

**Vesa Kiviniemi**

FL

Arviointipäällikkö, Fimea

**Antti Hyvärinen**

MMT, VTM

Lääketaloustieteilijä, Fimea

# Big data lääkeviranomaisen silmin

Big data on megatrendi, jonka avaamat mahdollisuudet eivät ole jääneet huomaamatta myöskään lääkealalla. Valtavat tietomäärät, monenkirjavat tietolähteet, vaativa analytiikka ja tulosten tulkinta haastavat myös lääkeregulaatiota. Big datan mahdollisuuksia ei kuitenkaan kannata jättää käyttämättä.

**B**ig dataa on kuvattu lukuisilla eri tavoilla. Sitä voidaan luonnehtia esimerkiksi seuraavasti: ”Big data on aineistoa, joka on kooltaan niin suurta ja rakenteeltaan niin monimutkaista, että sitä ei voi lähestyä perinteisin analyysimenetelmin eikä työkaluin.”

Kirjallisuudessa big dataa on usein kuvattu neljän englanninkielisen V-kirjaimen avulla: Massiivinen tietomäärä (volume), tiedon nopea kertyminen ja muuntuminen (velocity), aineiston ja tietolähteiden monimuotoisuus (variety) sekä aineiston epätäsmällisyys ja validiteettiongelmat (veracity).

## Big data tuo useita mahdollisuuksia lääkekehitykseen

Lääketutkimuksessa avautuu useita mahdollisuuksia suurten tietomasojen hyödyntämiselle innovatiivisia data-analytiikan metodeja käyttäen. Esimerkiksi kokogenomisekvensointi mahdollistaa uudentyyppisten, genomitietoa hyödyntävien sovellusten kehittämisen ja lääkehoitojen yksilöllistymisen. Reaaliaikaisen big data -analytiikan avulla voidaan ennustaa, vähentää ja kontrolloida tarttuvien sairauksien leviämistä. Valtavista tietomassoista voidaan tunnistaa hoitopolkujen mahdollisia epäkohtia, ohjata hoitoa tai optimoida

Kuvio 1. Big datan keskeiset tietolähteet lääkealalla ja HMA/EMA Big Data Task Forcen työryhmät.



resurssien käyttöä. Kuvantamistutkimuksista (esim. MRI, fMRI, PET, MEG) kertyvät aineistot mahdollistavat täysin uusia lähestymistapoja esimerkiksi neurodegeneratiivisten sairauksien tutkimuksessa.

Kliinisessä tutkimuksessa on mahdollista tietolähteiden ja analytiikan keinoin identifoida lääkkeen potentiaalinen kohdepopulaatio totuttua täsmällisemmin ja potilasrekrytointeja voidaan tehostaa. Heterogeenisten ja laajojen tietomassojen analysointi uusien keinoin voi mahdollistaa aiemmin tuntemattomien biomarkkereiden, lääkkeen vaikutuskohteiden ja uusien käyttötarkoitusten löytymisen. Lisäksi erilaisten taudin kulkua ennustavien mallien kehitys voi monipuolistua.

### Mistä big data tulee ja kuinka sitä voidaan käyttää?

Lääkettutkimuksessa big datan lähteitä ovat esimerkiksi genomiikkaan-, proteomiikkaan-, metaboliikkaan- ja kuvantamistietoihin liittyvät aineistot. Näiden tietovarantojen yhdistäminen hoitotuloksiin sekä muuhun terveydenhuollossa kertyvään demografiseen ja kliiniseen aineistoon (Real World Data, RWD) sekä erityisesti tähän liittyvän analytiikan sovellukset ovat vielä alkuvaiheessa. Esimerkki tällaisesta tietovarannosta on Ison-Britannian biopankkihanke. Myös suomalaiset biopankit ja terveydenhuollon rekistrit kartuttavat tällaista tietovarantoa tuleviin tutkimustarpeisiin.

## Lääkettutkimuksen ja -regulaation näkökulmasta sosiaalisen median tietovarantoihin ja spontaanien haittavaikutusten louhintaan liittyy käyttämätöntä potentiaalia.

