

Tarja Laitinen ja Arho Virkki

## Potilastiedon louhinta mullistaa tutkimuksen ja hoidon

Tekoäly on osa modernia tilastotiedettä, ja sitä on sovellettu pitkään monilla teollisuudenaloilla. Koneet voidaan ohjelmoida tekemään tarkkoja päätöksiä nopeasti ja väsymättä, kunhan merkityksellistä opeusmateriaalia on tarpeeksi. Tekoälyyn perustuvien tutkimusmenetelmien soveltaminen lääketieteeseen hyödyttää sekä potilasta että yhteiskuntaa, sillä näin pystytään tuottamaan entistä vaikuttavampia hoitoja nykyistä tehokkaammin. Tieteellisen avoimuus on välttämätöntä modernien analyysimenetelmien kehittämisessä, mutta avoimuus pitäisi laajentaa koskemaan myös oheisohjelmistoja. Näin voidaan riippumattomasti varmistaa, että menetelmien tuottamat tulokset ovat lääketieteellisesti oikeita, tilastollisesti pitäviä ja tietoturvallisesti tuotettuja. Tiedon hyödyntäminen on vasta alussa, mutta tekniikkaakin tärkeämpää on ymmärtää korkealaatuisen tiedon välttämättömyys hyvän hoidon tuottamisessa.

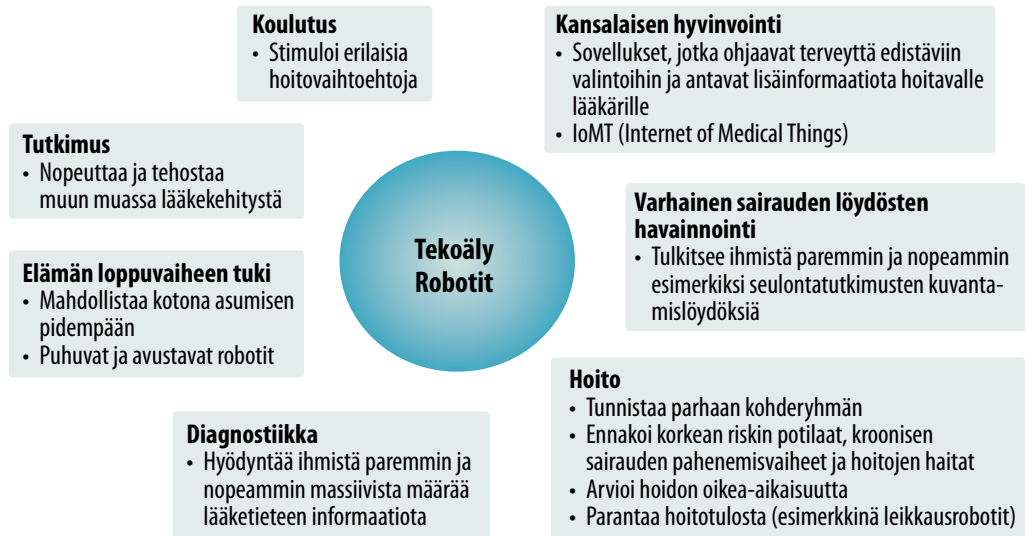
Lääketiede perustuu tutkittuun tietoon. Biologisten ilmiöiden tilastollista merkittävyyttä ja ennustevoimaa on tutkittu jo vuosisatoja. Havaintojen normaalijakauma kuvattiin ensi kertaa vuonna 1808 ja näin jakautuneiden muuttujien keskiarvojen eroja mittaava t-testi vuonna 1908. Yksittäisten biologisten muuttujien tilastollista ennustevoimaa populaatiossa on osattu tutkia jo pitkään.

