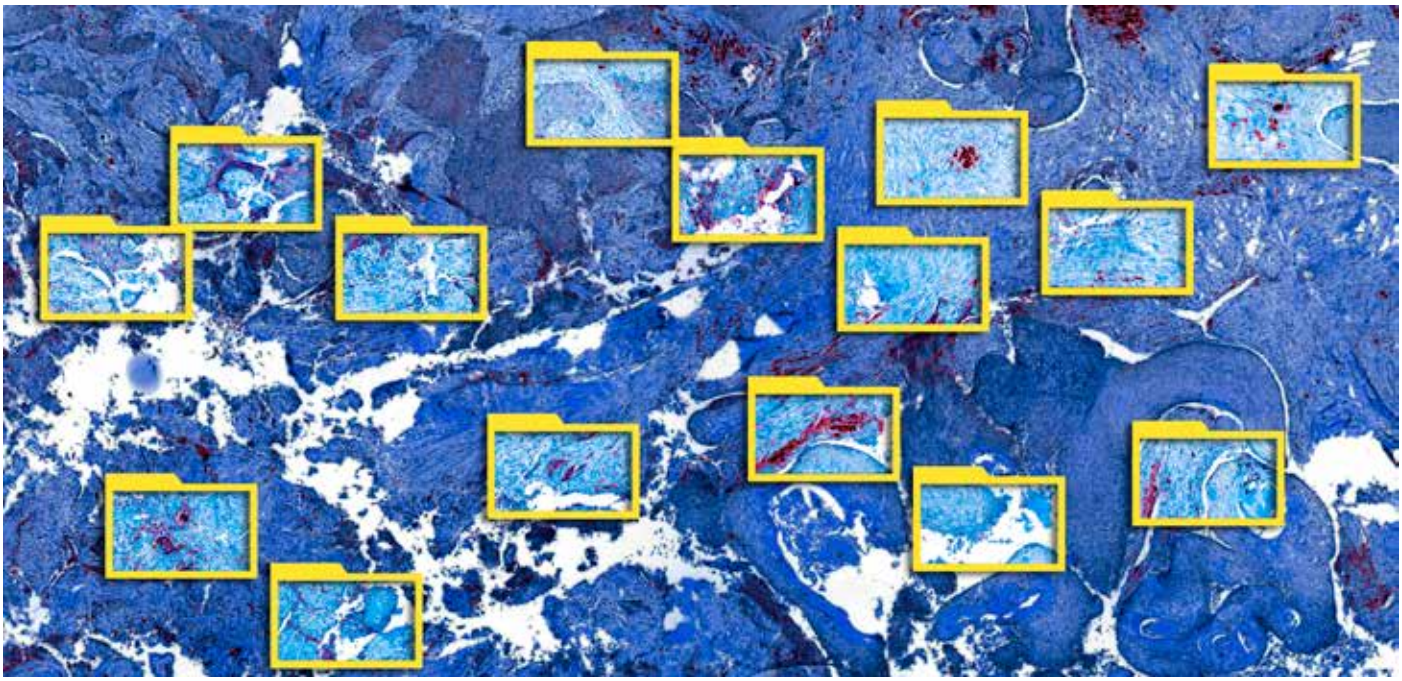
Tissue samples into digital images, interpreted by artificial intelligence

Turku University Hospital and Auria Biobank aim to have all tissue specimens in digital format. The samples would be scanned from glass slides, with diagnostics in pathology performed on computers. They will also develop artificial intelligence models, or classifiers, to identify e.g. cancer from digitalised samples.



Turku University Hospital alone takes 200,000 patient samples every year. Tissue samples are placed in formalin and cast into a paraffin block that can be sliced and subsequently examined under a microscope. The paraffin blocks are stored at the end of the process. Managing the samples is laborious and time-consuming. Systematic digitalisation of the samples makes the job easier.

”Since the samples are so numerous, metadata will enable us to quickly find the samples that we want,” says **Antti Karlsson**, data analyst at Auria Biobank.

