolleisuusriskeistä (eng. mortality risk), jotka yleensä kuvaavat ihmisen eliniän pituutta tiettyjen kliinisten tekijöiden perusteella (mm. (Yourman yms. 2012)). Myös psykiatriselle osastolle takaisinottoa (eng. psychiatric readmission) 30 päivän sisällä sairaalasta päästyä tutkittiin mallilla, jonka validointi tosin osoitti heikohkoja tuloksia (AUC = 0.631) (Vigod yms. 2015).

Sairaalahoitoon takaisinottoa (eng. hospital readmission) tutkiva kirjallisuuskatsaus (Kansagara yms. 2011) kävi läpi yhteensä 30 tutkimusta, joissa suurimmassa osassa laskettiin sairaalahoitoon takaisinottoa 30 päivän sisällä (yhteensä 14 tutkimusta, AUC = 0.55–0.65 eli huono erottelukyky). Muut tutkimukset käsittelivät suuressa riskissä olevien potilaiden tunnistamista (7 tutkimusta, AUC = 0.56–0.72), sairaalasta päästämistä (5 tutkimusta, AUC = 0.68–0.83) ja 6 tutkimusta teki mallien vertailua samassa populaatiossa, joista kahdessa todettiin funktionaalisten ja sosiaalisten muuttujien parantavan mallin erottelua.

3.2 Syviä neuroverkkoja hyödyntävät

Massadatan lisääntyessä ja datan käyttöönoton helpottumisen ansiosta voidaan neuroverkkoja suosia resursseiden salliessa niiden laskentatehon ja tarkkuuden vuoksi esimerkiksi ennuste- ja luokittelutehtävissä. Terveyden tutkimuksessa käytetyt syvät neuroverkot voidaan jakaa esimerkiksi eri datalähteiden perusteella sähköisiin terveysrekistereihin, kliiniseen kuvantamiseen, genomiikkaan ja matkapuhelindataan (Miotto yms. 2017). Tässä alaluvussa esitellään useampi sähköisiä terveysrekistereitä käyttänyt tutkimus, ja mainitaan lyhyesti tutkimuksesta muihin datalähteisiin liittyen.

Sähköiset terveysrekisterit sisältävät diagnoosikoodeja ja -toimenpiteitä, lääkitykseen ja väestöryhmään liittyvää dataa sekä laboratoriotuloksia sisältävää dataa (Yadav yms. 2018). Kun datalähteenä on käytetty sähköisiä terveysrekistereitä, voidaan syviä neuroverkkoja käyttävät kliiniset tutkimukset jakaa viiteen eri luokkaan: informaation eristykseen (eng. information extraction), piirreoppimiseen (eng. representation learning), tuloksen ennustamiseen (eng. outcome prediction), ilmiasuttamiseen tai fenotyypittämiseen (eng. phenotyping) ja tunnistuksen poistoon (eng. de-indentification) (Shickel yms. 2017). Näistä tässä keskitytään etenkin neuroverkkomalleihin, jotka ennustavat potilaiden tuloksia ja rikkaampien sairauskuvausten määrittelemiseen (fenotyypitys) pyrkiviin neuroverkkoihin.

Stanfordin, Kalifornian ja Chicago medicine yliopistojen sekä Googlen työntekijöiden yhteistyönä (Rajkomar yms. 2018) EHR-dataa pohjana käyttänyt tutkimus opetti syviä neuroverkkoja ennustamaan sairaalakuolleisuutta, pidentynyttä hoitoaikaa, kaikki potilaan diagnoosit sekä ei-suunniteltua sairaalaahoitoon takaisinottoa 30 päivän sisällä. Kolme käytettyä neuroverkkoarkkitehtuuria olivat pitkä lyhytkestoinen muisti (eng. long short-term memory, LSTM), tarkkaavaisuuspohjainen aikaa-autoenkoodaava neuroverkko (eng. attention-based TANN eli temporal autoencoding neural network) ja neuroverkko, joka oli tehostettu aikaan pohjautuvilla päätöstyngillä (eng. neural network with boosted time-based decision stumps). Neuroverkot opetettiin FHIR-formaattiin (fast healthcare interoperability resources) perustuvalla raakadatalla, jonka volyymi oli kooltaan 46 miljardia datapistettä. Neu-

roverkot opetettiin joka tehtävään erikseen yhdistämällä arkkitehtuurit (eng. ensemble learning) ja validoitiin käyttämällä tunnisteetonta (de-identified) dataa. Mittaamalla tilastolliset AUC-arvot huomattiin neuroverkkojen olevan tarkkoja tai hyvin tarkkoja jokaista mitattavaa arvoa kohden. Hyvin tarkoiksi (AUC yli 0.80) lukeutuivat sairaalakuolleisuus 24 tuntia sairaalaanoton jälkeen (AUC = 0.93–0.94), kaikki potilaan diagnoosit (potilaan sairaalasta päästyä) (AUC = 0.90) ja pidentynyt hoitoaika (AUC = 0.85–0.86), sekä tarkoiksi (AUC = 0.7–0.8) äkillinen sairaalaahoitoon takaisinotto 30 päivässä (AUC = 0.75–0.76) (luottamusväli kaikissa 95 %).