Todellisissa biologisissa ilmiöissä on kuitenkin valtavasti muuttujia, joiden yhteistä merkitystä on mahdoton tutkia yksi tekijä kerrallaan. Ongelmaan pureutuvaa tiedonlouhintaa kutsutaan toisinaan tilastolliseksi oppimiseksi, koneoppimiseksi tai tekoälyksi. Alan menetelmät perustuvat moderniin tilastotieteeseen ja monimuuttujamenetelmiin, joiden soveltaminen alkoi 1950-luvulla. Erityisesti matriisilaskennan sanotaan pelastaneen tilastotieteen, koska monimuuttujamenetelmien muotoilu ilman tätä matemaattista teoriaa on vaikeaa. Tekoälytutkimuksen historia on lyhydestään huolimatta vaiheikas, ja kokeilut ulottuvat sääntöpohjaisista päätöksenteon tuen järjestelmistä verkkoteoriaa hyödyntäviin hierarkkisiin malleihin (1,2). Esimerkiksi neuroverkkoja käytettiin jo 1980-luvun alussa innokkaasti, vaikka niiden rakenteellinen ennustevoima onnistut-

tiin todistamaan matemaattisesti vasta kyseisen vuosikymmenen lopulla (3). Kesti silti vielä kaksikymmentä vuotta, ennen kuin teknologia löi lopullisesti läpi algoritmien kehittyessä ja laskentatehon kasvaessa.

### Tekoäly lääketieteen palveluksessa

Nykyhetki eroaa kuitenkin monella tavalla menneestä. Ensinnäkin, niin sanotut syvät neuroverkot kykenevät löytämään itsenäisesti datasta ne piirteet, jotka ovat ennusteen kannalta merkittäviä. Tätä kutsutaan usein totunnaisista lääketieteen hypoteeseista vapaaksi tutkimukseksi, jossa selittävistä ilmiöistä ei tehdä etukäteisolettamuksia. Vielä joitakin vuosikymmeniä sitten tutkija joutui suunnittelemaan etukäteen laskentasäännöt niiden piirteiden muodostamiseksi, joilla oletettiin olevan merkitystä ilmiön ennustamisessa. Merkityksellisiksi piirteiksi voitiin arvella esimerkiksi sydämen sykevälin tai veren happisaturaation muutoksia. Laskennan jälkeen nämä etukäteen valitut muuttujat syötettiin tilastolliseen malliin, kuten regressiomalliin, yksinkertaiseen neuroverkkomalliin tai tukivektorialgoritmillemme (support vector machine) piirteiden ennustavuuden arvioimiseksi.



**KUVA.** Tekoälysovellukset ja robotiikka terveydenhuollon päätöksenteossa.

Toiseksi, nykyaikaisten tietokoneiden laskentateho riittää jo käytännössä itseoppivien tilastollisten mallien rakentamiseen. Modernit koneoppimisalgoritmit pystyvät tunnistamaan ihmistä nopeammin ja tarkemmin hahmoja äänestä, kuvasta tai videosta ja tekemään tämän perusteella tarkoituksenmukaisia päätöksiä. Esimerkkeinä tästä ovat itseajavat autot tai musiikkikappaleen tunnistus ääninäytteen perusteella. Lääketieteessä automatisoitua hahmontunnistusta on jo alettu käyttää radiologian tiedostojen ja muiden kuvantamistulosten, kuten digitoitujen patologisten leikkeiden tulokinnassa (4). Systeemibiologian omiikkatutkimuksetkin perustuvat eksperimentaaliseen ja laskentaintensiiviseen tilastotieteeseen. Tutuin esimerkki lienee genomien kattava assosiaatiotutkimus, jossa alttiusgeeneistä ei oleteta etukäteen mitään.

Kolmantena syynä nykyiselle tekoälyn voitokululle on datan saatavuuden parantuminen. Lääketiede on tekoälyn hyödyntämisessä muuta yhteiskuntaa jäljessä. Tähän ovat osittain vaikuttaneet sekä tekniset että lainsäädännölliset seikat. Terveydenhuollon tieto on pirstaleisesti tallennettu eri potilastietojärjestelmiin ja kokonaiskuvaa potilaasta on ollut melkein mahdollista saada. Samoin lainsäädäntö on jäänyt jäl-

keen terveydenhuollon digitalisaatiosta. Eduskunnan 13.3.2019 hyväksymä potilastiedon tietoturvalista hyödyntämistä selventävä lainsäädäntö pyrkii osaltaan korjaamaan tilannetta.

