Tidsskriftet Tidsskrift for Den norske legeforening

Machine learning in medical research

MEDICINE AND NUMBERS

GURO F. GISKEØDEGÅRD

guro.giskeodegard@ntnu.no
Guro F. Giskeødegård, associate professor in biostatistics at the K.G. Jebsen Center for Genetic Epidemiology, NTNU.
The author has completed the ICMJE form and declares no conflicts of interest.

STIAN LYDERSEN

Stian Lydersen, dr.ing., and professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU of Central Norway) at the Department of Mental Health, NTNU.
The author has completed the ICMJE form and declares no conflicts of interest.

Machine learning is used to find underlying patterns in data. This can be useful in medical research.

Machine learning is a form of artificial intelligence that is used to find underlying patterns in data. It can be based on statistical methods, or other techniques from mathematics or informatics that are not based on probability models. Machine learning is particularly useful for datasets with a large number of variables, and the learning entails training a model to identify associations between the variables. The purpose is often to build a model that can predict an outcome. A typical characteristic of many machine learning methods is their iterative training process, where a single change is made at a time, thereby improving the data adaptation step by step. Within machine learning, the term 'feature' is often used synonymously with the term 'variable' in statistics, and the term 'label' is synonymous with 'outcome variable'.

Unsupervised machine learning

A distinction is often made between unsupervised and supervised machine learning. In unsupervised learning, only the measured variables are used to identify associations and clusters in the data, without predicting an outcome variable. In a study that measures gene expression from blood samples of cancer patients and healthy controls, unsupervised machine learning will only use the gene expression data in the analysis, with no information about which samples belong to patients and controls. The model will tell us whether we have natural clusters in the data, for example that the age of the individuals has a strong correlation to the gene expression, or that samples from the patients are distinct from those of the controls. Unsupervised methods are also well-suited for detecting extreme values or extreme combinations of values. Hierarchical cluster analysis, principal component analysis and the relatively new UMAP (Uniform Manifold Approximation and Projection) method are examples of unsupervised machine learning methods (1). Common for these methods is their capability to visualise large datasets with a high number of variables in only a few dimensions.

Supervised machine learning

In supervised machine learning, one or more labels (outcome variables) are used to train the model, and the result is a model that is optimised to find the association between the explanatory variables and some characteristic or outcome of interest.

The outcome variable can be a continuous or a categorical variable. In the example of gene expression data, a supervised machine learning model will be informed about which samples belong to patients and which belong to controls. This enables us to build a model that can predict whether a given gene expression belongs to a cancer patient or a control.

Some supervised machine learning models function as a black box, where the model can predict the status of a new sample but does not provide information about *why* the sample is given this status. Neural networks and XGBoost (Extreme Gradient Boosting) are examples of popular black box models. Such models are less useful if we want to understand the underlying biology that distinguishes patients from controls. However, there is a major focus within research on 'opening up' such black boxes with a view to understanding their modelling processes, and thus growing interest in the field of explainable artificial intelligence (XAI). Some supervised machine learning methods automatically provide information about which variables are important for prediction, such as certain regression models.

Validation is crucial

A complex supervised machine learning model can identifying patterns in data very well. It can actually do the job so well that it is overfitted to the data it has learned from. An overfitted model will describe the data it has learned from well, but it will perform poorly with new data. Thus, biological interpretations of such a model will yield imprecise or incorrect information. It is therefore crucial that supervised machine learning models are robustly validated (Figure 1). This requires training and optimising the model using a training set of data, such as a random sample of 80 % of the data. The final model is then validated using data that were not used in the learning process (often referred to as the validation set).

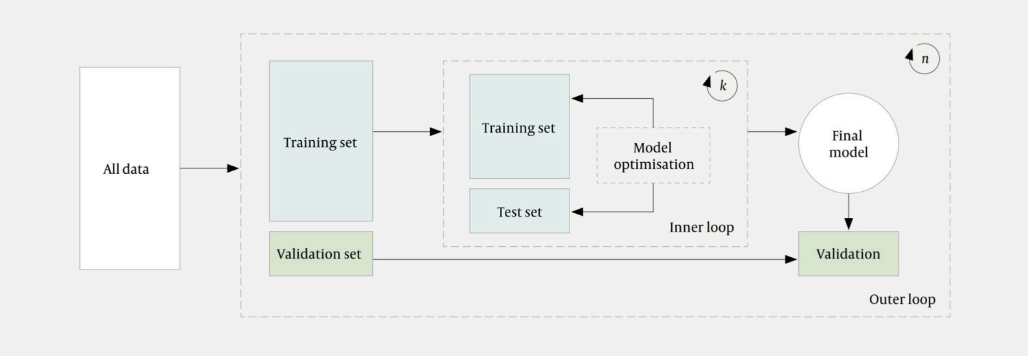

Figure 1 During validation of a supervised machine learning model, the data are split into a training set and a validation set. The model is optimised through a learning process, and the final model is validated using the validation set. The training set is often split into further training and test sets in an inner loop, where the model is optimised. The inner loop is usually repeated several times (here: k times) with a random division into training and test sets. The outer loop can also be repeated n times to estimate the variation in the model's performance.

REFERENCES

1. Røislien J, Langaas M. Klynger. Tidsskr Nor Legeforen 2022; 142: 1586. [CrossRef]

Publisert: 29 May 2023. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.23.0196
Copyright: © Tidsskriftet 2026 Downloaded from tidsskriftet.no 14 February 2026.