Myös RETAIN (reverse time attention model eli ajassa taaksepäin tarkasteleva) -malli (Choi yms. 2016) on kehitelty ennustamaan erilaisia terveysongelmia, kuten sydämenpysähdyksiä, kun taustalla olevat merkittävät kliiniset muuttujat ovat esimerkiksi edellisiin diagnooseihin, lääkärikäynteihin ja niistä johdettuihin tietoihin liittyviä. Tätä kaksikerroksista neuroverkkoa (neural attention model) testattiin suurella terveysjärjestelmällä, joka sisälsi dataa 263 000 potilaasta ja 14 miljoonasta sairaalakäynnistä. Mallin tarkkuus mitattiin negatiivisellä log-todennäköisyydellä (eng. negative log-likelihood, NLL) ja AUC:lla. Kun verrattiin eri menetelmien suoriutumista sydämenpysähdyksen ennustamisessa, logistiseen regressioon (AUC = 0.7789–0.8011, NLL = 0.3164–0.3374) verrattuna RETAIN suoriutui paremmin (AUC 0.8624–0.8786, NLL = 0.2479–0.2645), mutta RETAIN:in kanssa vertailtavana voidaan pitää myös rekurrentteja neuroverkkoja (eng. recurrent neural network, RNN) hyödyntäviä menetelmiä, RNN+α_M (AUC = 0.8545–0.8703, NLL = 0.2609–0.2773) ja RNN+α_R (AUC = 0.8637–0.8797, NLL = 0.2517–0.2693).

IBM:n ja Connecticutin yliopiston työntekijöiden (Cheng yms. 2016) opettama konvoluutioneuroverkko (eng. convolutional neural network, CNN) pyrki tarkempien sairauskuvausten määrittelemisen fenotyypittämisellä. Fenotyypittämisellä viitataan yleensä uusien fenotyyppien eli ilmiasujen löytämiseen tai olemassaolevien kuvausten tarkentamiseen niin, että ne ovat hienojakoisempia ja entistä tarkempia. Ennen neljän kerroksen neuroverkon rakentamista jokainen potilas esitettiin temporaalimatriisina heidän EHR-datojensa perusteella. Mallin validointi tuotti melko tarkkoja tuloksia, kun fenotyypityksen kohteena olivat sydämen vajaatoiminta (eng. con-

gestive heart failure, CHF) ja keuhkoahtaumatauti (eng. chronic obstructive pulmonary disease, COPD). Kun opetusdatasta käytettiin 90%, oli SF-CNN tarkin menetelmä sydämen vajaatoiminnan (AUC = 0.7223–0.8125) ja keuhkoahtaumataudin (AUC = 0.6838–0.79382) kuvauksen laatimisessa.

Muissa tutkimuksissa EHR-dataa on käytetty esimerkiksi dynaamisen muistimallin (LSTM nimeltä DeepCare) ennustettavaan lääkitykseen potilashistorian perusteella (Pham yms. 2016).

Kliinisessä kuvantamisessa voidaan hyödyntää monia kuvantamismenetelmiä, kuten magnetoenkefalografiaa (MEG), enkefalografiaa (EEG) tai vaikka magneettikuvantamista (eng. magnetic resonance imaging, MRI) ja PET-kuvausta (positroniemissiotomografia), jolla on pystytty seuraamaan Alzheimerin taudin etenemistä aivokuvia ottamalla. Liu työryhmineen (2014) eristi MRI-kuvista aivojen harmaat alueet ja PET-kuvista CMRGlc-arvot (eng. cerebral metabolic rate for glucose consumption eli aivoperäinen/serebraalinen metabolinen aste glukoosin kulutukselle), jotka toimivat piirteinä Alzheimerin taudin variaatioiden eri tasojen tunnistamisessa. Tutkimuksessa diagnoosikuvauksen koostamista varten kehitetty osaohjattu (eng. semi-supervised) neuroverkko koostui pinotuista harvoista autoenkoodereista (eng. stacked sparse autoencoders, ssAE) ja softmax regressiokerroksesta. Tutkimuksessa tunnistettiin yhteensä 83 ROI-aluetta (eng. region of interest), joissa Alzheimerin taudin eteneminen näkyy. Validointi suoritettiin mittaamalla tarkkuus (eng. accuracy, ACC), herkkyys (eng. sensitivity, SEN) ja spesifisyys (eng. specificity, SPE) ja tutkimuksessa verrattiin neuroverkkoa myös yksittäiskerneliseen sekä monikerneliseen tukivektorikoneeseen (eng. single/multi-kernel support vector machine). Neuroverkko (ACC = 87.76, SEN = 88.57, SPE = 87.22) suoriutui yksikernelistä (ACC = 84.40, SEN = 86.64, SPE = 84.31) ja monikernelistä tukivektorikonetta (ACC = 86.42, SEN = 84.98, SPE = 87.83) hieman paremmin (Liu yms. 2014).

