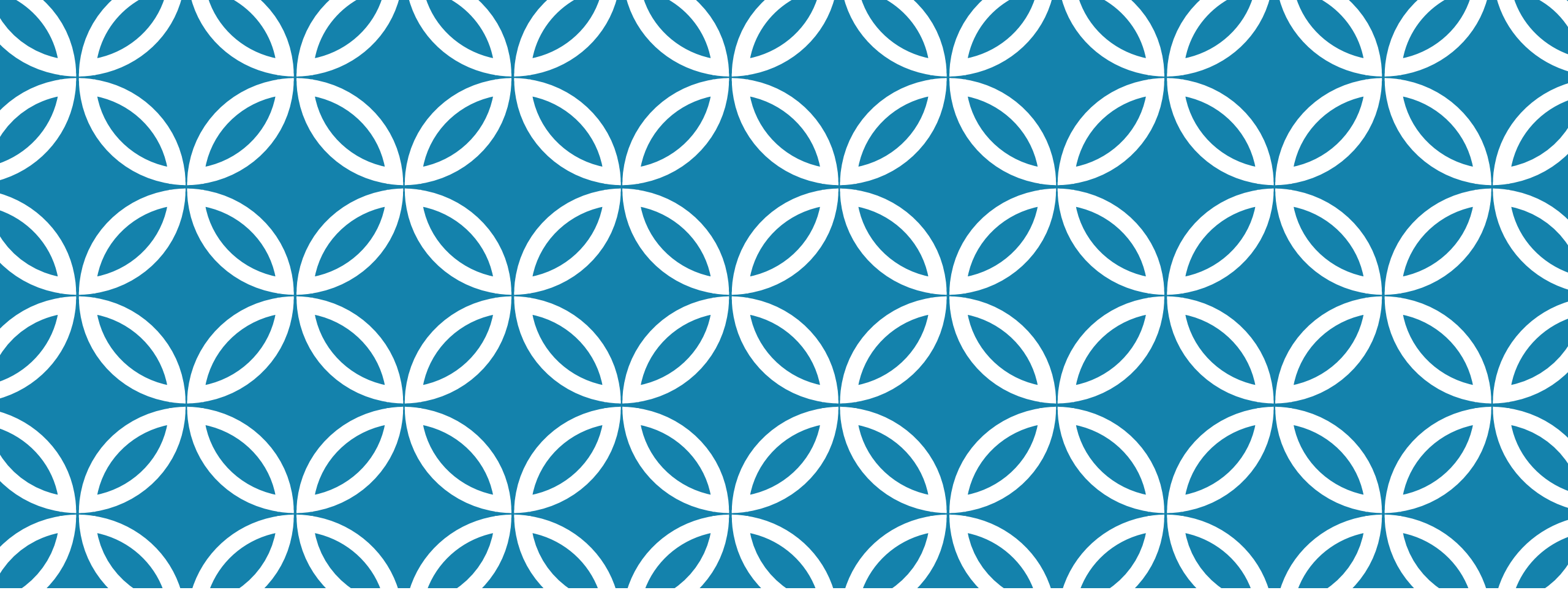


APPLICATIONS OF MACHINE LEARNING IN DRUG DISCOVERY AND DEVELOPMENT

Presentation by:

- Carina Luchsinger Salinas
- Joshua Kanters
- Samuel Börlin

USES OF MACHINE LEARNING IN DRUG DISCOVERY

- Identify novel targets
- Improve compound design and optimization
- Developing new biomarkers
- Improving analysis of biometric data
- And more…