Osa suomalaisia sairaanhoitopiirejä on jo saatu kehitettyä tietoaltaita. Näiden avulla sairaalat voivat hallinnoida tuottamaansa potilastietoja aivan uudella tavalla sekä avata aineistoja tiedon louhintaan (5). Tietoturva tulee ottaa huomioon monella eri tasolla alkaen käyttäjien koulutuksesta ja teknisistä lokikirjoista fyysiseen kulunvalvontaan. Laajojen potilasaineistojen käsittely tietoturvallisesti ja joustavasti mahdollistuu parhaiten sairaalan tarjoaman tietoturvallisesta etäympäristön avulla. Tässä analyysiympäristössä tutkimusryhmän on mahdollista asentaa tai ohjelmoida itse uusia algoritmeja aineiston tutkimiseksi ilman, että datasta muodostetaan fyysistä kopiota sairaalan ulkopuolelle (6). Suurimmat heikkoudet tietoturvaan syntyvät tunnetusti ihmisten ymmärtämättömyydestä tai pahansuovasta käytöksestä. Sairaala on tiedon tuottajana ja rekisterinpitäjänä kuitenkin vastuussa aineistonsa käytöstä ja käyttöluvista. Tästä syystä tarvittavaa tietoteknistä osaamista ei pidä ulkoistaa. Laskentaympäristö kannattaa myös rakentaa avoimeen lähdekoodin komponenteista, jolloin

sen toiminta voidaan aidosti arvioida joko itse tai riippumattomien tietoturva-auditointien avulla. Avoin lähdekoodi varmistaa tieteen avoimuuden, riippumattomuuden ja vertaisarvioinnin, jotka ovat luovuttamattomia arvoja. Mikäli dataa ei saada laajaan käyttöön, yksilöllisen lääketieteen kehitys on vaarassa pysähtyä ja potilaiden eriarvoisuus kasvaa.

Tekoälyllä tulee olemaan monia erilaisia käyttökohteita terveydenhuollossa (**KUVA**). Lisäksi lääkekehitys hyötyy datasta. Uudet lääkkeet eivät ole enää niin sanottuja kassamagneetteja, joiden käyttäjäkunta ja hyötyjät olisivat laajoja potilasjoukkoja. Tulevaisuuden lääkkeet ovat entistä spesifisempiä, ehkä vain yhteen signalointireittiin vaikuttavia, jolloin entistä harvemmat potilaat hyötyvät niistä. Toisaalta oikeassa kohderyhmässä ja oikeaan aikaan käytettynä niiden hyöty voi olla erittäin merkittävä.

## Uudet haasteet tiedon määrässä ja laadussa

Uusi täsmälääkekehitys ja yksilöllistetty hoito ovat tehneet tarpeelliseksi erottaa molekyyligeneettisesti erilaiset sairaudet nykyisten saateenvarjodiagnoosien alta, jotta oikeaan alaryhmään liittyvä täsmäterapia voidaan aloittaa. Tämä vaatii reaaliaikaista vaikuttavuuden seuranta ja ehkä entistä laajempaa potilaan monitorointia sekä uusia laskentamenetelmiä tunnistamaan potilaan yksilöllinen riskiprofiili.