Kliinisten kokeiden yhteydessä kertyneet tutkimusaineistot ovat myös yksi keskeinen big datan lähde. Esimerkiksi ClinicalStudyDataRequest-tietokanta sisältää 13 lääketyhtiön tekemän yli 3 000 kliinisen kokeen potilaskohtaisen aineiston. Pro-Act-tietokantaan on puolestaan kerätty yli 11 000 ALS-potilaan yksilöllinen aineisto 23 kliinisestä kokeesta. Yksittäisessä kliinisessä kokeessakin voi syntyä valtavia tietomassoja, esimerkiksi tilanteissa joissa koehenkilöiltä kertyy sensorien keräämää signaalidataa. Tällaisen datan hyötykäyttö edellyttää big data -analytiikan soveltamista.

Lääkettutkimuksen ja -regulaation näkökulmasta sosiaalisen median tietovarantoihin ja spontaanien haittavaikutusten louhintaan liittyy käyttämätöntä potentiaalia. Näihin tietovarantoihin kohdistettu analytiikka voi edesauttaa reaaliaikaisempien turvallisuussignaalien löytymisessä sekä potilasnäkökulman paremmassa ymmärtämisessä.

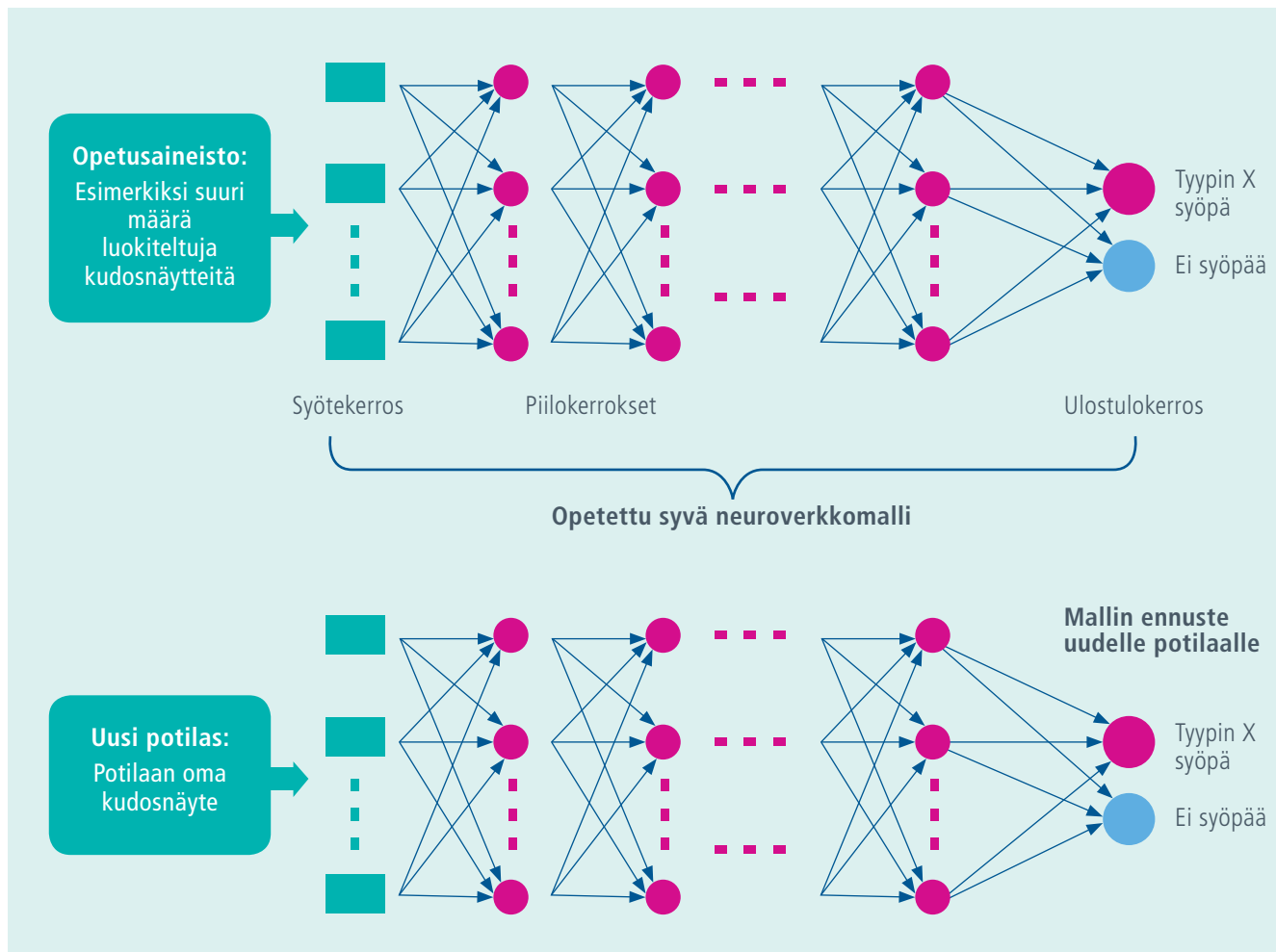
Big datan keskeiset tietolähteet lääketekesektorilla on esitetty **kuviossa 1**.

