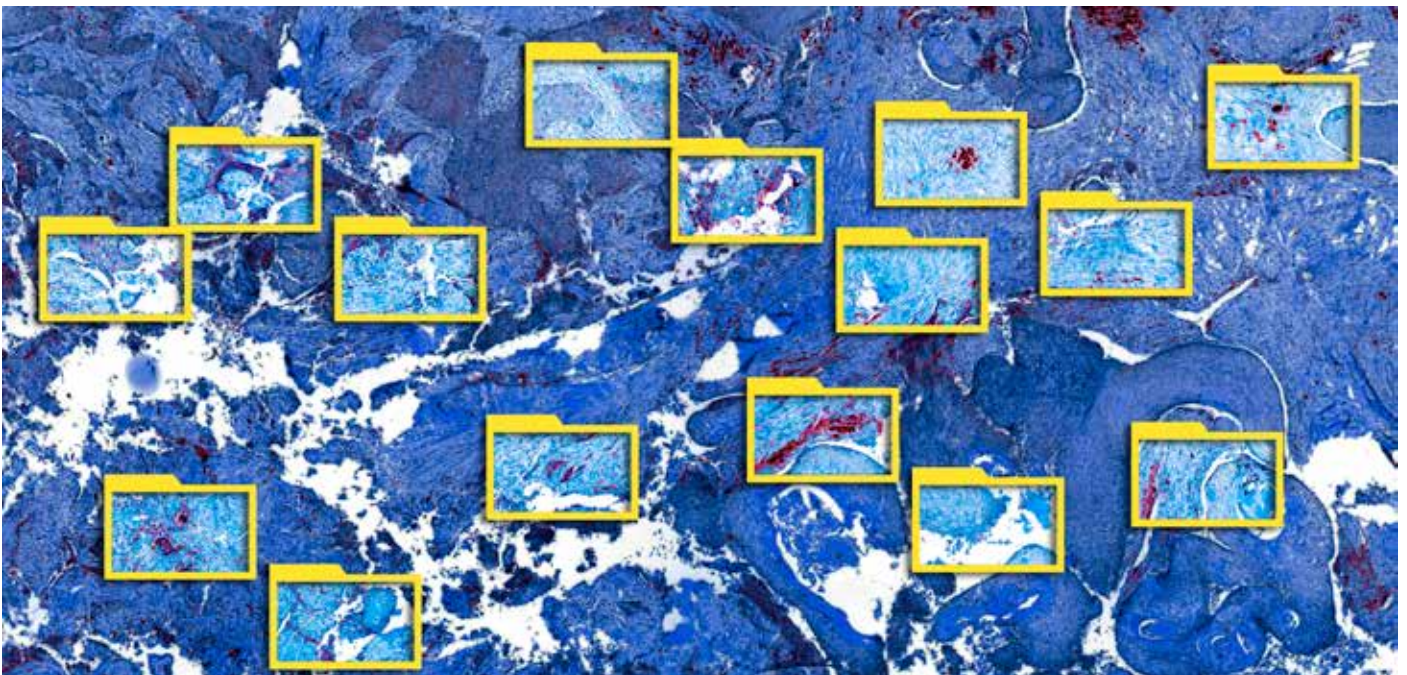


Kudosnäytteet digitaalisiksi kuviksi ja tekoäly tulkitsemaan

Turun yliopistollisen keskussairaalan ja Auria Biopankin tavoitteena on saada kaikki kudosleikenäytteet digitaaliseen muotoon. Näytteet skannattaisiin lasilevyiltä ja jokapäiväisessä patologian työssä diagnostiikka siirtyisi tietokoneelle. Samalla kehitetään tekoälymalleja eli luokittelijoita, jotka tunnistavat digitoituista näytteistä esimerkiksi syövän.



Pelkästään Turun yliopistollisessa keskussairaalassa (TYKS) otetaan potilailta 200 000 kudosnäytettä vuosittain. Kudosnäyte käsitellään formaliinissa ja valetaan parafiiniblokkiin, jonka jälkeen siitä voidaan höylätä leikkeitä mikroskoopilla tarkasteltavaksi. Lopulta parafiiniblokki varastoidaan. Näytteiden hallinnoiminen on työlästä ja vie paljon aikaa. Näytteiden järjestelmällinen digitoiminen tuo tähän apua.

”Koska näytteitä on paljon, metadatatiedon avulla löydetään halutut ja oikeat näytteet nopeasti”, sanoo Auria Biopankin data-analyytikko **Antti Karlsson**.

