

## What is Machine Learning?

1 message

FindingPheno <newsletters@findingpheno.eu>  
Reply-To: FindingPheno <newsletters@findingpheno.eu>  
To: findingpheno.eu@gmail.com

Tue, Mar 22, 2022 at 2:30 PM

[View this email in your browser](#)

Newsletter 7 | March 2022



We have been making explainer videos for FindingPheno and, while they are being animated, I have been writing blog posts to support some of the learning objectives for the videos. This is the first one, it was too much fun writing this to hold it back until the videos are done!

Note: I have been having problems with these newsletters going into spam, so if you didn't receive one about fish a while back, take a look in your junk folder!



### What is Machine Learning?

**Machine learning, or ML, is a computer algorithm which is able to adapt and learn based on the data you give it.** This is a type of artificial intelligence where the algorithm is able to take in large amounts of data and make predictions or decisions based on that data. By adding the ability to learn, ML technology become very flexible, able to work with a wide range of data types or different tasks, and can result in an analytical tool that is more efficient and powerful than a human can design. ML is particularly suitable for analysing large, complex, interacting data sets such as the omics data found in bioinformatics, where the ML system learns how to find patterns in the data that are not obvious or linear, untangling the interactions to make meaningful suggestions.

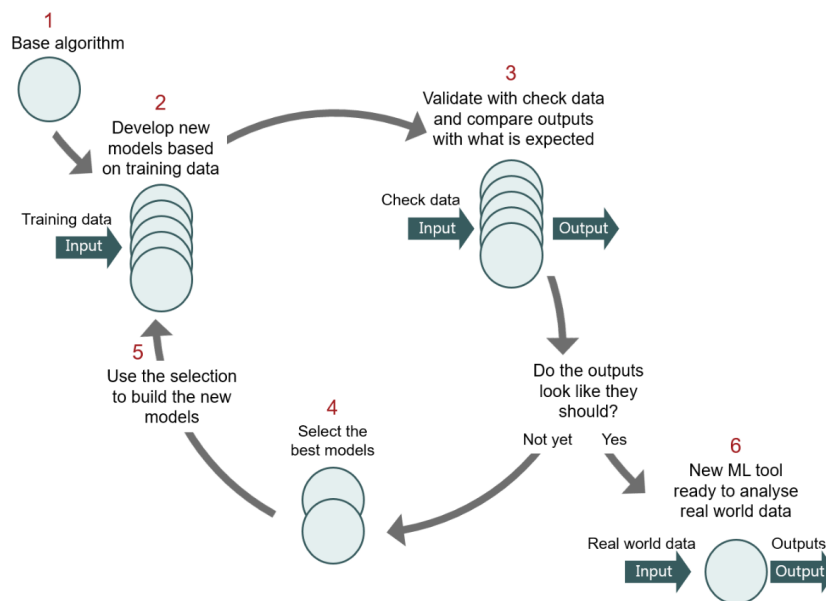
An ML system has three parts: **data input**, one or more **hidden processing layers**, and

## outputs.

**Data input** is where the information we want to analyse is listed in a format that the machine can understand and then uploaded into the computer for analysis. It also often involves simplifying the data in some way (aka "dimensionality reduction"), filtering out noise and streamlining the analysis to make it faster.

The **outputs**, meanwhile, usually involve making some kind of categorisation or prediction based on what the model has learned. This information allows us to answer biological questions like: is a sample cancerous or healthy, what signalling pathways are involved in a disease, or will changing an animal's diet make it grow more or less? It is important to define the type of outputs you are looking for before you start developing the system as this what guides the learning process for the machine.

The most important part of ML is the **hidden layer**. This is where the analysis model is found, i.e. the algorithm or set of calculations that turns data into outputs, and is the part that learns. To build this hidden layer, we start with a basic analysis framework then feed in lots and lots of training data -i.e. data sets of the type we want to analyse but where we know what the outputs from that data should be- and teach the system what we are looking for. There are multiple ways this can be done depending on which system you choose, but generally this is a cyclic process where the system is trained then tested then trained again (and so on), allowing it to evolve and adapt to the data until the outputs best match what we expect. One example is genetic algorithms, where Darwinian selection is used to evolve the algorithm. The general workflow for this evolution is described in the diagram below:



1. Set up the software with a basic algorithm as the analysis model.
2. Input training data and generate new analysis models.
3. Check the models against a second set of data and measure the quality of their outputs.
4. Select one or more of the best models and throw away the rest.
5. Starting from this selection, input training data again and develop more models.
6. Continue the cycle until output quality is good and the ML system is ready.

**There are two main types of ML: supervised learning or unsupervised learning**, with each one defined by the type of data that is used to train the system.

**Supervised learning** uses labelled data, i.e. where you tell the machine what each sample or data point really is. So for example, you feed in a lot of different patient data sets that are labelled to say which ones have cancer and which ones don't, then the system learns to tell the difference between the two. Once it has been trained to make this

distinction, now you can input data from a new patient and have the system tell you if they have cancer or not. The benefit of this system is that it can become better at distinguishing between different sample types than a human, improving speed and accuracy and maybe even detecting cancer at an earlier stage.

**Unsupervised learning**, meanwhile, uses unlabelled data where only the measurements of interest are included without indication of what they may be. This allows the machine to look at the structures within that data and figure out for itself what is important. This type of ML is good at finding hidden patterns or clusters within the data and working out how they relate to each other. For example, you can input data describing gene expression in cancer samples and let the system find which genes or groups of genes might be working together to cause the cancer. This ability to uncover unexpected or novel patterns within very complex data makes unsupervised learning particularly suitable for analysing large omics data sets to find previously unknown biological mechanisms.

So now I hope you have an idea of what Machine Learning really is. There are more posts to come giving examples of how this technology can be applied to interesting questions using biological big data.

[Read and comment on the full post here.](#)

---

If you like interesting things sent to your inbox, [take a look at The Sample](#). They will send you a different newsletter every day, week, or month based on your preferences. I found some really cool things there already and its free and super easy to use.

---



---

Copyright © 2022 FindingPheno, All rights reserved.

---



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952914.

---

This email was sent to [findingpheno.eu@gmail.com](mailto:findingpheno.eu@gmail.com)  
[why did I get this?](#) [unsubscribe from this list](#) [update subscription preferences](#)  
FindingPheno · Øster Farimagsgade 5A · København N 1353 · Denmark

Grow your business with  mailchimp