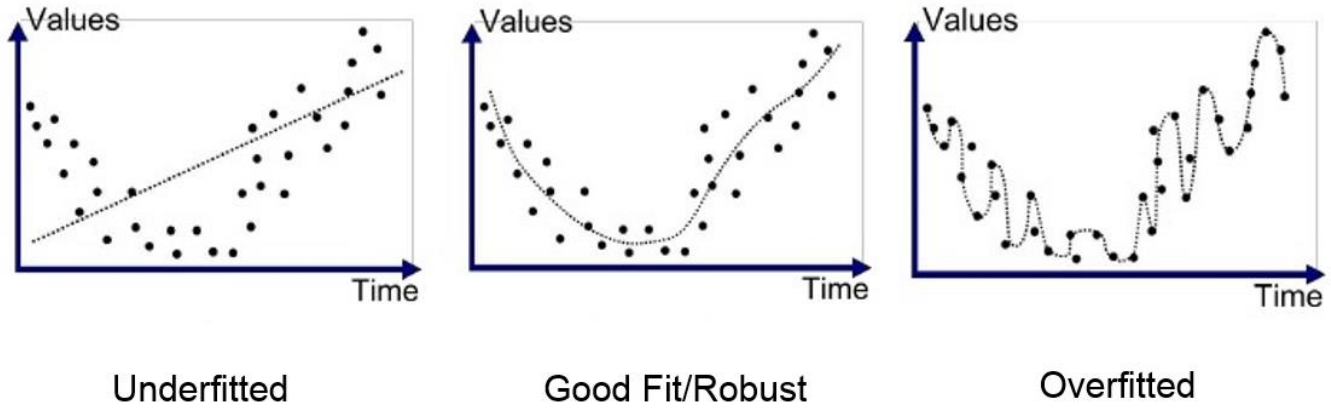
→ Employed in nearly all stages of drug discovery and development

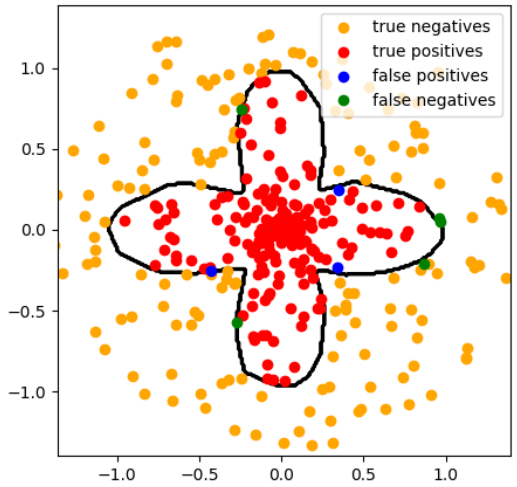
HOW IS MACHINE LEARNING (ML) USED?

1. Choose suitable ML algorithm
2. Feed it a lot of data to make it learn e.g. statistical distributions or decision boundaries
3. Run trained ML algorithm on new data to make classifications or predictions
4. If results are not satisfactory, go back to 1. and try again with a different algorithm or parameters

→ ML algorithm *learns from training data*,
instead of hand-crafting some classifier or predictor