### Tekoäly oppii tulkitsemaan aineistoa esimerkitapausten avulla

Big datan analytiikan yhteydessä puhutaan usein tiedon louhinnasta, koneoppimisesta ja tekoälystä. Koneoppimisella tarkoitetaan sellaista ohjelmistoa, joka algoritmien avulla oppii esimerkitapauksista samoin kuin ihmiset oppivat kokemusten kautta. Tällaisen ohjelmiston voidaan odottaa kehittyvän sitä paremmaksi, mitä enemmän esimerkitapausta sille annetaan oppimista varten. Erityisen menestyksekkäiksi ovat 2010-luvulla osoittautuneet niin sanotun syväoppimisen (deep learning) sovellukset. Tämän päivän tekoälysovelluksilla tarkoitetaan tyypillisesti keinotekoisii neuroverkkoihin pohjautuvia sovelluksia, jotka sisältävät useita kerroksia.

Keinotekoisien neuroverkkosovellusten tavoitteena on, että saadut tulokset olisivat yleistettävissä uusien, opetusaineiston ulkopuolisten, vastaavien tapausten ennustamiseen. Jos kun ohjelmisto on opetettu esimerkitapausten avulla luokittele-

Kuvio 2. Yksinkertaistettu havainnekuva kerroksittaisesta neuroverkosta syöpädiagnostiikassa.



maan kuvia tai tunnistamaan niistä jonkin ominaisuuden, se kykenee tekemään saman myös uusille kuville, mitä se ei ole aiemmin nähnyt. Kliinisissä syväoppimisen sovelluksissa esimerkiksi kudospäytteet voidaan luokitella sen mukaan, näkyykö niissä tietyn tyyppinen syöpäkasvain vai ei (kuvio 2).

Erona syväoppimismallin ja perinteisen mallinnuksen välillä on, että käyttäjän ei tarvitse tarkasti määrittellä kuvista etsittäviä ominaisuuksia ja näihin liittyviä luokittelusääntöjä. Tämä annetaan koneen tehtäväksi,

jolloin syväoppiva sovellus oppii itse tunnistamaan luokittelun kannalta parhaat ominaisuudet opetusaineistosta. Mallia muodostaessaan algoritmi käy läpi lukemattoman määrän eri vaihtoehtoja pyrkien koko ajan parantamaan luokittelun tarkkuutta. Lopputuloksena syntyy sovellus, jonka tuottamaa informaatiota voidaan käyttää tehokkaasti lääkärin tekemän diagnostiikan tukena. Terveysteknologian alalla syväoppimista on sovellettu ensimmäisenä juuri kuvantamistutkimusten analyysissä ja genomikassa.

### Big data haastaa lääke-regulaatiota

Big data on eräänlainen megatrendi, joka on saanut myös viranomaiset miettimään sen mahdollisuuksia lääkekehityksessä ja siihen liittyvässä viranomaistoiminnassa.

Vuonna 2017 aloitettiin Euroopan lääkevirastojohtajien (HMA) ja Euroopan lääkeviraston (EMA) yhteinen Big Data -työryhmä. Ryhmän tehtäväksi asetettiin lääkekehityksen ja regulaation kannalta merkityksellisten big datan lähteiden

## *Big data -analytiikassa satunnaisvaihtelun tuottamia näennäisiä signaaleja erehdytään toisinaan pitämään todellisina kausaalivaikutuksina.*

kartoitus, aineistojen käytettävyyden ja sovellettavuuden arviointi sekä nykytilan ja tulevaisuuden tarpeiden arviointi. Tämän ohella työryhmän tulee esittää big datan hyötykäyttöön tähtäävä suosituksensa. Ryhmä on jakautunut big datan lähteitä karakterisoivan kuvan 1 mukaisiin alatyöryhmiin. Ryhmä jatkaa toimintaansa vuoden 2018 viimeiselle neljännekselle asti.