Lähdetieto voi olla luonteeltaan hyvin erilaista (**TAULUKKO**). Sairauskertomukseen kirjatut löydökset ja tehdyt tutkimukset on suunnattu aluksi työdiagnoosin ja myöhemmin todetun hoitodiagnoosin mukaisesti. Kun diagnostinen polku on oikea, löydöksillä on usein hyvä ennustearvo, mutta tehdyt testit ovat voimakkaasti valikoituneet. Sairauskertomusmerkinnät ovat usein jo johtopäätöksiä: kokonaisesta digitaalisesta sydänkäyrästä merkitään esimerkiksi vain joitakin tunnuslukuja, joita käytetään diagnostiikassa. Kroonisissa sairauksissa potilastieto kattaa useita vuosia. Jos dataa voidaan yhdistää eri lähteistä, kuten perusterveyden- ja työterveyshuollosta ja erikoissairaanhoidosta ja Kelasta, tiedon avulla voidaan ehkä tunnistaa

**TAULUKKO.** Esimerkkejä takautuvaan potilastietoon perustuvan tutkimuksen lähdedatasta, päätetapahtumista ja analyysimenetelmien haasteista.

### Lähdedata

Potilaskertomus (diagnoosit, muut löydökset, hoidot, haitat, hoitovaste)

Potilaan lähettämä seurantatieto (kyselytutkimukset, liikunta-, uni, ravintopäiväkirjat)

Valtakunnallinen terveydenhuollon rekisteri (lääkeostot, sairauspäivät, muut etuudet)

Kuvantamistiedosto (tietokonetomografia, kaiku- ja magneettikuvaukset)

Digikuvatiedosto (silmänpohjakuva, digitalisoitu histologinen näyte)

Seurantalaitteen lähettämä biosignaali (vitaalisen toimintojen ja veren glukoosipitoisuuden seuranta, potilaan kotona käyttämien laitteiden lähettämä biosignaali)

Laaja yksittäinen tai toistettu laboratoriotutkimus (omiikatutkimukset)

### Esimerkkejä mallinnettavista päätetapahtumista

Mitkä tekijät vaikuttavat optimaaliseen hoitotulokseen?

Mitkä tekijät selittävät mahdollisia haittoja?

Mitkä tekijät selittävät parhaiten eloonjäämistä?

Mitkä tekijät selittävät parhaiten tutkittavan terveyteen liittyvää elämänlaatua sairaudesta tai toimintakyvyn muutoksesta tai hoidosta aiheutuvia haittoja?

### Vaikeudet

Yksittäiset tapahtumat, kuinka muodostetaan pitkittäistarkasteluun käyttökelpoisia muuttujia?

Kuinka hallitaan epävarmuuksia?

Kuinka hallitaan samanaikaisuuksia ja eriaikaisuuksia?

Voiko puuttuvaa tietoa täydentää (imputoida) kliinisen todennäköisyyden pohjalta (ei tietoa sairastetusta tulehduksesta on yhtä kuin ei sairastettua tulehdusta)?

sairauden luonnollisen kulun ja hoitovasteiden perusteella samankaltaisesti käyttäytyvät potilaat ja paljastaa siten sairauden uusia alaryhmiä (7,8) Vaikka dataa kustakin potilaasta voi olla niukasti ja valikoidusti, suuret potilasmäärät ja pitkät seuranta-ajat tekevät aineiston mielenkiintoiseksi ja informatiiviseksi.

Tekoäly yhdistetään usein tilanteisiin, joissa kustakin potilaasta on suuri määrä tietoa käytettävissä, kuten esimerkiksi kuvantamistutkimuksen algoritmisessa tulkinassa. Tavoitteena voi olla joko vain ihmistyön korvaaminen tai kuvantamislöydöksen ihmissilmää tarkempi tulkinta. Tällainen tekoälyyn perustuva algo-

## Ydinasiat

- ▶ Potilastiedon määrä ja kompleksisuus kasvaa huomasti tulevina vuosina.
- ▶ Jotta sairaalat voivat hyödyntää tietoa oman toimintansa optimoimisessa sekä uuden diagnostiikan ja hoitoalgoritmien kehityksessä, tarvitaan paljon uudenlaista ajattelua ja osaamista.
- ▶ Digitaalisuus helpottaa tietojen keräämistä ja tekoäly tiedon tulkintaa.
- ▶ Matematiikka ja ohjelmistotekniikka ovat jatkossa entistäkin keskeisempiä työvälineitä myös biotieteissä.