CHOOSING THE RIGHT ML ALGORITHM



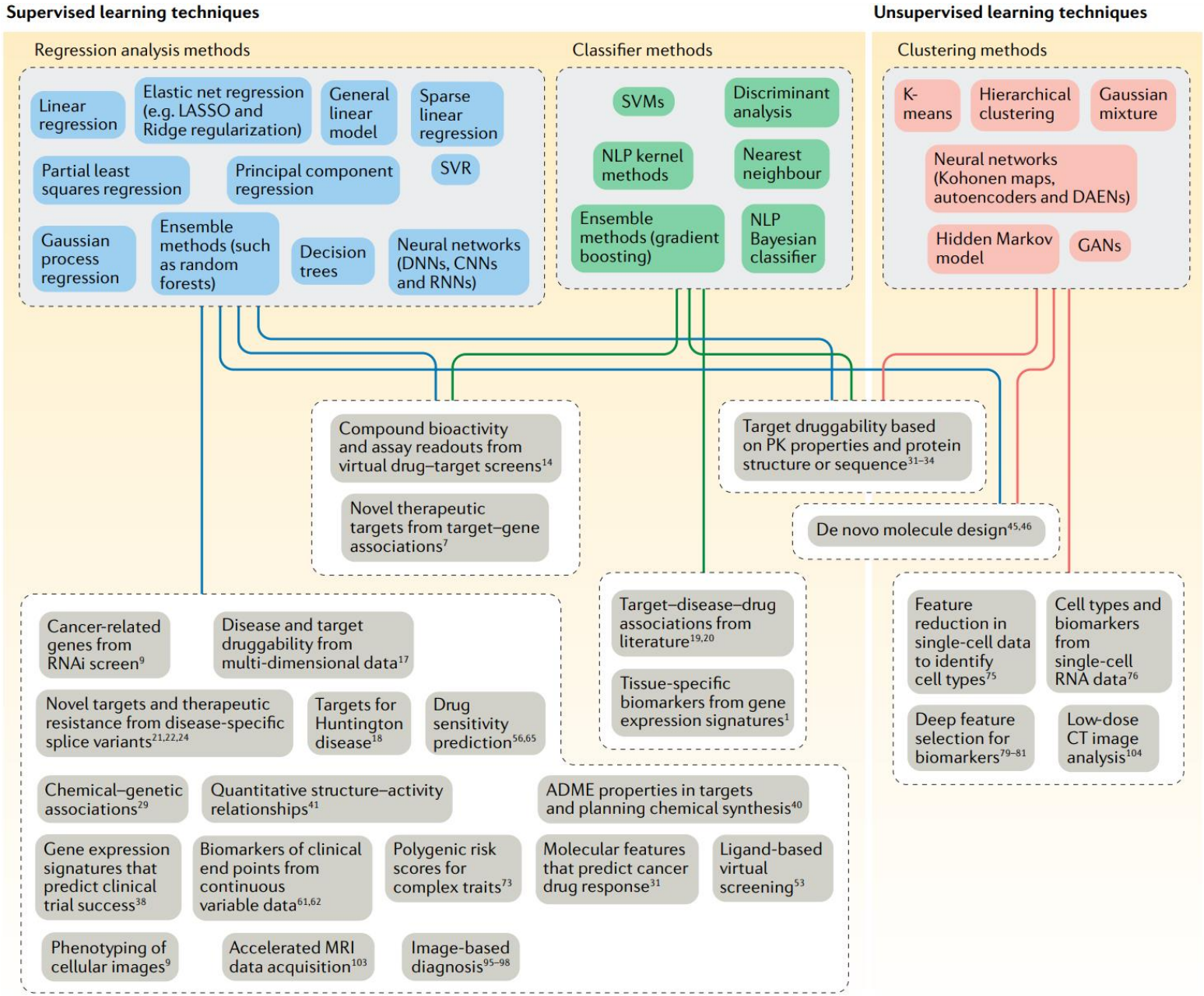


Goals:

- Good generalization, prevent overfitting (e.g. with dropout)
- Good classification or prediction accuracy

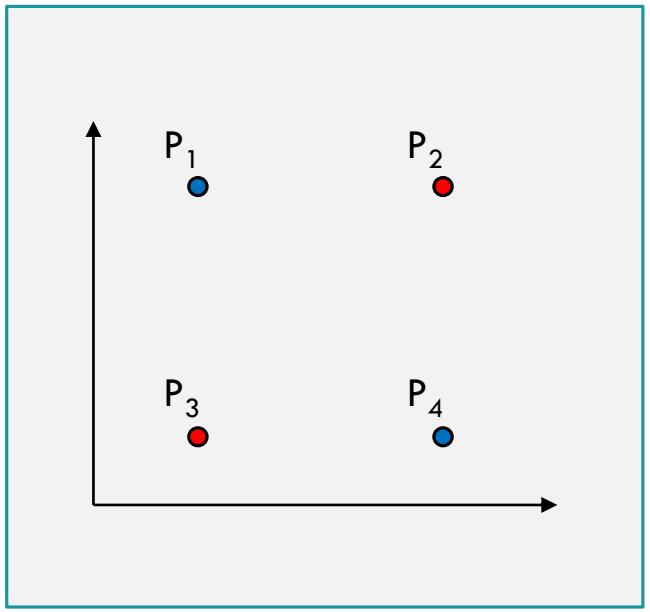
Two main categories: Supervised and Unsupervised

Source of figures: https://towardsdatascience.com/ and Pattern Recognition (10907) lecture exercise