Tietokantaan voi siis tehdä esimerkiksi haun, joka etsii kaikki ne näytteet, joissa on rintasyöpäkasvainta. Hakua voi metatiedon avulla tarkentaa, jolloin voidaan etsiä esimerkiksi 60-vuotiaiden rintasyöpäpotilaiden näytteet, joissa on tietty reseptoristatus.

Digitaalisen patologian hankkeessa mikroskooppilaseilla olevat näytteet skannataan. Sitten patologit voivat katsoa tietokoneella näytteet ja kuvailla ja luokitella ne. Nämä ns. annotaatiotiedot ovat olennaisia, jotta voidaan opettaa tekoälyä automaattisesti tunnistamaan näytteistä esimerkiksi syöpäsolut. Tämä nopeuttaisi huomattavasti patologin työtä jatkossa. Auria Biopankissa on panostettu data-analytiikkaan, algoritmien kehitykseen ja koneoppimismalleihin.

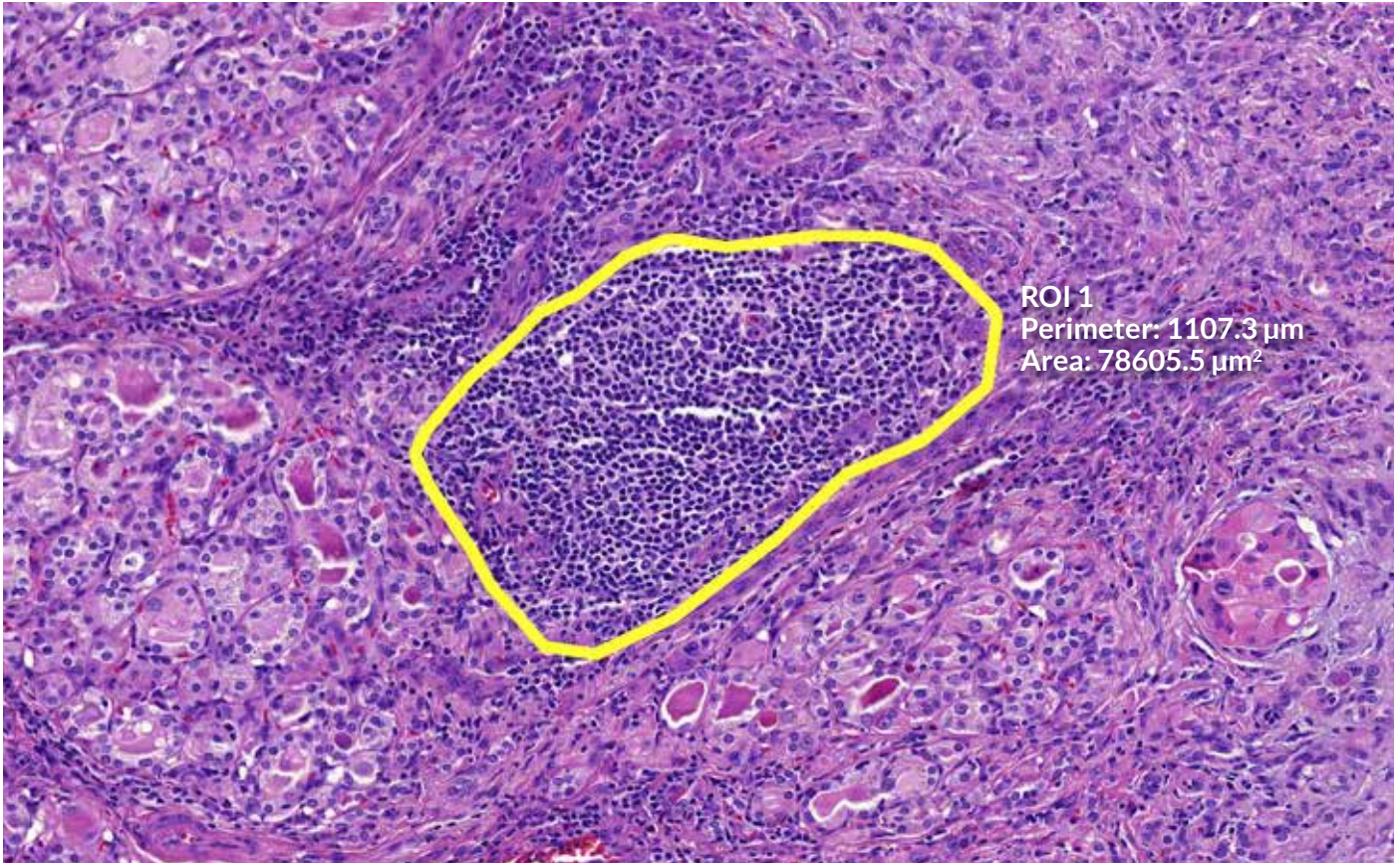
Kielimalli avuksi metadatan kuvailuun

Turun yliopistollisessa keskussairaalassa on valtavasti mikroskooppilaseille säilöttyjä kudosleikkeitä. Ongelma on, että lasiin ei saa tallennettua metadatat, joka saataisiin siirrettyä tietokantoihin automaattisesti. Nyt tarkoituksena on, että uusiin näytteisiin

patologit merkitsevät metadatan kuvankäsittelyohjelman avulla.