Tämän lisäksi neuroverkkoja on käytetty esimerkiksi dermatologi-tasoisen ihosyövän luokitteluun (CNN:llä) (Esteva yms. 2017).

DeepBind on syvä konvoluutioverkko, jonka Alipanahi tutkimusryhmineen (2015)

on kehittänyt ennustamaan dna ja rna-sidosproteiinien sekvenssispesifisyyksiä haastamaan perinteiset positiopainomatriiseja (eng. position weight matrices, PWM) käyttävät lähestymistavat. Kehon säätelysysteemien toimintaa kuvaavia spesifisyyksiä on visualisoitu myös tietyssä sekvenssissä geenivariaatioiden vaikutuksia tapahtuvaan sitomisen osoittavilla mutaatiokartoilla sekä PWM-kokonaisuuksilla. Vertailussa muihin vastaaviin algoritmeihin ohitti DeepBind ne sekä PBM- ja AUC-arvojen vertailussa (kalibrointivaiheen riistinvalidoinnissa AUC = 0.70 ja testauksen loppuvaiheessa AUC = 0.93) (Alipanahi yms. 2015).

Genomiikan tutkimuksessa neuroverkkoja on käytetty terveydenhuollossa myös esimerkiksi syövän luokittelussa geenin ilmentymisen profiileja käyttämällä (pinotulla harvalla autoenkooderilla, eng. stacked sparse autoencoder) (Fakoor yms. 2013) ja erilaisten kromatiinimerkkien esiintyvyyden estimoinnissa (CNN) (Koh yms. 2016).

Fotopletysmografiasignaalien (eng. photoplethysmography, PPG) tunnistamista on käytetty terveyden tarkkailussa Jindalin (2016) johtamassa tutkimuksessa, jossa rajoitetun Boltzmannin koneen (eng. restricted Boltzmann machine, RBM) ja syväuskoverkon (eng. deep belief network, DBN) yhdistelmäarkkitehtuuria käytettiin biometriikan tunnistustehtävässä. Esikäsittelyvaiheen jälkeen tutkijat käyttivät medoidien ympärysten ositusklusterointia (eng. partition around medoids clustering), jonka jälkeen he hyödynsivät RBM-verkkoa ohjaamattomasti sekä ohjatusti. Malli sai luokiteltua PPG-signaalit hyvin tarkasti (ACC = 0.961) (Jindal yms. 2016).

RBM:n lisäksi konvolutiivisia neuroverkkoja on käytetty esimerkiksi energiankulutuksen estimoinnissa puettavia sensoreita käyttämällä (CNN) (Zhu yms. 2015) ja unenlaadun ennustamisessa valveillaoloaikaan kerätyn fyysisen aktiivisuuden perusteella (CNN) (Sathyanarayana yms. 2016).

4 Menetelmien analysointi

Niin kuin minkä tahansa mallin kehittaminen, myös ennustavien mallien rakentaminen on moniportainen prosessi. Kehittämisen vaiheet voi jakaa esimerkiksi viiteen: ennustemallin valmisteluun (suunnitteluvaihe), mallin opettamisessa ja testaamisessa käytettävän datan valitsemiseen ja siitä saatavien muuttujien käsittelyyn sekä mallin generointiin ja evaluointiin sekä validointiin (Lee yms. 2016). Menetelmästä riippuen kuitenkin mallin rakentamisessa vaiheet eivät välttämättä seuraa toisiaan lineaarisesti vaan voivat myös limittyä tai toistua. Tässä luvussa käydään lyhyesti läpi (logistisen) regressiomallin ja (konvolutiivisten) syvien neuroverkkojen perusominaisuuksia pitäytyen terveydenhoidon kontekstissa.