Source of figure: doi.org/10.1038/s41573-019-0024-5

NEURAL NETWORKS: STRUCTURE

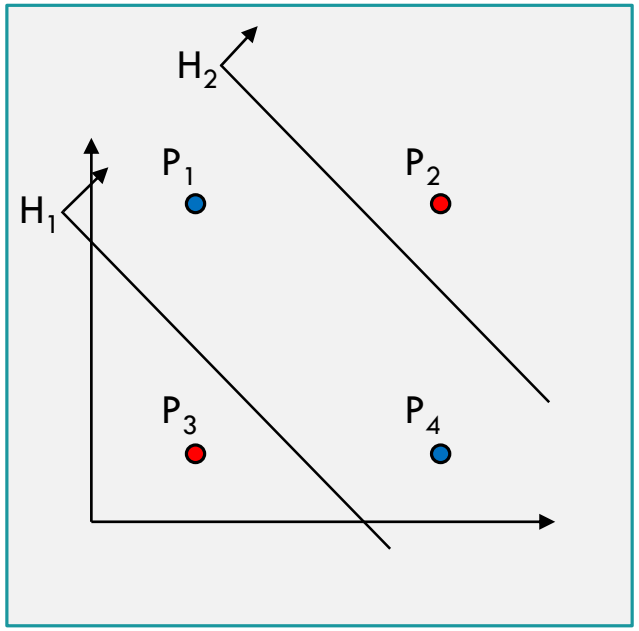


- Class 1 ($P_{1/4}$)
- Class 2 ($P_{2/3}$)

4 points and two classes

How can we classify these points accurately using just hyperplanes (in 2D: lines)?

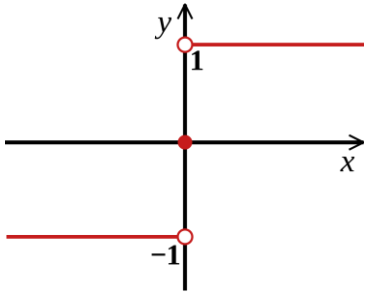
NEURAL NETWORKS: STRUCTURE



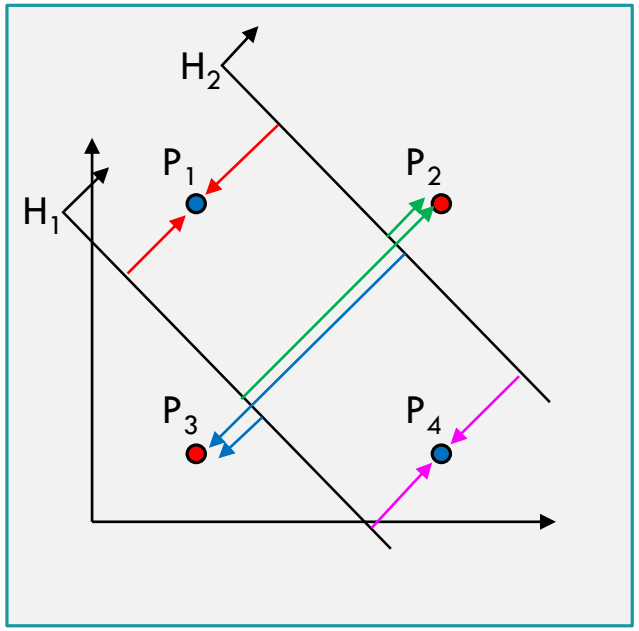
- Class 1 ($P_{1/4}$)
- Class 2 ($P_{2/3}$)

Calculate distances from hyperplanes

and run distances through activation function



Signum activation function



- Class 1 ($P_{1/4}$)
- Class 2 ($P_{2/3}$)