You can, for example, search the database for all samples indicating breast cancer tumours. By using metadata, searches can be narrowed down to pinpoint, say, samples with a certain receptor status from 60-year-old breast cancer patients.

In the digital pathology project, samples on microscope slides are scanned. Then a pathologist can view the samples on a computer screen and describe and classify them. All this annotation data is relevant to teaching artificial intelligence to automatically detect abnormalities such as cancer cells in the samples. This would considerably speed up the work of pathologists. Auria Biobank has invested in data analytics, development of algorithms, and machine learning models.

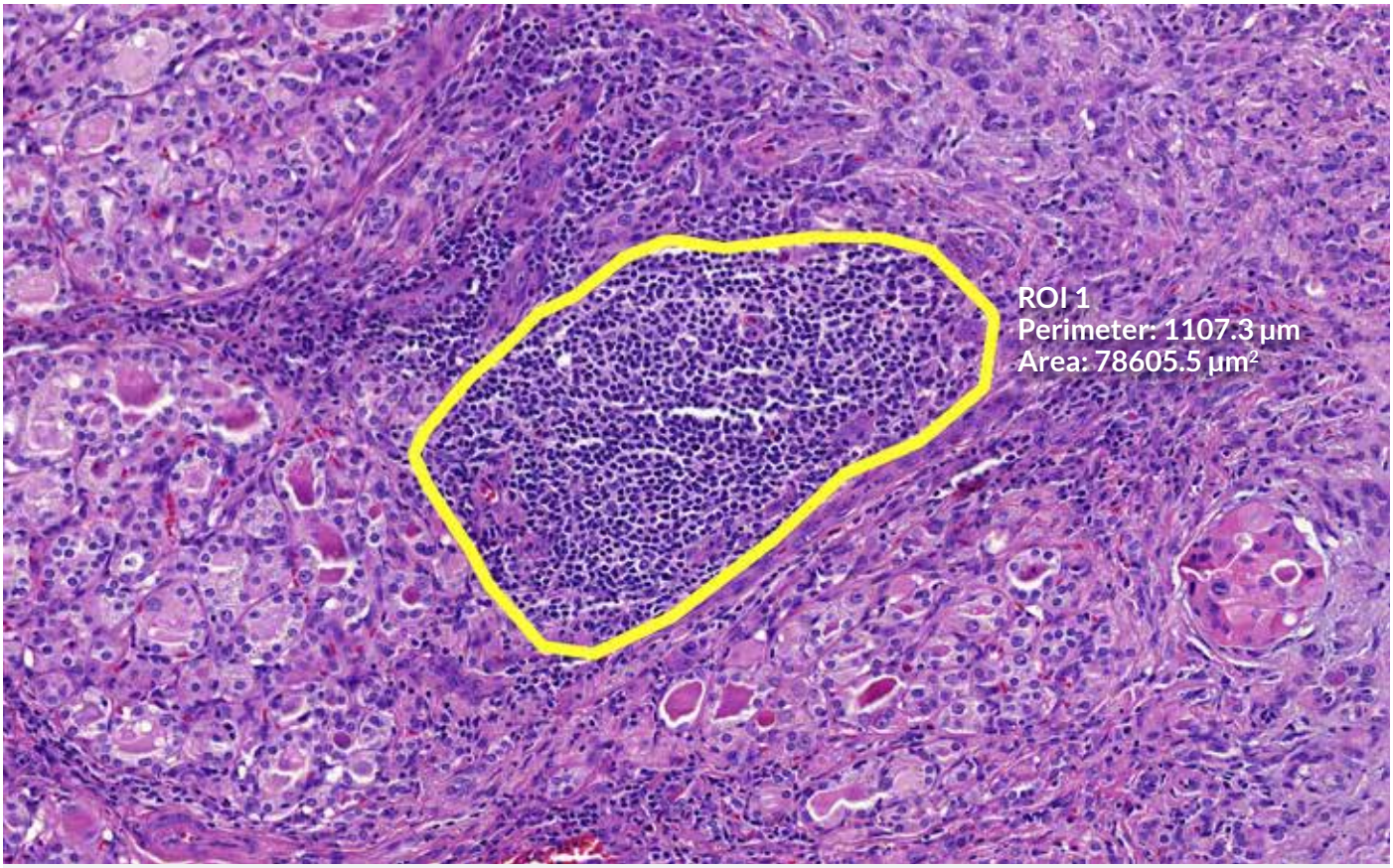
Language model in support of metadata description

