Deep learning algorithms help in breast cancer screening

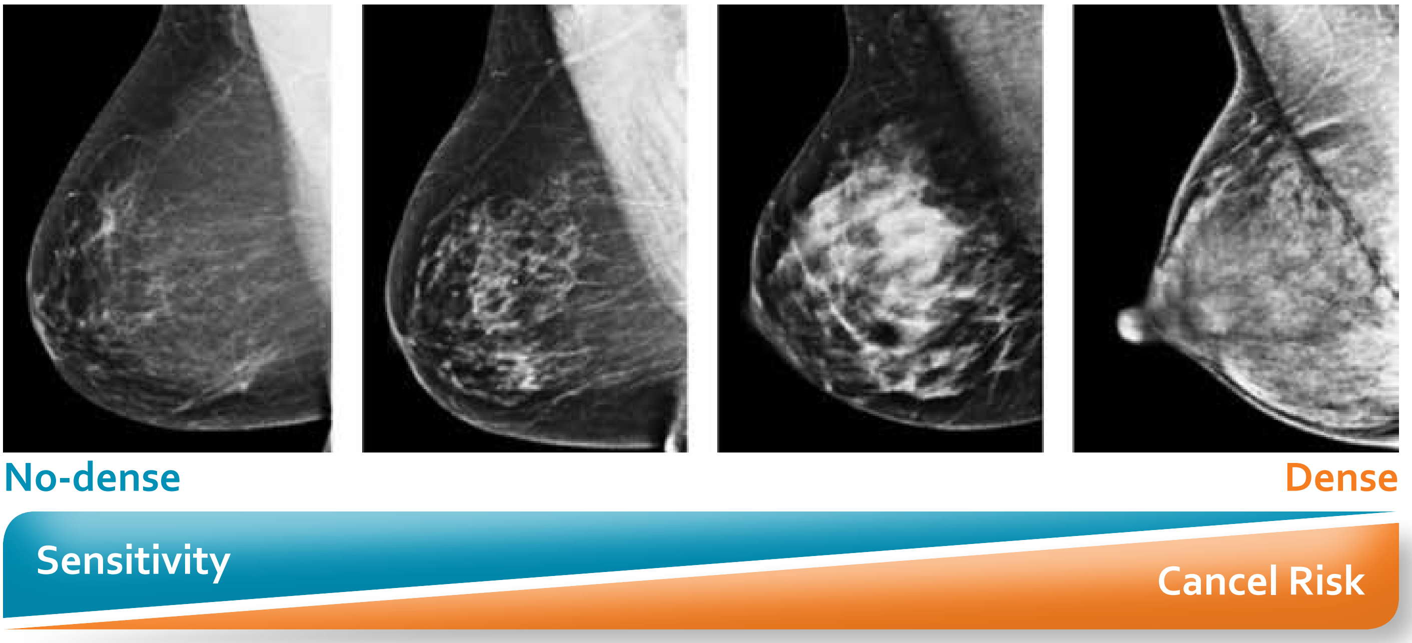
Researcher Raju Gudhe has studied computer science with a focus on intelligent systems. He is now developing deep learning algorithms for breast cancer risk analysis using radiology and clinical data. These algorithms have been trained using massive data sets from Kuopio University Hospital to predict the density of breasts on mammograms.

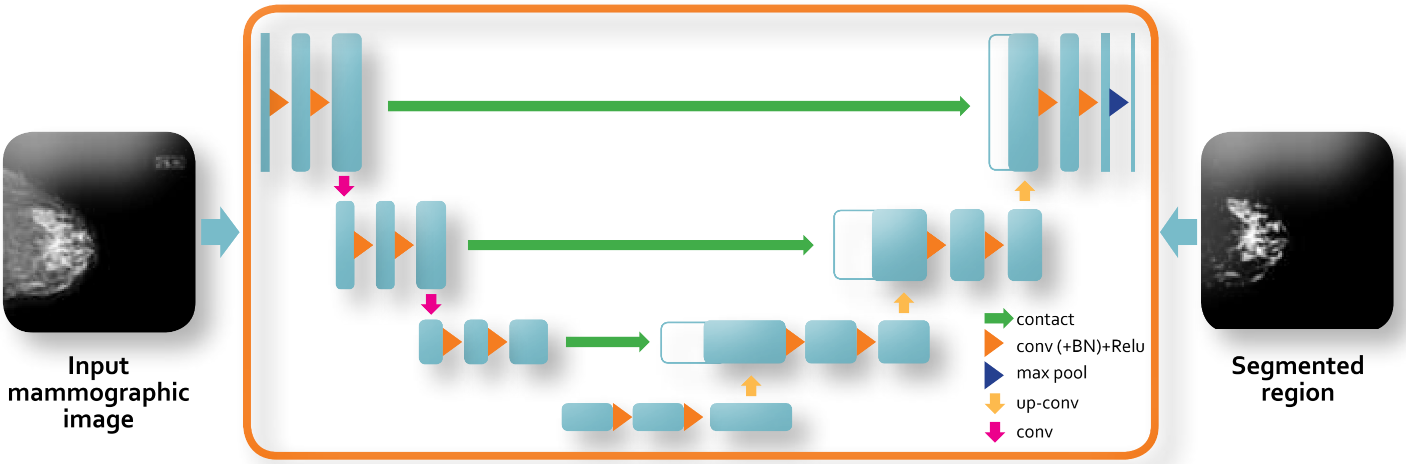
"We try to localise regions of interest on a mammogram and classify the tumour type based on features extracted using deep learning algorithms," says Gudhe, who works as a data analyst at the Institute of Clinical Medicine of the University of Eastern Finland, in Kuopio.