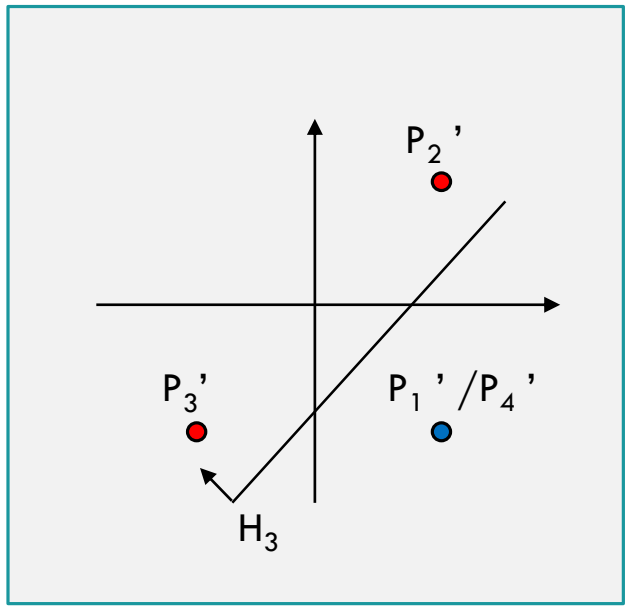
$P_1' = (1, -1)$

$P_2' = (1, 1)$

$P_3' = (-1, -1)$

$P_4' = (1, -1)$

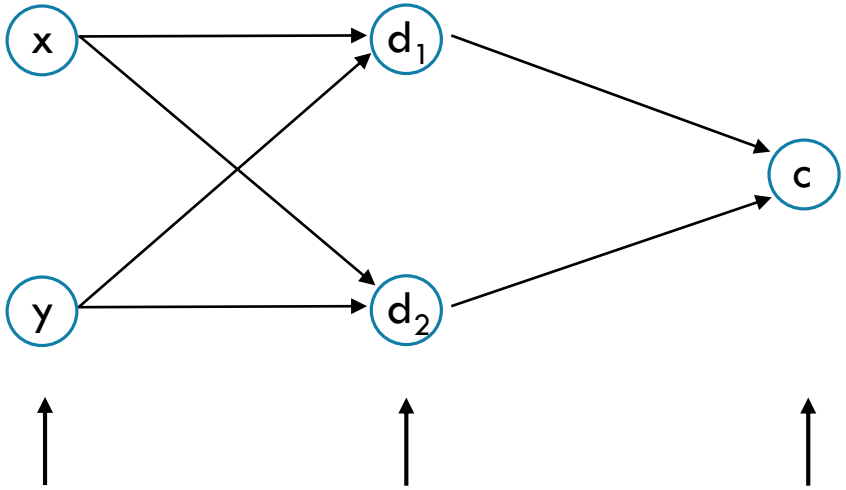
NEURAL NETWORKS: STRUCTURE



- Class 1 ($P_{1/4}$)
- Class 2 ($P_{2/3}$)

Now we can classify by checking on which side of the line the dots are!

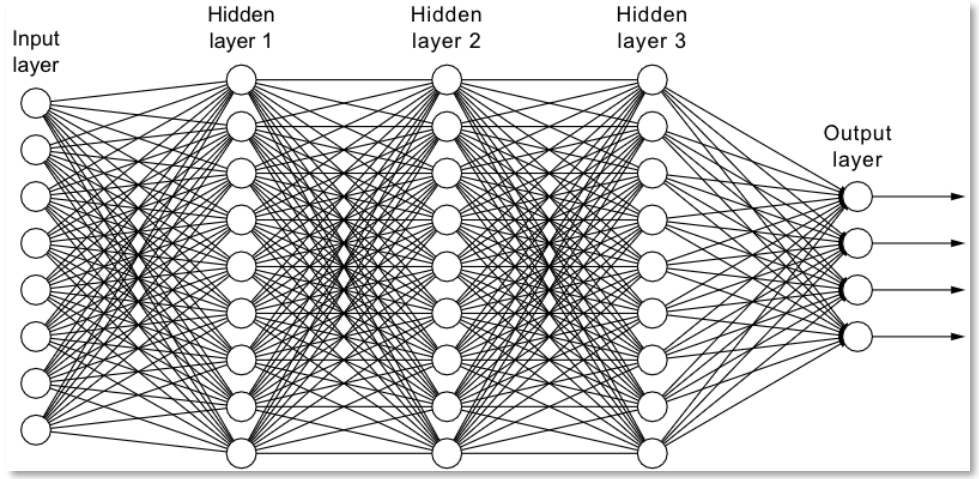
P_i coordinates Distances from $H_{1/2}$ Distance from H_3 → Classification!