Turku University Hospital has huge numbers of tissue specimens stored on microscope slides. The problem is that no metadata can be stored on the slides and transferred into databases automatically. The idea now is that pathologists add metadata to new samples by means of graphics software.

Karlsson says that the work is mechanical to begin with. Pathologists use graphics software to indicate the points where e.g. cancer is found on the scanned samples.

Description data is also required. This is where neural-network language models might come in. A pathologist would add information about the sample directly into a computer. This has been studied in cooperation with **Filip Ginter’s** research team at the Department of Future Technologies of the University of Turku. The research team has focused on how computer software can be used to analyse natural text and speech. From a large amount of unclassified text, the language model learns how a spoken language seems to work statistically. Auria Biobank and Turku University Hospital are interested in how medical statement texts could be formed into classified and structural information by means of language models.

”One application of digital pathology could be to mine various types of informa-



The most common staining method when determining basic structures of tissues is hematoxin-eosin staining that can be used to stain various structures in tissues on the basis of their pH. Alkaline hematoxyline stains the acidic nuclei of cells violet, while acidic eosin stains the alkaline support structures of the cell — such as the connective and muscle tissue — red. The image shows HE-stained tissue, with the potentially interesting structure indicated. The pathologist draws an area in the image and names it appropriately. By creating enough examples like this, we can train artificial intelligence models to create similar descriptions and classifications automatically.

tion from statements, such as which part of the sample contains interesting tissue, making sample selection for research purposes easier. We could also develop a model that automatically structures regular medical statement texts. Pathologists could speak in 'prose', which artificial intelligence would collect and compile into a structured table."

According to Karlsson, such tables are already used relatively frequently, for example when pathologists have agreed on what aspects of each tumour should be reported.

"At the moment we are experimenting with these models, for example to detect and classify smoking data from among hundreds of thousands of statements, and to detect cancer metastases, symptoms related to hospital infections and various diagnoses."

The challenge is that data comes in a variety of forms. For example, scanners by different equipment manufacturers produce different kinds of data that should be presented in a systematic way.

Artificial intelligence model identifies cancer automatically from samples

Metadata and digitalised sample material are used to develop artificial intelligence applications, for example, which are taught to automatically classify the locations with cancer cells in images. To teach the artificial intelligence system, we require some material classified by pathologists. According to Antti Karlsson, you do not in fact need very many images for the algorithm to start learning.

"A few dozen images is enough to get started. A single, whole slice image may yield a thousand small images that can be used to train the models."