Eurooppalaisen lääkeregulaation ohella FDA on tarttumassa big datan mahdollisuuksiin. FDA:n suunnitelmat liittyvät erityisesti terveydenhuollon arjesta kertyvän tutkimustiedon (RWD) hyödyntämiseen regulatorisessa päätöksenteossa. FDA:ta ohjaava 21st Century Cures Act edellyttää jatkossa mahdollistamaan menettelyjä ja prosesseja, joissa indikaation laajennukset tai myyntiluvan myöntämisen jälkeiset tutkimusvelvoitteet voitaisiin toteuttaa soveltuvin osin RWD:tä hyödyntäen.

### **Big datan hyödyntämisen karikat**

Aineiston saatavuus, sovellettavuus ja kaupalliset intressit ohjaavat paljon sitä, mitä big datalla tehdään. Nämä eivät aina parhaalla mahdollisella tavalla kohtaa niitä tarpeita, joita terveydenhuollolla, lääkekehityksellä tai lääkeregulaatiolla on. Joissakin sovelluksissa ongelmia syntyy myös siksi, että kysymyksen haetaan ratkaisua siihen soveltumattomalla aineistopohjalla. Nykyisistä koneoppimisen sovelluksista suuri osa pohjautuu niin sanottuun ohjattuun oppimiseen, jolloin mallin

opettaminen vaatii runsaasti valmiiksi luokiteltua aineistoa, jota ei välttämättä ole saatavilla.

Toinen keskeinen haaste big data -analytiikassa on tulosten ja johtopäätösten validiteetti, toistettavuus ja yleistettävyyden. Koneet oppivat erinomaisen hyvin sen aineiston, jota ne käyttävät. Ilman ihmisen asettamia tai yleisesti hyväksytyihin teorioihin perustuvia rajoitteita, koneilla on taipumus ylioppia aineistosta, jolloin tulosten yleistämisen opetusaineiston ulkopuolelle tai niiden toistaminen muussa aineistossa kärsii.

Edellisten ohella kehittyneimmät sovellukset tarjoavat käyttäjälle usein lopputuloksen, mutta ei sen täsmällistä tulkintaa. Näin ollen malli on eräänlainen ”musta laatikko”, jonka tuottamien lopputulosten taustalla olevia tekijöitä ei voida välttämättä täsmällisesti jäljittää. Big data -analytiikassa myös satunnaisvaihtelun tuottamia näennäisiä signaaleja erehdytään toisinaan pitämään todellisina kausaalivaikutuksina.

### **Mitä tästä seuraa?**

Big data antaa lupauksia uudenlaisesta päättelystä ja tiedon kehittymisestä. Lääkealalla big data ei ainakaan lähitulevaisuudessa tule silti korvaamaan satunnaistettuja ja kontrolloituja kliinisiä tutkimuksia, mutta kehittyvä analytiikka voi täydentää myös kliinisten tutkimusten tulosten analysointia. Lupaavimmat sovellusalueet löytyvät tällä hetkellä genomiikan, kuvantamisen ja RWD:n alueilta.

Big datan ja koneoppimisen tuomia mahdollisuuksia terveydenhuollossa ei kannata jättää käyttämättä. Big datan ja sen sovellusten myötä lääkekehitykseen syntyy valtavasti mahdollisuuksia, mutta myös runsaasti epävarmuutta. Tämän päivän tekoäly on monessa mielessä varsin rajoittuneesti älykäästä. Siksi se tulee useimmissa tilanteissa nähdä apuvälineenä, ei niinkään varsinaisena ongelman ratkaisijana. ●

### **KIRJALLISUUTTA**

- Big data technologies in health-care – needs, opportunities and challenges - [www.bdva.eu](http://www.bdva.eu)
- Chen H, ym. The rise of deep learning in drug discovery. *Drug Discovery Today* 2018. Painossa.
- ClinicalStudyDataRequest. [www.clinicalstudydatarequest.com](http://www.clinicalstudydatarequest.com)
- Identifying opportunities for big data in medicines development and regulatory science (EMA): [www.ema.europa.eu](http://www.ema.europa.eu)
- LeCun Y, ym. Deep learning. *Nature* 2015; 521: 436–44.
- UK biobank. [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)
- 21st Century Cures Act. <https://www.congress.gov>