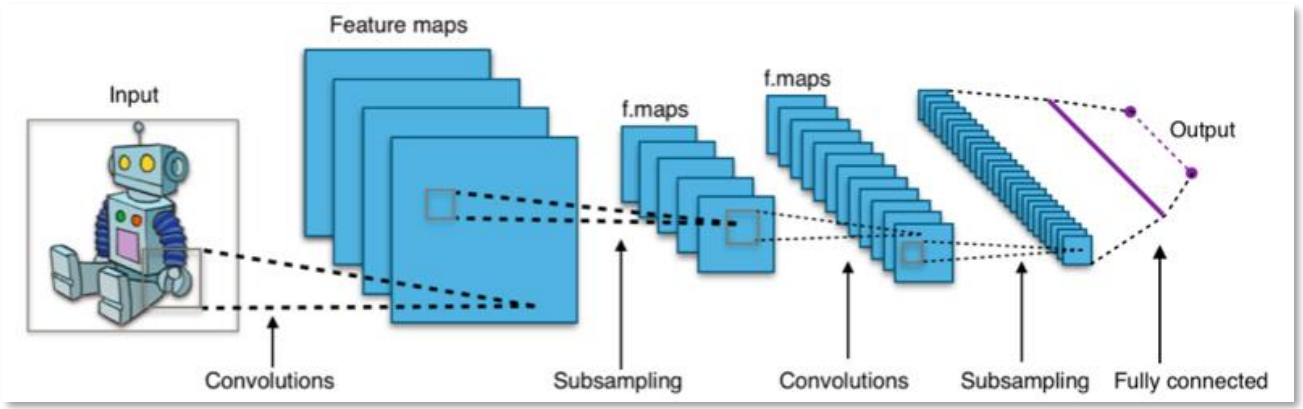
Input layer Hidden layer Output layer

These columns are called «layers» in machine learning terminology.
Circles are called «neurons».

NEURAL NETWORKS



Fully connected neural network



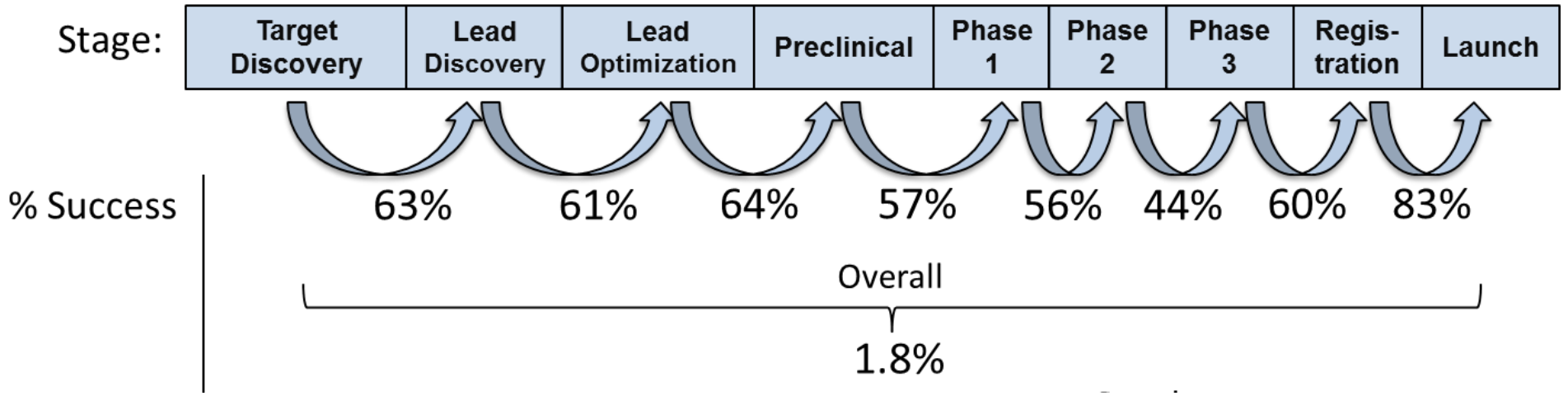
Convolutional neural network

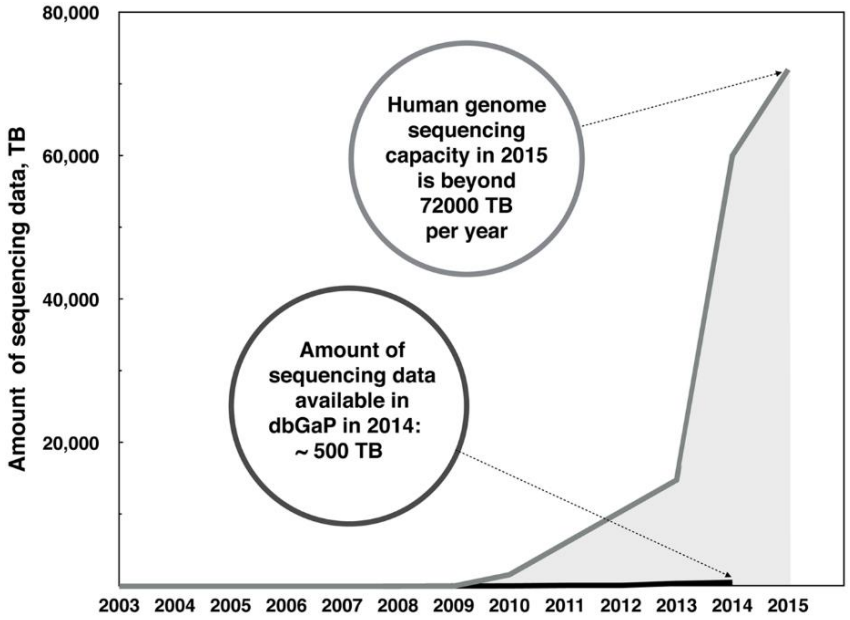
And *many* more: Deep NNs, Recurrent NNs, Long-short-term-memory NNs, etc…