Karlssonin mukaan työ on ensin mekaanista. Patologi käyttää kuvankäsittelyohjelmaa, jonka avulla piirtää skannattuihin näytteisiin ne kohdat, joissa on esimerkiksi syöpää.

Tämän lisäksi tarvitaan kuvailutietoja. Tässä apuna olisivat neuroverkko-kielimallit. Patologi voisi kuvailla suoraan tietokoneelle näytteen tietoja. Aihetta on tutkittu Turun yliopiston tulevaisuuden teknologioiden laitoksen **Filip Ginterin** tutkimusryhmän kanssa. Tutkimusryhmä on keskittynyt siihen, miten tietokoneohjelmia voidaan käyttää luonnollisen tekstin ja puheen analysointiin. Kielimalli oppii isosta määrästä luokittelu- ja puhe-tekstiä, miten jokin puhuttu kieli näyttää tilastollisesti toimivan. Auria Biopankki ja TYKS ovat kiinnostuneita siitä, miten lääkärintulosten teksteistä saataisiin



Yleisin värjäys kudosten perusrakenteiden selvittämisessä on HE eli hematoksyliini-eosiini -värjäys, jolla voidaan värjätä erilaisia rakenteita kudoksista pH:n mukaisesti. Emäksinen hematoksyliini värjää solun happamat tumat violeteiksi ja hapen eosini värjää solun emäksiset tukirakenteet, kuten side- ja lihaskudoksen j punaisiksi. Kuvassa on HE-värjättyä kudosta, johon on merkitty mahdollisesti kiinnostava rakenne. Patologi piirtää kuvaan alueen, jonka nimeää haluamallaan tavalla. Tällaisia esimerkkejä tarpeeksi kartuttamalla voidaan kouluttaa tekoälymalleja, jotka tekevät vastaavia kuvailuja ja luokitteluita automaattisesti.

muodostettua luokiteltua ja rakenteistettua tietoa kielimallien avulla

”Digipatologiassa yksi sovellus voisi olla se, että jälkikäteen louhitaan lausuntoteksteistä erilaisia tietoja, kuten vaikka missä näytteen osassa on mitään kiinnostavaa kudosta, jolloin näytteiden valinta tutkimusten tarpeisiin helpottuu. Lisäksi voitaisiin kehittää vapaata lausuntotekstiä automaattisesti rakenteistavaa mallia. Patologi voisi lausua ’proosaa’, jonka tekoäly sitten keräisi ja koostaisi rakenteiseksi taulukoksi.”

Karlssonin mukaan tällaisia taulukoita käytetään jo nyt aika paljon esimerkiksi silloin, kun patologit ovat sopineet, mitkä kaikki asiat kustakin kasvaimesta pitää raportoida.

”Tällä hetkellä kokeilemme jo näitä malleja esimerkiksi tupakointitiedon löytämiseen ja luokitteluun satojentuhansien lausuntotekstien sisältä, sekä syövän metastasointitietojen, sairaalainfektioihin liit-

tyvien oireiden ja erilaisten diagnoosien löytämiseen.”

Haasteena on vielä monimuotoinen data. Esimerkiksi eri laitevalmistajien skannerit tuottavat erilaista dataa, joka pitäisi luotettavasti yhteensovittaa.

Tekoälymalli tunnistaa syövän automaattisesti näytteestä

Metadatan ja digitoidun näytemateriaalin avulla kehitetään esimerkiksi tekoälysovelluksia, jotka opetetaan luokittelemaan automaattisesti, missä kohtaa kuvassa on syöpäsoluja. Tekoälyn opettamiseen tarvitaan patologien luokittellemaa materiaalia. Antti Karlssonin mukaan kuvia ei tarvita itse asiassa kovinkaan paljon, jotta algoritmi oppisi.

”Kymmenillä kuvilla päästään jo alkuun. Yksi iso leikekuva voi tuottaa tuhat pientä kuvaa, joilla voi kouluttaa malleja.”