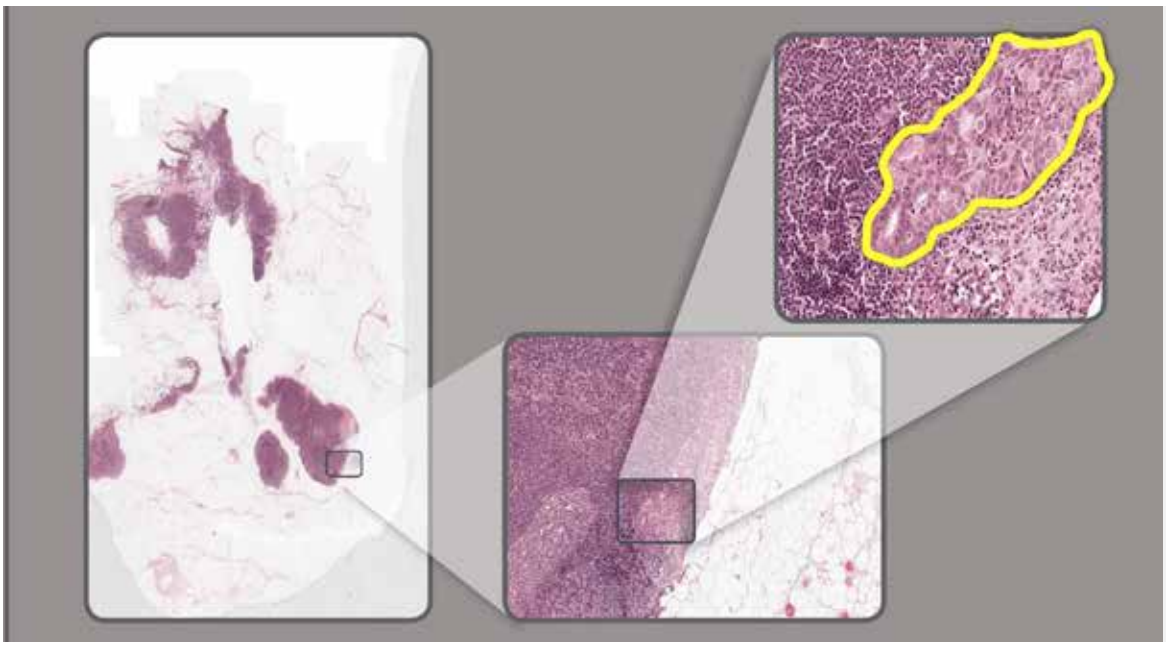
This means that up to 10,000 small images can be obtained from 20 patients.

"A large image cannot be used with an algorithm, because no computer has the kind of graphics processor memory required to deal with it."

Karlsson stresses that artificial intelligence models which examine images are different to models examining texts.

"Clearly, they are all manifestations of artificial intelligence, and even of neural networks, but their structures and operating principles are quite different. Artificial intelligence is actually more like a collection of tools, each one of which is useful for a specific application."

Auria Biobank's Director Lila Kallio says that, in addition to genome data being used for research purposes, digital pathology making use of data analytics is a focus area at Auria.



European storage place for digital pathology data is being planned.

”There is growing interest in how digitalised cancer tissue samples can be used to identify various issues. We are involved in studies where we try to use an algorithm to examine an image of a primary cancer tumour and predict how it will respond to treatment, or whether the primary cancer tumour will metastasise. There are indications that the algorithm may be able to predict something that is not otherwise visible from a histological image.”

One-stop service

According to **Lila Kallio**, Finland has been a pioneering country in data management and sharing. The Finnish Biobank Act has enabled research and the combination of data from various registers. It is critically important that clinical information can be connected to samples.

”Services for researchers have been provided on a one-stop basis. Biobanks take care of the permits, collect the samples and combine all clinical data related to the research. All this can then be combined with other data, such as genetic data.”

Researchers get all the samples through a biobank.

”Biobanks cooperate in Finland. Researchers can request samples from all Finnish biobanks with a single request made to the Finnish Biobank Cooperative.

According to Lila Kallio, the challenge now and in the future is data storage and management.

”Data is stored inside firewalls in hospital districts. If diagnostic samples will be digitalised on a larger scale within pathology, the storage capacity problem must be solved as well. In addition, the image sizes are so huge that they cannot be transferred on ordinary data networks.

The Finnish ELIXIR Center CSC plays an important role in terms of computing power, and safe storage and usage environments.

Ari Turunen

MORE INFORMATION:

Auria Biobank
https://www.auria.fi

CSC – IT Center for Science
is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.
http://www.csc.fi
https://research.csc.fi/cloud-computing

ELIXIR
builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.
http://www.elixir-finland.org
http://www.elixir-europe.org