4.1 Logistinen regressiomalli

Regressiomalli on yksi suosituimmista kahden tai useamman tekijän suhdetta mallintavista menetelmistä. Käyttökelpoisuus ja tehokkuus mallissa vähentyy yleensä yli 10 muuttujan malleissa (Lee yms. 2016), joten monimutkaistenkin ilmiöiden kuvaaminen pyritään tekemään vähemmällä muuttujamäärällä monimuuttujamalleissa. Siinä missä lineaarinen regressiomalli näyttäytyy koordinaatistossa lineaarisena suorana, Bernoullin jakaumaa noudattava aineisto muodostaa logistisessa regressiomallissa S-kirjaimen muotoisen kuvaajan. Logistista regressiomallia hyödynnetään tapauksissa, joissa riippuva muuttuja on dikotominen (saa arvon 1 tai 0) ja mallin kertoimia voidaan käyttää estimoimaan ristitulosuhde (eng. odds ratio) jokaiselle mallin riippumattomalle muuttujalle (Alexopoulos 2010), joten se sopii tällöin etenkin terveydenhuollon mallien kehittämiseen. Ristitulosuhteet kertovat riippumattoman muuttujan kyvystä selittää riippuvaa muuttujaa. Tässä kohtaa on pidettävä mielessä odds-termin käännös, joka voi vaihtua kontekstista riippuen (ei varsinaisesti viittaa todennäköisyyteen niin kuin termi probability). Logistisen mallin vahvuus on kaikkien muuttujien yhtäaikainen käsittely, jolloin voidaan välttää tulosten vääristymistä (eng. confounding; riippumattomien muuttujien vaikutus toisiinsa ja riippuvaan niin, että se aiheuttaa tulosten väärentymistä) (Sperandei

2014).

Regressiomallin kehittämistä varten voidaan joko kerätä kaikki mahdolliset ilmiötä kuvaavat muuttujat ja vähentää/pitää mukana/lisätä niiden määrää erilaisilla menetelmillä, tai aloittaa vain vakiotermillä (eng. intercept) ja rakentaa malli lisäämällä muuttujia testaamalla ne yksittäin (Sperandei 2014). Sopivan regressiomallin löytämistä varten voidaan käyttää esimerkiksi askeltavaa regressiota (eng. stepwise criteria), jossa joka askeleella lisätään olennaisia muuttujia kuitenkin ennen lisäyksiä poistamalla mahdollisesti tarpeettomia muuttujia. Uuden mallin sovituksessa jokainen lisäyksistä ja poistoista on erillinen askel. Esimerkiksi Louis'n (2014) tutkimuksessa askeltavan regressiolla saatiin kehiteltyä malli (Louis yms. 2014), jolloin vähemmällä muuttujamäärällä saatiin robustimpi joukko olennaisia riippumattomia muuttujia. Regressiomallia käyttäessä sosioekonominen asema voisi olla esimerkki sellaisesta muuttujasta, joka ei yksinään kuvaa ilmiötä tarpeeksi kattavasti tai tuo paljoa lisäarvoa (Lee yms. 2016), joten se on yleensä kuolleisuusriskiä kuvaavissa malleissa ohitettukin kliinisiin muuttujiin enemmän keskittyen.

Logistisen regressiomalli on yleensä muotoa (Sperandei 2014)

$$\text{todennäköisyys} = \frac{mahdollisuus}{1 + mahdollisuus}, \tag{4.1}$$

jossa todennäköisyys (eng. probability) esitetään ja mahdollisuus (eng. chance).

Eksponenttifunktion avulla esitettynä logistinen funktio on muotoa (Šarlija yms. 2018)

$$p = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_r x_r}}{1 + e^{\beta_0+\beta_1 * x_1+\beta_2 x_2+\ldots+\beta_r x_r}}, \tag{4.2}$$

missä p on todennäköisyys, β_0 on vakiotermi, β_1, β_2, ..., β_r ovat regressiokertoimet ja x_1, x_2, ..., x_r ovat riippumattomat muuttujat, kun $r \in \mathbb{N}$. Tässä todennäköisyys p on se todennäköisyys, että riippuva muuttuja on 1. Regressiokertoimien estimointia varten lasketaan yleensä logistinen muunnos

$$logit(p) = ln(\frac{p}{1-p}) = ln(e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_r x_r}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_r x_r, \tag{4.3}$$

missä $\frac{p}{1-p}$ on odds eli vetokertoimien/todennäköisyyksien suhde ja sen logaritmia kutsutaan log odds:ksi.

Mallin suorituskyvyn arvioimisessa apuna käytetään tilastollisia mittoja (mallin evaluointi ja validointi), joista mainittakoon etenkin edellä käytetyt erottelu (ROC/AUC-käyrä), herkkyys ja spesifisyys (eng. sensitivity and specificity, SEN ja SPE), kalibraatio (eng. calibration: calibration plot, Hosmer-Lemeshow testi), (positiivinen ja negatiivinen) uskottavuusosamäärä (eng. likelihood ratio) sekä tarkkuus (eng. accuracy) (Lee yms. 2016).

Regressiomallin vahvuutena ovat melko tarkat ja helposti tulkittavat tulokset, joten sen suosio tuskin täysin hiipuu ja väistyy uusien menetelmien tieltä. Regressiomallia käyttäessä on kuitenkin hyvä varoa alisovittumista (eng. underfitting) ja multikollineaarisuutta (eng. multicollinearity) sekä koittaa pitää otosjoukko mahdollisimman suurena ja havainnot riippumattomina tarkempien tulosten ja virheiden sattumisen varalta (Garson 2016).