ritmi voi esimerkiksi seuloa etukäteen laajan aineiston kuvantamistutkimuksia, jolloin alan asiantuntija voi keskittyä ainoastaan merkityksellisiin havaintoihin. Signaalidataa voidaan käyttää myös ennustamaan tulevia tapahtumia. Esimerkiksi vitaalien elintoimintojen seuranta voi auttaa havaitsemaan suuren riskin potilaiden, kuten keskosten sepsiksen tai aivovammapotilaiden tilan akuutin huononemisen (9,10). Näissä tapauksissa yksilöllistä dataa on paljon, ja oletamme, että tekoälysovellukset ensimmäiseksi käyttöön juuri näillä alueilla.

Datan laatu herättää oikeutetusti huolta. Laatu ei kuitenkaan ole vain hyvää tai huonoa, vaan on aina suhteessa tutkimuskysymykseen. Alan tieteellisissä julkaisuissa, joissa analyysit perustuvat esimerkiksi sähköiseen potilaskertomukseen, datan kulku lähdejärjestelmästä analyysiympäristöön (ja siihen liittyvä tekninen extract, transfer and load eli ETL-prosessi) kuvataan usein vain pintapuolisesti. Jotta pitkittäisseurantatutkimukset saisivat paremman arvostuksen lääketieteellisissä julkaisuissa, lähdedatan laadun kuvaukseen ja validointiin tulisi kiinnittää erityistä huomiota (**TAULUKKO**).

Lääketieteellisen datan laatustandardit eivät ole vielä vakiintuneet, mutta datan kattavuus, virheettömyys, yhdenmukaisuus eri lähteistä, uskottavuus ja ajantasaisuus ovat ainakin tärkeitä tekijöitä. Monet potilasjärjestelmät pyr-

ivät mahdollisimman rakenteiseen tietoon, jota pidetään luotettavampana kuin vapaata kertomustekstiä. Asia ei liene noin yksinkertainen: kertomus voi parhaimmillaan antaa hyvän kokonaiskuvan potilaasta siinä missä rakenteenkin tieto voi olla epävarmaa tai virheellistä. Tässä tärkeintä on uusien työntekijöiden perehdytys hoitosuosituksiin ja yhtenäisiin kirjauskäytäntöihin. Toisaalta erilaisilla tekstintulouhinnan sovelluksilla on mahdollisuus syvälliseen tekstin asiayhteyden ymmärtämiseen. Parhaimmillaan algoritmit voisivat rakentaa kertomustekstin jo sitä kirjattaessa. Mikäli ohjelma osaisi ehdottaa tilanteeseen sopivia rakenteisia kirjauksia arvoineen, ihmisen tehtäväksi jäisi luokitusten tarkastus ja hyväksyntä. Hyvin kirjattu sairauskertomus ohjaa tekijäänsä oikeisiin ratkaisuihin ja on avain datan ja sitä kautta hoidon laadun parantamiseksi.

Seurantatutkimuksen tilastollisiin analyyseihin liittyy aina epävarmuutta, koska niissä ei ole käytettävissä satunnaistamista kuten etenevässä tutkimuksessa. Samoin sairauksien diagnostiset kriteerit ja hoito ovat jatkuvassa muutoksessa. Systemaattista, mutta tuntematonta syytä potilaiden valikoitumiselle on edelleen vaikea tunnistaa. Suuret otoskoot auttavat kuitenkin satunnaisen virheen hallinnassa. Virhelähteiden punninta tulisi esittää vähintään pohdintana tulosten julkaisun yhteydessä, mutta myös laskennallinen arvio tai esimerkiksi uudelleenotantamenetelmiin (kuten bootstrap) perustuva arvio analyysin herkkyydestä antaisi tutkimuksille lisäarvoa. Suomalaisella tutkimuksella olisi mahdollisuus tuoda tässä uusia käytäntöjä kansainväliseen alan kirjallisuuteen ja nostaa katsausartikkelien tasoa merkittävästi.