Source of figures: https://freecontent.manning.com/ and https://www.wikipedia.org/

Application in Drug discovery

Target Identification and Validation

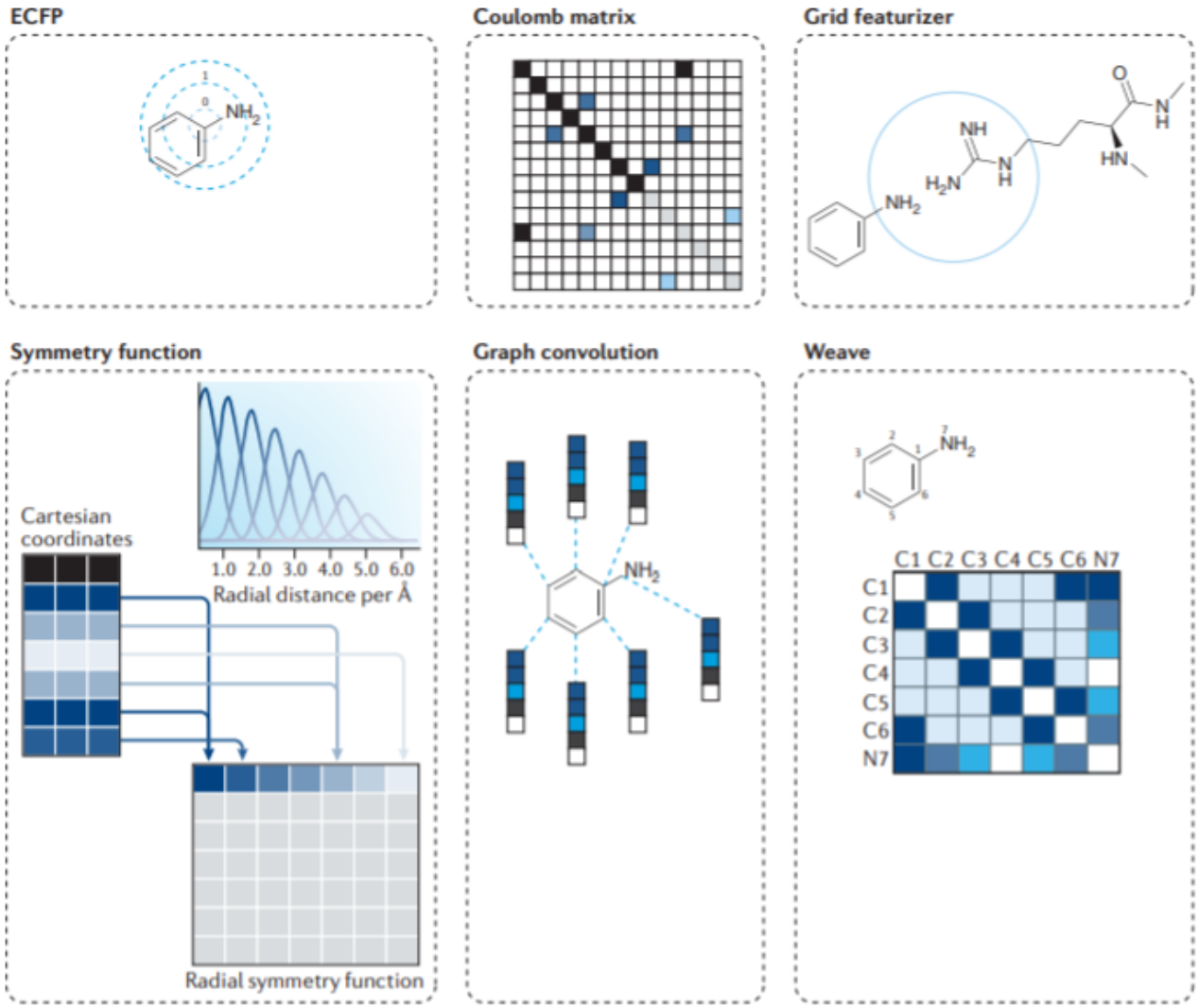




Source of figures: doi.org/10.1038/s41573-019-0024-5