4.2 Syvät neuroverkot

Laskennallisesti raskaan datan (eng. big data) prosessointia varten on tarvittu menetelmiä, jotka analysoivat ja oppivat datasta luotettavasti myös sen raa'assa, käsittelemättömässäkin muodossa (ilman tunnisteita, ohjaamaton oppiminen, eng. unsupervised learning; vrt. ohjattu oppiminen, eng. supervised learning). Tätä varten on kehitelty esimerkiksi (teko)neuroverkkoja (eng. artificial neural network; yleensä yhden piilotetun kerroksen sisältäviä kolmikerroksisia tai vielä yksinkertaisempia neuroverkkoja) ja monimutkaisempia syviä neuroverkkoja, jotka eroavat niistä huomattavasti suuremmalla kerrosmäärällään (Miotto yms. 2017). Eri tyyppisiä datoja varten onkin kehitelty monenlaisia eri neuroverkkoja, joita ovat esimerkiksi monikerroksinen perseptroni (lyhenne MLP), rekurrentti neuroverkko (RCC; esim. LSTM), konvolutiivinen neuroverkko (CNN), erilaiset Boltzmannin koneet (BM; esim. RBM) sekä autoenkooderit (AE; esim. stacked sparse autoencoders, ssAE). Näiden arkkitehtuurien lisäksi on kehitetty erilaisia variaatioita ja neuroverkkoarkkitehtuurien yhdistelmiä datan ominaisuuksista ja tutkimuksen tavoitteista riippuen. Tässä kirjallisuuskatsauksessa huomioon otetut terveydenhuollon tutkimukset ovat suosineet etenkin konvolutiivisia neuroverkkoja, jota käydään läpi myös

tässä hieman tarkemmin.

Konvolutiivinen neuroverkko on kehitetty erityisesti tunnistamaan kaksiulotteisia muotoja olemalla korkealla tasolla muuttumaton erilaisille vääristymisen muodoille, kuten vinoutumiselle, väärin skaalautumiselle tai kääntämiselle (Haykin 2009). Esimerkiksi Chengin (2016) tutkimuksessa rakennettiin 4-tasoinen konvoluutioneuroverkko (esitelty luvussa 3.2) tautien kuvausten määrittelemistä varten suoraan sähköisistä terveysrekistereistä eli EHR-datasta. Fenotyypitykseen pyrkivässä neuroverkon ensimmäisessä kerroksessa jokaisen potilaan EHR-data esitettiin temporaalimatriisina (dimensiot: aika ja tapahtuma), josta neuroverkon toisen kerroksen yksipuolinen konvoluutiokerros irroitti fenotyyppejä. Kolmannessa kerroksessa, maksimivarantointikerroksessa (eng. max pooling layer), esitellään havaittujen fenotyyppien harvuutta, jotta vain tärkeimmät fenotyypit saadaan eroteltua ja jätettyä malliin. Viimeisessä kerroksessa yhdistyvät kerrokset softmax-ennustekerrokseksi.

Syvät neuroverkot ovat menetelmiä, jotka ovat laskennallisesti kalliimmasta päästä ja tarvitsevat paljon dataa tarkkojen tulosten takaamiseksi. Tästä huolimatta niiden käyttäminen terveydenhuollon sovellutuksissa voi osoittautua ongelmalliseksi, sillä vain pieni osa terveysdatasta on saatavilla ja senkin opettaminen on haastavaa datan heterogeenisyyden (eng. heterogeneity), tulkinnanvaraisuuden, kohinaisuuden (vääristynyt tai korruptoitunut data; eng. noisiness) ja monimuotoisuuden takia (Miotto yms. 2017; Alipanahi yms. 2015; Cheng yms. 2016). Datan heterogeenisyys, eli selittävien tekijöiden paljous, sekä tiedon puuttuminen sairauksien syistä ja etenemisestä kielii ilmiön monimutkaisuudesta. Datan laadussa ongelmia voi tuottaa lisäksi korkeadimensioisuus (eng. high-dimensionality), vinous (eng. bias), toisteisuus (eng. redundancy), puuttuvat arvot ja harvuus (eng. sparsity). Terveysdatan käsittelyssä olisi hyvä ottaa huomioon sen temporaalisuus (eng. temporality, time dependant), millä viitataan terveydentilan muuttumiseen ajan suhteen. Monissa syvissä neuroverkoissa käytetään staattisia vektoripohjaisia syötteitä, joten ajan epädeterministisyyttä ei oteta huomioon vaan aikaa käsitellään epäluonnollisesti (Miotto yms. 2017). Kehittämisen varaa on myös tulkittavuuden lisäämisessä, joka olisi erityisen tärkeää terveydenhoidossa ja siinä, että mallien tekemät päätökset

tehtävistä toimenpiteistä (kuten lääkkeiden määräämisen tai tietyn sairastumisriskin laskemisen) saisivat tukea myös terveydenhuollon ammattilaisilta (Miotto yms. 2017).