## Uudet analyttiset mahdollisuudet

Englanninkielisessä kirjallisuudessa termejä AI, data mining, pattern mining, machine learning, deep learning käytetään usein melko villisti lääketieteellisessä tutkimuksessa ja usein niillä tarkoitetaan jonkinlaista uutta, laskentaintensiivistä analyysiä. Ensimmäinen hyöty tiedonlouhinnasta on hahmontunnistus ilman käsin tehtyä piirteiden hakua (feature engineering). Tämä tarkoittaa tutkimushypoteesien tuottamista

suoraan datasta, kuten eritavoin ilmentyvien tai käyttäytyvien potilasryhmien identifioiminen. Siinä tietokone ohjelmoidaan oppimaan tilastollisesti merkitsevät piirteet suoraan tuhansien muuttujien joukosta tai suoraan signaaleista ja kuvista. Tämä ei tarkoita totunnaisen tilastotieteen hylkäämistä. Aineisto on tärkeää jakaa tilastollisen mallin sovitusta ja testaamista varten riippumattomiin osajoukkoihin, jotta muodostettu malli ei ylisovitu. Jos vahva tilastollinen riippuvuus löytyy, on syytä epäillä mekanistista riippuvuutta. Tämä tarkoittaa, että jokin ilmiö aiheuttaa toisen, kuten verisuonten supistuminen yhtäällä voi aiheuttaa hapenpuutteen toisaalla. Tällöin ilmiötä kannattaa siirtyä tarkastelemaan systeemidynamiikan keinoin, kuten esimerkiksi kemiallisten reaktioiden, sähköisten ilmiöiden tai aineen mekaanisen kulkeutumisen, kuten diffuusiomallien kautta (11).

Tiedonlouhintaa voidaan hyödyntää myös silloin, kun tutkittava päätetapahtuma on ennalta sovittu. Päätös voidaan tehdä suoraan tilastollisen analyysin perusteella. Tällöin laajasta aineistosta lasketaan yksilöllinen indeksi lukuisten riskitekijöiden ja suojaavien tekijöiden avulla ja joka on (yksittäisen biomarkkerin tapaan) vahvasti sidoksissa johonkin päätetapahtumaan, vaikka tarkkaa vaikuttavaa mekanismia ei tunneta. Esimerkinä tästä ovat esimerkiksi polygeeniset riskiarviot ja populaatiotason riskimittarit (12,13).

Näiden lisäksi koneoppimista voidaan käyttää päätöksenteon tukena, kunhan kaikki yksittäisetkin muuttujat näytetään päätöstä tukevana graafisena esityksenä. Näin vastuun kantava ihminen voi varmistua halutessaan kaikista niistä tekijöistä, joihin päätös perustuu. Kone-

oppimista voidaan käyttää luokittelun ohella myös tekstin louhintaan ja kirjausten rakenteistamiseen. Kirjoittajat ovatkin testanneet alustavasti semanttisten verkkojen avulla tahtuvaa tekstin luokittelua ja anonymisointia tutkimustarkoituksiin. Koneoppimista voisi käyttää mahdollisesti myös synteettisten tutkimusaineistojen tuottamiseen opetus- ja tuotekehityskäyttöön.

## Lopuksi

Roy Amaran lakina (14) tunnettu lausahdus kiteyttää väitteemme laskennallisten tieteiden aiheuttamasta yhteiskunnallisesta murroksesta: ”Yliarvioimme lyhyen aikavälin ja aliarvioimme pitkän aikavälin vaikutukset.” Jotta potilaiden hoidosta kertyvää tietoa voidaan täysimääräisesti hyödyntää, meidän on rakennettava ja opeteltava hyödyntämään tietoaltaita. Tarvitsemme myös muutosta tiedon laadun, kirjaamisen, keräämisen ja analysoinnin arvotuksessa. Tieto voi olla tallennettu keskitetysti tai hajautetusti, mutta korkeatasoiset valmiudet tiedon käsittelemiseksi ja ymmärtämiseksi ovat kummassakin tapauksessa ehdoton edellytys mielekkäille sovelluksille. Nämä valmiudet ovat tulevana vuosina tärkeä osa kliinistä päätöksentekoa. Tiettyjen tieteenalojen, kuten lääketieteen ja biologian luonne on muuttunut vääjäämättömästi: tutkimusta ei ole mahdollista tehdä ilman merkittävää matematiikan ja ohjelmistotekniikan osaamista ja toiminnan resursointia. Suomen opetushallitus on tämän jo havainnut, ja siksi esimerkiksi ohjelmointi on otettu osaksi pakollista peruskoulun oppimäärää. ■

