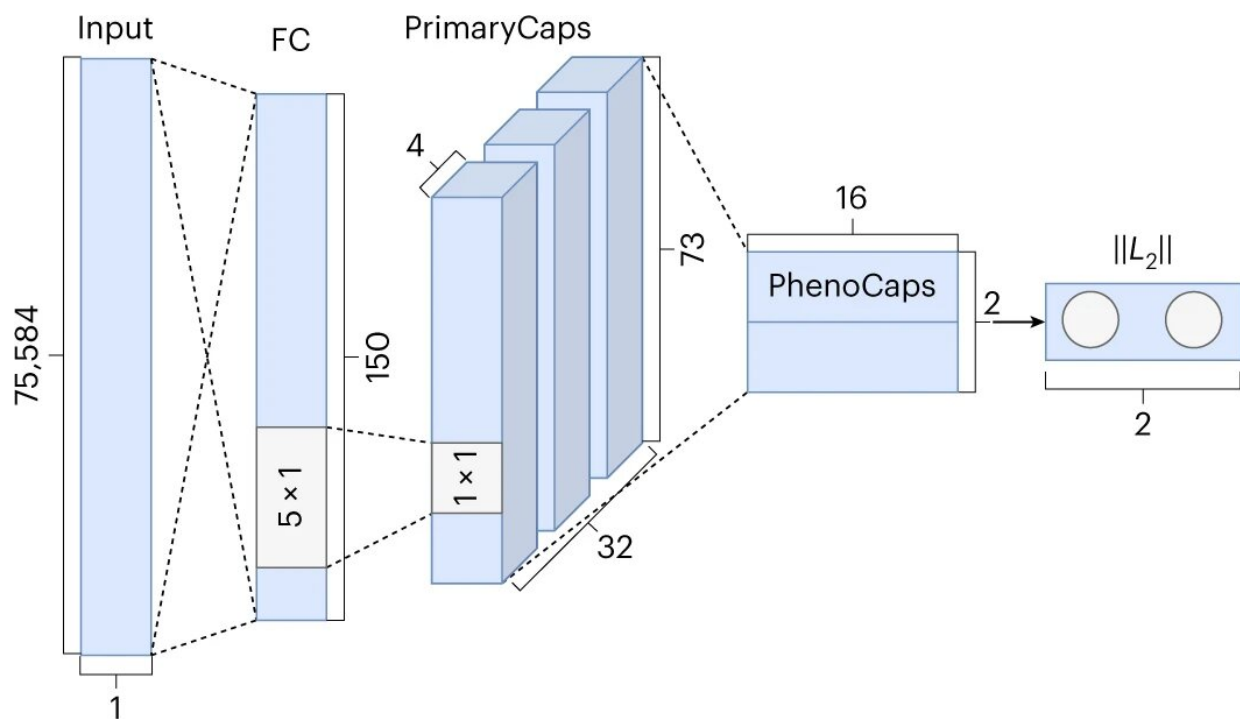# Predicting development of ALS disease with AI methods

February 28 2023

The architecture of DiseaseCapsule. The input is the concatenation of the compressed features from all Gene-PCA models, where each feature corresponds to one Gene-PCA. The number of Gene-PCAs is 75,584, so the dimensionality of the input is 75,584 × 1. DiseaseCapsule consists of three layers: a fully connected layer (FC), a primary capsule layer (PrimaryCaps) and a phenotype capsule layer (PhenoCaps). The FC layer consists of 150 neurons followed by ReLU as activation function. The PrimaryCaps is composed of 32 primary capsules. Each of them involves four different convolutional filters (kernel size 5 × 1, stride 2, no padding). PhenoCaps consists of two 16-dimensional vectors. Each phenotype capsule receives input from all 32

primary capsules. The output is a binary classification label (Healthy or ALS). Credit: *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-022-00604-2

Amyotrophic lateral sclerosis (ALS)—a severe disease of the motor nervous system—is hereditary, but a large part of this heritability has previously been unexplained. Using artificial intelligence (AI) methods, researchers led by Professor Dr. Alexander Schönhuth from Bielefeld University's Faculty of Technology have succeeded in recording and deciphering the genotype profiles of 3,000 ALS patients and thus learning more about the development of the disease.

With the new method, it is possible to predict with 87 percent accuracy whether or not people will develop ALS. The researchers have presented their study results in the journal *Nature Machine Intelligence*.

In his Genome Data Science working group, bioinformatician Professor Dr. Alexander Schönhuth is developing methods and tools to work with tens of thousands of genomes and analyze their data. Alongside the Faculty of Technology, the working group also belongs to Bielefeld University's Center for Biotechnology (CeBiTec).

A current focus of research in Schönhuth's team is the disease [amyotrophic lateral sclerosis](link) (ALS), also known as motor neuron disease. In ALS, the patients' ability to move decreases over time until it comes to a standstill. At the same time, however, the brain's activity remains completely functional.

"There is still a great deal about ALS that we do not understand," says Alexander Schönhuth who led the study on diagnosing ALS. He started working on the topic four years ago when he was still head of a research group at the Centrum Wiskunde & Informatica (CWI) in Amsterdam,

the Dutch national research center for mathematics and computer science. "We know that ALS is a hereditary disease, but 80 percent of this heritability is still unexplained," says Schönhuth.

The genetic architecture of ALS is complex

"Many hereditary diseases reveal overlapping, so-called additive effects of genetic factors. An example is schizophrenia," Schönhuth explains. "The more of these factors the genes reveal, the more likely it is that a person will develop schizophrenia. We can therefore easily recognize the genetic disposition on the basis of the genes. With ALS, in contrast, things are much more complex."

Schönhuth and his team assume that individual factors alone are very likely to lead to ALS. However, if these factors occur together, the opposite is the case: no disease occurs. This assumption would explain why so little is still understood about ALS.

Analyzing the genetic data of 3,000 ALS patients with AI methods

The central method from [artificial intelligence](#) (AI) that Schönhuth and his team applied is called capsule networks. It was used to evaluate [genetic data](#) from 3,000 ALS patients and 7,000 people not suffering from ALS.

"The great advantage of this method is that it can capture overlapping processes." Classical methods cannot cope with the amount of data and do not deliver clear results. "Our AI method, in contrast, clearly and comprehensibly shows which genes and their processes are particularly important for the development of ALS," says Schönhuth.

More than 900 genes found to play a role in the development of ALS

The scientists' results show an 87 percent accuracy in predicting whether or not people will develop ALS. "Our method can make predictions about the disease, and it is much more accurate than other methods. We have found more than 900 genes that play a role in identifying the disease and 644 genes that interact in complex ways. These need to be investigated further in other fields of research," Schönhuth continues.

"Each gene is engaged in different biological processes: the more we learn about the genes, the more we learn about the processes. In this way, our results will help people affected by ALS to adapt their lifestyle and reduce their risk of suffering from the disease. In addition, drugs could also be developed that influence specific processes," Schönhuth explains.