Tehokkaampien ja robustimpien mallien kehittämistä varten kannattaa siis kiinnittää huomiota datan laatuun ja määrään. Malleja voisi parantaa esimerkiksi rajoituksilla, maksimiaktivaatiolla (eng. maximum activation), kvalitatiivisella klusteroinnilla ja mimiikkaoppimisella (eng. mimic learning) (Shickel yms. 2017). Syviä neuroverkkomalleja voitaisiin kehittää myös piirteiden rikastamisella (eng. feature enrichment), liittoumapäättelyllä (eng. federated inference), ajallisuuden mallinnuksella sekä kiinnittämällä huomiota mallin yksityisyyteen ja asiantuntijoiden tiedon sisällyttämiseen (Miotto yms. 2017).

5 Yhteenveto

Lisäämällä ennustavia malleja terveydenhuoltoon voitaisiin ennakoida terveysongelmia ja eritellä terveydenhoitokuluja entistä tehokkaammin. On hyvä kuitenkin pohtia eri menetelmien käytön luotettavuutta, tarkkuutta ja vaaroja, sekä terveysriskien ehkäisyn vaikutusta jo olemassaoleviin järjestelmiin.

Vaikka tietoisuuden lisääminen oman terveyden tilasta muuttaa itsessään terveyskäyttäytymistä (Sieck ja Dembe 2014), on vähintään yhtä tärkeää tietää miten tehdä preventioita yhdenmukaisesti ja laadukkaasti. Preventioiden toteuttamista käsittelevä Fineberg (2013) luettelee sairauksien ehkäisyn ongelmakohdiksi esimerkiksi sen, että preventiivinen hoito voi olla ristiriidassa kaupallisten intressien kanssa. Kaupallisista intresseistä riippuen tuottoon pyrkivät yritykset saattavat puoltaa tai vastustaa ehkäisevää hoitoa, sillä hyödyt eivät usein keräänny prevention maksajalle vaan asiakkaille (Fineberg 2013). Terveydenhoitojärjestelmä perustuu enemmän parantavan hoitoon, mikä näkyy myös terveysasemien, sairaaloiden ja terveydenhoidon ammattilaisten vastuissa ja työnkuvassa. Prevention lisäämistä varten muutoksia olisi tehtävä siis suuremman kaavan mukaan, joten niiden toteuttaminen voisi osoittautua lopulta hyvin mittavaksi.

Kustannustehokkuuden lisäksi ehkäisevän hoidon suunnittelussa on otettava huomioon sattuviin virheisiin kohdistuvat ennakkoasenteet, jatkuva terveyskäyttäytymisen tarkkailu ja palkitsemisen viive (Fineberg 2013). Terveyden aikasidonnaisuus vaikuttaa (algoritmien kehittämisen lisäksi) myös preventioon, sillä se on terveyden ylläpitoon pyrkivä jatkuva prosessi. Palkitseminen ei tässä toimi myöskään samalla lailla kuin parantavassa hoidossa (saa hoitoa -> olo tokenee) vaan terveyskäyttäytymisen muuttamisessa olennaisessa osassa on terveyskasvatus. Terveyskasvatuksen vaikutus ei yksilöllisistä eroista johtuen ole kaikilla sama, joten luottamus tiedonantajiin voi loppujen lopuksi olla preventiossakin yksi määrittelevimmistä tekijöistä. Luottamuksen lisäksi erilaiset kulttuurilliset, uskonnolliset ja henkilökohtaiset uskomukset vaikuttavat ratkaisevasti ehkäisevään hoitoon, eikä eettisiltä kysymyksiltäkään vältytä. Esimerkiksi järjestelmien rajattua käyttöä ja tietojen väärinkäytön

mahdollisuutta on mietittävä ja sitä, kuinka pitkälle itsemääräämisoikeus ulottuu. Kaikesta tästä huolimatta on muistettava, että terveysongelmien ehkäisy lähtee ensisijaisesti omasta itsestä ja terveyden ylläpitoon vaikuttavat vahvasti myös yksilön omat motiivit ja tausta.