**TARJA LAITINEN, LT, tutkimusjohtaja**  
Tays

**ARHO VIRKKI, FT, johtaja**  
Auria tietopalvelu

**VASTUUTOIMITTAJA**  
Tuomas Mirtti

## SIDONNAISUUDET

**Tarja Laitinen:** Luento-/asiantuntijapalkkio (Teva Oy)  
**Arho Virkki:** Luentopalkkio (Medafcon)

**KIRJALLISUUTTA**

1. Goodfellow M, Bengio Y, Courville A. Deep learning. MIT Press 2016. [www.deeplearningbook.org](http://www.deeplearningbook.org)
2. Chollet F, Allaire JJ. Deep learning with R. Manning 2018.
3. Hornik K, Stinchcombe, White H. Multi-layer feedforward networks are universal approximators. *Neur Network* 1989;2: 359–66.
4. Coudray N, Ocampo PS, Sakellaropoulos T, ym. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018. DOI: 10.1038/s41591-018-0177-5.
5. Hammis A, Varjonen J, Virkki A. The clinical data refinery: management and administration of the analytics environment. *Auria Clinical Informatics*. Turku University Central Hospital 2018. <https://github.com/auriadih/data-refinery-book/>.
6. Laajala TD, Murtojarvi M, Virkki A, ym. ePCR: an R-package for survival and time-to-event prediction in advanced prostate cancer, applied to real-world patient cohorts. *Bioinformatics* 2018;34:3957–9.
7. Ahlqvist E, Storm P, Käräjämäki A, ym. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6: 361–9.
8. Koskela J, Kupiainen H, Katajisto M, ym. Individual fev1 trajectories can be identified from a COPD cohort. *J Chr Obstruct Pulm Dis* 2016;13:425–30.
9. Mani S, Ozdas A, Aliferis C, ym. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J Am Med Inform Assoc* 2013;21: 326–36.
10. Agoston DV, Langford D. Big Data in traumatic brain injury; promise and challenges. *Concussion* 2017;2:CNC44.
11. Virkki A, Polo O, Gyllenberg M, ym. Can carotid body perfusion act as a respiratory controller? *J Ther Biol* 2007;249:737–48.
12. Natarajan P, Peloso GM, Zekavat SM, ym. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat Commun* 2018. DOI: 10.1038/s41467-018-05747-8.
13. Gormley P, Kurki MJ, Hiekkala ME, ym. Common variant burden contributes to the familial aggregation of migraine in 1,589 families. *Neuron* 2018;99:743–53.
14. Ratcliffe S, toim. *Roy Amara 1925–2007, American futurologist*. Oxford essential quotations. 4. painos. Oxford University Press 2016.

**SUMMARY**

**Real world data – data mining revolutionizes research and care**

Machine learning is a discipline of statistics with a long tradition in many fields of engineering. Computers can be programmed to tirelessly perform precise and rapid decisions, provided that there is an abundance of information-rich data available. Machine learning based research methodologies benefit both the patient and the society, enabling the development of more effective treatments. The principle of scientific openness is a key element in developing modern analysis methods, and the accompanying software should also be free and open-sourced. This enables us to independently ensure that all methods are medically sound, statistically valid and adherent to high information security. The era of data utilization is still at its early stages, but beyond the technology, it is essential to understand the value of high-quality data in providing high-quality patient care.