Tällöin 20 potilaasta saadaan jopa 10 000 pientä kuvaa.

”Isoa kuvaa ei sellaisenaan pysty vielä lykkäämään algoritmeille, koska minkään tietokoneen grafiikkaprosessorin muisti ei riitä siihen.”

Karlsson haluaa tähdentää sitä, että kuvia katsovat tekoälymallit ovat eri asia kuin tekstiä katsovat mallit.

”Ne ovat toki kaikki tekoälyä ja vieläpä neuroverkkoja, mutta rakenteeltaan ja toimintaperiaatteeltaan erilaisia. Tekoäly on ennemminkin kokoelma työkaluja, joista jokainen on sitten käyttökelpoinen omaan tiettyyn sovellukseensa.”

Data-analytiikka yksi biopankin painopisteistä

Auria Biopankin johtaja **Lila Kallio** toteaa, että genomidatan tutkimuskäytön lisäksi digipatologiaa hyödyntävä data-analytiikka on yksi keskeisiä Aurian painopisteitä.



Patologiselle datalle suunnitellaan digitaalista tallennuspaikkaa Eurooppaan.

”Entistä enemmän ollaan kiinnostuneita siitä, miten digitoidusta syöpäkudosleikkeestä voidaan tunnistaa eri asioita. Olemme mukana tutkimuksissa, joissa pyritään algoritmin avulla ennustamaan primäärisyöpäkasvaimen näytteen kuvasta esimerkiksi taudin hoitovastetta tai sitä, tuleeko primäärisyöpäkasvain levittämään etäpesäkkeitä. On viitteitä siitä, että algoritmi pystyisi ennustamaan histologisesta kuvasta sellaista, mikä ei silmämääräisesti ole nähtävissä.”

Yhden luukun palvelu

Suomessa on Lila Kallion mielestä oltu datan hallinnoimisessa ja jakamisessa hyvin edistyksellisiä. Suomen biopankkilaki on

mahdollistanut tutkimuksen ja tiedon yhdistelemisen eri rekistereistä. Erityisen tärkeää on, että kliininen tieto voidaan yhdistää näytteisiin.

”Palvelua tutkijoille on voitu toteuttaa yhden luukun periaatteella. Biopankki hoitaa luvat, kerää näytteet ja yhdistää niihin tutkimukselle oleellisen kliinisen tiedon. Tämä kaikki voidaan sitten yhdistää muuhun dataan, esimerkiksi geenitietoihin.”

Biopankin kautta tutkija saa tarvitsemansa näytteet.

”Biopankit tekevät Suomessa yhteistyötä. Tutkija voi pyytää näytteitä kaikista Suomen biopankeista Suomen biopankin osuuskunnan kautta yhdellä pyynnöllä.”

Haasteena nyt ja tulevaisuudessa on Lila Kallion mielestä datan tallentaminen ja hallinnoiminen.

”Dataa tallennetaan sairaanhoitopiirin palomuurien sisälle. Jos patologian diagnostisia näytteitä ruvetaan rutiinomaisesti digitoimaan, tulee myös tallennuskapasiteetti ratkaista. Lisäksi kuvien koko on niin valtava, etteivät ne helposti siirry tavallisten tietoverkkojen kautta.

Laskentateho ja tietoturvalliset tallennus- ja käyttöympäristöt Suomen ELIXIR-keskuksen CSC:n kanssa tulevat tässä tärkeään rooliin.

Ari Turunen

LISÄTIETOJA:

Auria Biopankki
<https://www.auria.fi>

CSC – Tieteen tietotekniikan keskus Oy
on valtion omistama, opetus- ja kulttuuriministeriön hallinnoima, voittoa tavoittelematon osakeyhtiö. CSC ylläpitää ja kehittää valtion omistamaa keskitettyä tietotekniikkainfrastruktuuria.
<http://www.csc.fi>
<https://research.csc.fi/cloud-computing>

ELIXIR
rakentaa infrastruktuurin bioalan tutkimuksen tueksi. Se yhdistää 21 Euroopan maan ja Euroopan molekyylibiologian laboratorion EMBL:n johtavat organisaatiot yhteiseksi biologisen informaation infrastruktuuriksi. Sen Suomen keskus on CSC – Tieteen tietotekniikan keskus Oy.
<http://www.elixir-finland.org>
<http://www.elixir-europe.org>

SUOMEN ELIXIR

Puh. +358 9 457 2821 • e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR PÄÄMAJA

EMBL-European Bioinformatics Institute
www.elixir-europe.org