Kirjallisuutta

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). *Socioeconomic status and health: the challenge of the gradient*. American psychologist, 49(1), 15 http://dx.doi.org/10.1037/0003-066X.49.1.15

Alexopoulos, E. C. (2010). *Introduction to multivariate regression analysis*. Hippokratia, 14(Suppl 1), 23. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049417/

Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). *Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning*. Nature biotechnology, 33(8), 831. https://www.nature.com/articles/nbt.3300

Cheng, Y., Wang, F., Zhang, P., & Hu, J. (2016). *Risk prediction with electronic health records: A deep learning approach*. In Proceedings of the 2016 SIAM International Conference on Data Mining (pp. 432-440). Society for Industrial and Applied Mathematics. https://pdfs.semanticscholar.org/8942/804fe4e2425758ab68df4ff80a2cac1987b8.pdf

Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., & Stewart, W. (2016). *Retain: An interpretable predictive model for healthcare using reverse time attention mechanism*. In Advances in Neural Information Processing Systems (pp. 3504-3512). http://papers.nips.cc/paper/6321-retain-an-interpretable-predictive-model-for-healthcare-using-pdf

Eggebraaten, J. T., Tenner, J. & C. Dubbels, J. (2007). *A health-care data model based on the HL7 Reference Information Model*. IBM Systems Journal. 46. 5 - 18. https://ieeexplore.ieee.org/document/5386594/

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, 542(7639), 115. https://www.nature.com/articles/nature21056

Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013, June). *Using deep learning to enhance cancer diagnosis and classification*. In Proceedings of the International Conference on Machine Learning (Vol. 28). https:

//www.researchgate.net/publication/281857285_Using_deep_learning_to_enhance_cancer_diagnosis_and_classification

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., ... & Boyle, M. (2002). *Multiattribute and single-attribute utility functions for the health utilities index mark 3 system*. Medical care, 40(2), 113-128. https://www.ncbi.nlm.nih.gov/pubmed/11802084

Fineberg, H. V. (2013). *The paradox of disease prevention: celebrated in principle, resisted in practice*. Jama, 310(1), 85-90. https://www.ncbi.nlm.nih.gov/pubmed/23821092

Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). *The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies*. Annals of medicine, 33(5), 375-384 http://www.chepa.org/files/working%20papers/01-02.pdf

Ganna, A., Rivadeneira, F., Hofman, A., Uitterlinden, A. G., Magnusson, P. K., Pedersen, N. L., ... & Tiemeier, H. (2013). *Genetic determinants of mortality. Can findings from genome-wide association studies explain variation in human mortality?*. Human genetics, 132(5), 553-561. https://www.ncbi.nlm.nih.gov/pubmed/23354976

Garson, G. D. (2016). *Logistic Regression: Binomial and Multinomial*. 2016 Edition. Asheboro, NC: Statistical Associates Publishers.

Haykin, S. (2009). *Neural networks and learning machines*. (Vol. 3). Upper Saddle River, NJ, USA:: Pearson.

Jindal, V., Birjandtalab, J., Pouyan, M. B., & Nourani, M. (2016). *An adaptive deep learning approach for PPG-based identification*. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 6401-6404. https://www.ncbi.nlm.nih.gov/pubmed/28269713

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). *Risk prediction models for hospital readmission: a systematic review*. Jama, 306(15), 1688-1698. https://jamanetwork.com/journals/jama/fullarticle/1104511

Kestilä, L. (2008). *Pathways to health: determinants of health, health behaviour and*

health inequalities in early adulthood. https://helda.helsinki.fi/handle/10138/23542

Koh, P. W., Pierson, E., & Kundaje, A. (2016). *Denoising genome-wide histone ChIP-seq with convolutional neural networks*. Bioinformatics, 33(14), i225-i233. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870713/

Lee, Y., Bang, H., & Kim, D. J. (2016). *How to Establish Clinical Prediction Models*. Endocrinology and Metabolism, 31(1), 38–44. http://doi.org/10.3803/EnM.2016.31.1.38

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., & Aryee, M. (2012). *A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010*. The lancet, 380(9859), 2224-2260. https://www.sciencedirect.com/science/article/pii/S0140673612617668?via%3Dihub#

Lim, S. S., Carnahan, E., Nelson, E. C., Gillespie, C. W., Mokdad, A. H., Murray, C. J., & Fisher, E. S. (2015). *Validation of a new predictive risk model: measuring the impact of the major modifiable risks of death for patients and populations*. Population health metrics, 13(1), 27. https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-015-0059-8

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., & Feng, D. (2014). *Early diagnosis of Alzheimer's disease with deep learning*. In Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on (pp. 1015-1018). IEEE. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6868045

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., ... Memish, Z. A. (2012). *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010*. The Lancet, 380(9859), 2095-2128. https://www.sciencedirect.com/science/article/pii/S0140673612617280?via%3Dihub

Louis, D. Z., Robeson, M., McAna, J., Maio, V., Keith, S. W., Liu, M., ... & Grilli, R. (2014). *Predicting risk of hospitalisation or death: a retrospective population-based analysis*. BMJ open, 4(9), e005223. https://www.ncbi.nlm.nih.gov/pmc/

articles/PMC4166245/

Massaro, J. M., Murabito, J. M., Au, R., Carnahan, E., Morgan, T. S., Murray, C., ... & D'Agostino Sr, R. B. (2017). *Evidence on the Validity of a Comprehensive Health Risk Index and Implications for Ambulatory Care and Population Health Management*. The Journal of ambulatory care management, 40(4), 297-304. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1732044/

