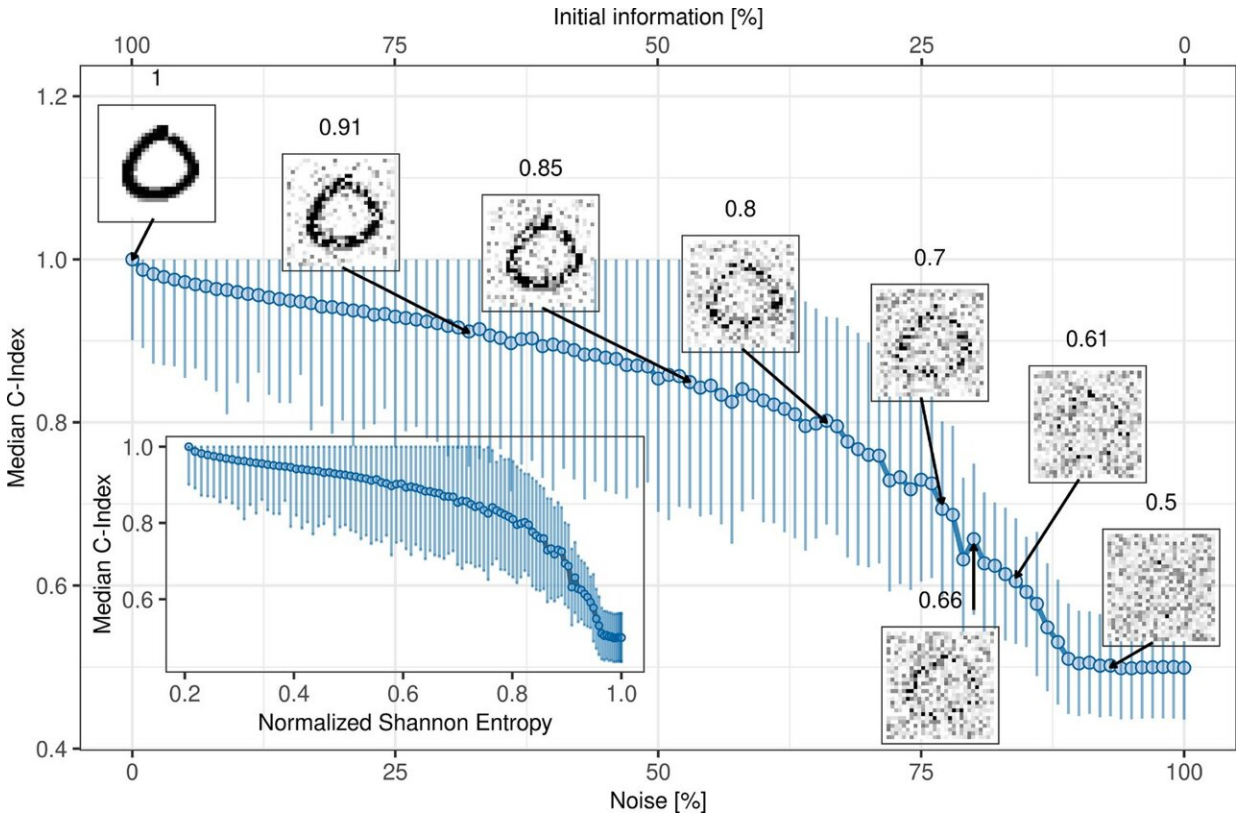
Medical Xpress

New insights into the prognostic power of gene expression signatures in breast cancer

October 5 2023, by Nina Vogt



Median C-indices of simulated survival time predictions based on MNIST data. We randomly sample 2,000 MNIST images and add varying levels of noise to them. We then reduce the images to 2-dimensional data points, which are used to simulate survival times using Cox proportional hazards models. The C-indices are computed and this process is repeated 100 times, with statistics recorded. The true survival times are based on the simulated survival times of images without noise. The initial information of an image is represented as 100%−noise[%]. The number above each exemplary image indicates the

calculated C-index of the corresponding image. In the inset, the dependency on the normalized Shannon entropy is shown, with error bars indicating the standard error of the median based on 100 random samples. Credit: *Scientific Reports* (2023). DOI: 10.1038/s41598-023-41090-9

When a woman is diagnosed with breast cancer, the difficult question arises as to which type of treatment is the right one. Gene expression testing is one of the methods used by doctors to help make a prognosis about the course of the disease and, based on this, to select a suitable therapy. However, the reliability of these tests has not been fully established.

Scientists from Leipzig University and the Pathologie Hamburg-West institute have now used machine learning to analyze large amounts of data on this question and found that gene expression signatures offer a high degree of certainty in prognosis, but not complete certainty.

Gene expression signatures are descriptions of the activity patterns of genes. When a person is diagnosed with cancer, these signatures can be used to make predictions about how tumors will develop. As such, they are crucial for classifying different types of cancer, determining prognosis and defining treatment strategies.

The current study by Dimitrij Tschodu, a doctoral researcher at the Peter Debye Institute for Soft Matter Physics at Leipzig University, was carried out in close collaboration with Professor Axel Niendorf from the Pathologie Hamburg-West institute and was recently published in *Scientific Reports*. Tschodu and his colleagues analyzed about 10,000 signatures based on breast cancer databases using various machine learning models to thoroughly assess their prognostic ability.

The results of the study show that the gene expression signatures examined lead to a correct patient prognosis in no more than 80% of cases. The researchers also point out that prognoses based on gene expression signatures alone take into account less than 50% of the potentially available information. They therefore recommend using other parameters in addition to gene expression tests.

"Although our results confirm the importance of gene expression signatures in predicting patient prognosis, they also highlight the urgent need for a holistic approach that takes into account molecular, clinical, histological and other complementary factors to ensure an accurate prognosis," explains Tschodu.

Need for a holistic approach to prognosis

"The results of this study are crucial for understanding the limitations of gene expression signatures in cancer prognosis," adds Professor Josef Käs, head of the Soft Matter Physics Division at Leipzig University. "While gene expression signatures are undoubtedly valuable, our findings show that a holistic approach is needed to ensure an accurate prognosis and to make informed decisions about treatment."

The publication comes from the physics of cancer research field, which looks at cancer from a physical perspective and also examines the mechanics of cells and tissues. Käs says, "This new study underlines the importance of the 'physics of cancer' in the medical field and the need for interdisciplinary collaboration to find innovative solutions to the challenges in cancer treatment."

Only recently, a research group led by Professor Käs and Professor Niendorf published new findings in this field that could promote more precise diagnostics of the spread and formation of metastases in breast tumors.

More information: Dimitrij Tschodu et al, Re-evaluation of publicly available gene-expression databases using machine-learning yields a maximum prognostic power in breast cancer, *Scientific Reports* (2023). DOI: 10.1038/s41598-023-41090-9

Provided by Leipzig University