Mammography, a low dose x-ray imaging technique, is one of the most widely used methods of detecting early-stage breast cancer. The early detection of breast cancer significantly reduces mortality rates. In 1987, Finland was the first country in the world to begin a country-wide cancer screening programme. However, mammography is not perfect. Mammograms are not particularly sensitive and can miss cancer cases, or can appear normal even when cancer is present.

Fully automated model for estimating breast density

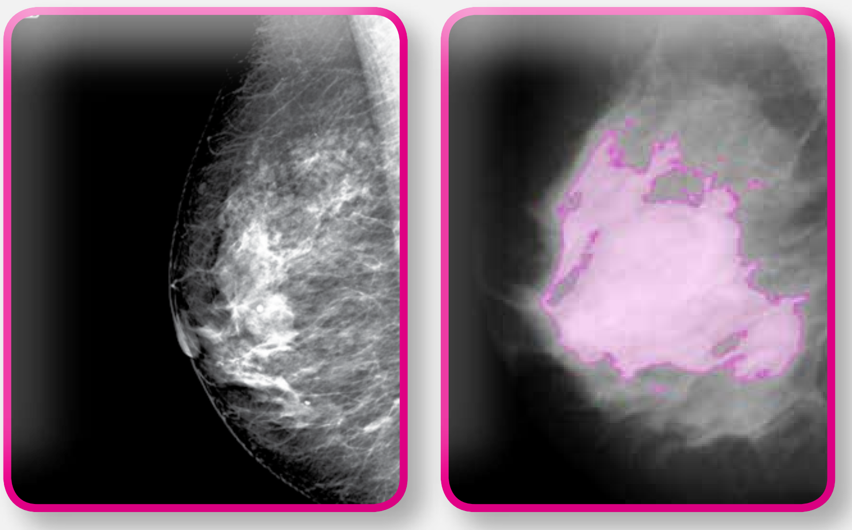
Breasts consist of variable portions of adipose (fat) and fibroglandular (dense) tissues. Being denser, fibroglandular tissues appear white (bright) on mammograms due to the attenuation of x-rays. Most cancers occur in the fibroglandular tissues, the brightness of which hide around 25% of cancers in mammograms.





Algorithms can identify tumour from mammograms. Breast density is one of the most commonly used breast cancer risk factor. The denser the breast, the greater the risk. Deep learning algorithms can assist radiologists to accurately predict the percentage of breast density. Mathematical operations performed to extract features from an image automatically.

Breast cancer usually occurs in the fibroglandular tissues. Due to the brightness of such tissues in mammograms, about 25% of cancers are not detectable through mammography. The brightness can mask cancer cases.



“This brightness can mask the presence of cancers; it is like finding a snowman in a dense snow cloud”, says Gudhe.

Based on fibroglandular tissue patterns and their distribution, radiologists categorise breasts as either ‘dense’ or ‘fatty’. Women with extremely dense breast tissue have a higher risk of developing breast cancer.

Researchers at the University of Eastern Finland and Kuopio University Hospital are interested in developing a fully automatic model for estimating breast density. Breast density, one of the strongest risk factors in breast cancer, is a measure of the relative amount of fibroglandular tissues. Accurate segmentation of fibroglandular tissues on a mammogram can reduce the likelihood of false diagnosis.

Algorithms developed by University of Eastern Finland can assist radiologists in the accurate estimation of breast density. The major challenge of using deep learning models is the massive amount of training data they require. In addition, in medicine the acquisition of images with accurate annotations adds to the complexity of data.

“We use thousands of mammogram images, which are manually annotated by an expert radiologist to generate accurate training set’s classification (e.g. ground truth labels)for our deep-learning models. We have developed a novel architecture based on the U-Net model, a state-of-the-art solution for the medical image segmentation of fibroglandular tissues”, says Gudhe

Because mammogram images are high-resolution, high computation power is required to train the related deep learning models. In this regard, the services provided by the Finnish ELIXIR centre CSC are used to handle sensitive data efficiently, training the models on graphic prosessing units provided by CSC.

Raju Gudhe emphasises that, in order to make a robust model for clinical implementations, researchers must integrate different imaging modalities and other clinical features with their algorithms. These are - in addition to mammograms - ultrasound, and magnetic resonance imaging. The next step is to integrate imaging and genomic data in order to analyse the cancer risk.

“Using mammograms, we can identify the density of a breast and, based on the density value, can generate the next image modality. We cannot rely on a single modality for the images, which is why the information cannot be used directly in clinical practice. To obtain an end-to-end model with good classification and diagnosis, we also need genomic data”.

Ari Turunen

MORE INFORMATION:

More information:

School of Medicine,
University of Eastern Finland
https://www.uef.fi/en/web/laake

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