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). *Deep learning for healthcare: review, opportunities and challenges*. Briefings in bioinformatics. http://dudleylab.org/wp-content/uploads/2017/05/Deep-learning-for-healthcare-review-opportunities-and-challenges.pdf

Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2016). *Deepcare: A deep dynamic memory model for predictive medicine*. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 30-41). Springer, Cham. https://arxiv.org/abs/1602.00357

Phillips, S. M., Glasgow, R. E., Bello, G., Ory, M. G., Glenn, B. A., Sheinfeld-Gorin, S. N., ... & Krist, A. H. (2014). *Frequency and prioritization of patient health risks from a structured health risk assessment*. The Annals of Family Medicine, 12(6), 505-513. http://www.annfammed.org/content/12/6/505.full

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Sundberg, P. (2018). *Scalable and accurate deep learning with electronic health records*. npj Digital Medicine, 1(1), 18. https://www.nature.com/articles/s41746-018-0029-1

Sanders, M. F., & Bowman, J. L. (2014). *Genetic analysis: An integrated approach*. Pearson Education.

Šarlija, N., Bilandžić, A., & Stanic, M. (2018). *Logistic regression modelling: procedures and pitfalls in developing and interpreting prediction models*. Croatian Operational Research Review, 8(2), 631-652. https://hrcak.srce.hr/ojs/index.php/crorr/article/view/5311

Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., ... & Taheri, S. (2016). *Correction of: sleep quality prediction from wearable*

data using deep learning. JMIR mHealth and uHealth, 4(4). https://www.ncbi.nlm.nih.gov/pubmed/27815231

Sieck, C. J., & Dembe, A. E. (2014). A 3-year assessment of the effects of a self-administered health risk assessment on health care utilization, costs, and health risks. Journal of occupational and environmental medicine, 56(12), 1284-1290. https://www.ncbi.nlm.nih.gov/pubmed/25479298

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). *Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis*. IEEE Journal of Biomedical and Health Informatics. https://arxiv.org/pdf/1706.03446.pdf

Shonkoff, J. P., Garner, A. S., Committee on Psychosocial Aspects of Child and Family Health, & Committee on Early Childhood, Adoption, and Dependent Care. (2011). *The lifelong effects of early childhood adversity and toxic stress*. Pediatrics, peds-2011. https://pediatrics.aappublications.org/content/129/1/e232

Stringhini, S., Carmeli, C., Jokela, M., Avendaño, M., Muennig, P., Guida, F., ... & Chadeau-Hyam, M. (2017). *Socioeconomic status and the 25 x 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women*. The Lancet, 389(10075), 1229-1237. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5368415/

Sperandei, S. (2014). *Understanding logistic regression analysis*. Biochemia medica: Biochemia medica, 24(1), 12-18. https://www.ncbi.nlm.nih.gov/pubmed/24627710

Sussman, S., Dent, C. W., Stacy, A. W., Burton, D., & Flay, B. R. (1995). *Psychosocial predictors of health risk factors in adolescents*. Journal of Pediatric Psychology, 20(1), 91-108. https://www.researchgate.net/publication/15308472_Psychosocial_Predictors_of_Health_Risk_Factors_in_Adolescents

Vigod, S. N., Kurdyak, P. A., Seitz, D., Herrmann, N., Fung, K., Lin, E., ... & Gruneir, A. (2015). *READMIT: a clinical risk index to predict 30-day readmission after discharge from acute psychiatric units*. Journal of psychiatric research, 61, 205-213. https:

//www.ncbi.nlm.nih.gov/pubmed/25537450

Wendler, D. (2017).*The Ethics of Clinical Research*, The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/clinical-research/

Yadav, P., Steinbach, M., Kumar, V., & Simon, G. (2018). *Mining electronic health records (EHRs): a survey*. ACM Computing Surveys (CSUR), 50(6), 85. https://arxiv.org/pdf/1702.03222.pdf

Yourman, L. C., Lee, S. J., Schonberg, M. A., Widera, E. W., & Smith, A. K. (2012). *Prognostic indices for older adults: a systematic review*. Jama, 307(2), 182-192. https://jamanetwork.com/journals/jama/fullarticle/1104837

Zhu, J., Pande, A., Mohapatra, P., & Han, J. J. (2015). *Using deep learning for energy expenditure estimation with wearable sensors*. In E-health Networking, Application & Services (HealthCom), 2015 17th International Conference on (pp. 501-506). IEEE. https://ieeexplore.ieee.org/document/7454554