Application in Drug discovery

Small-molecule design and optimization

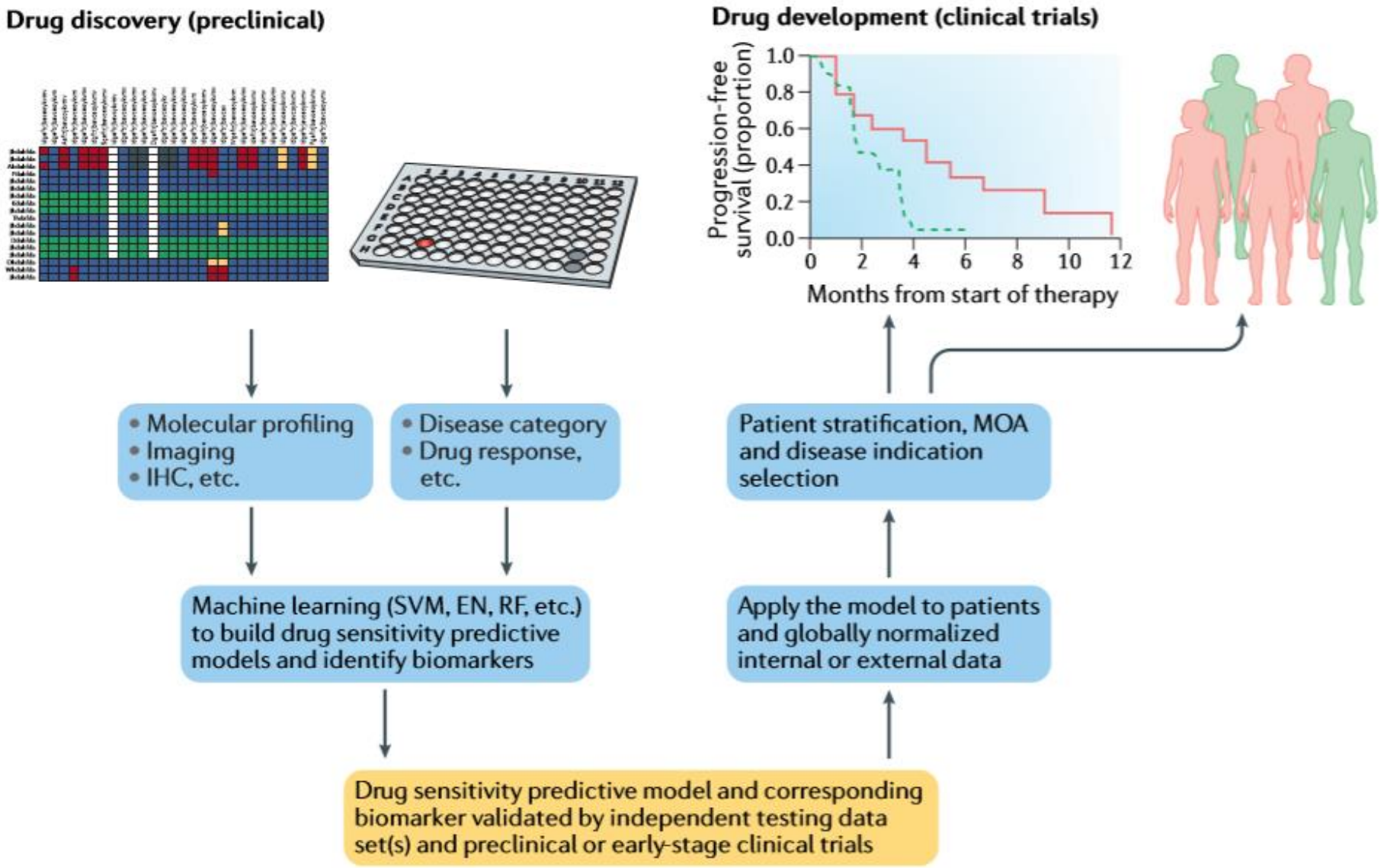


Source of figure: [doi.org/10.1038/s41573-019-0024-5](doi.org/10.1038/s41573-019-0024-5)

ML-BASED BIOMARKER DISCOVERY

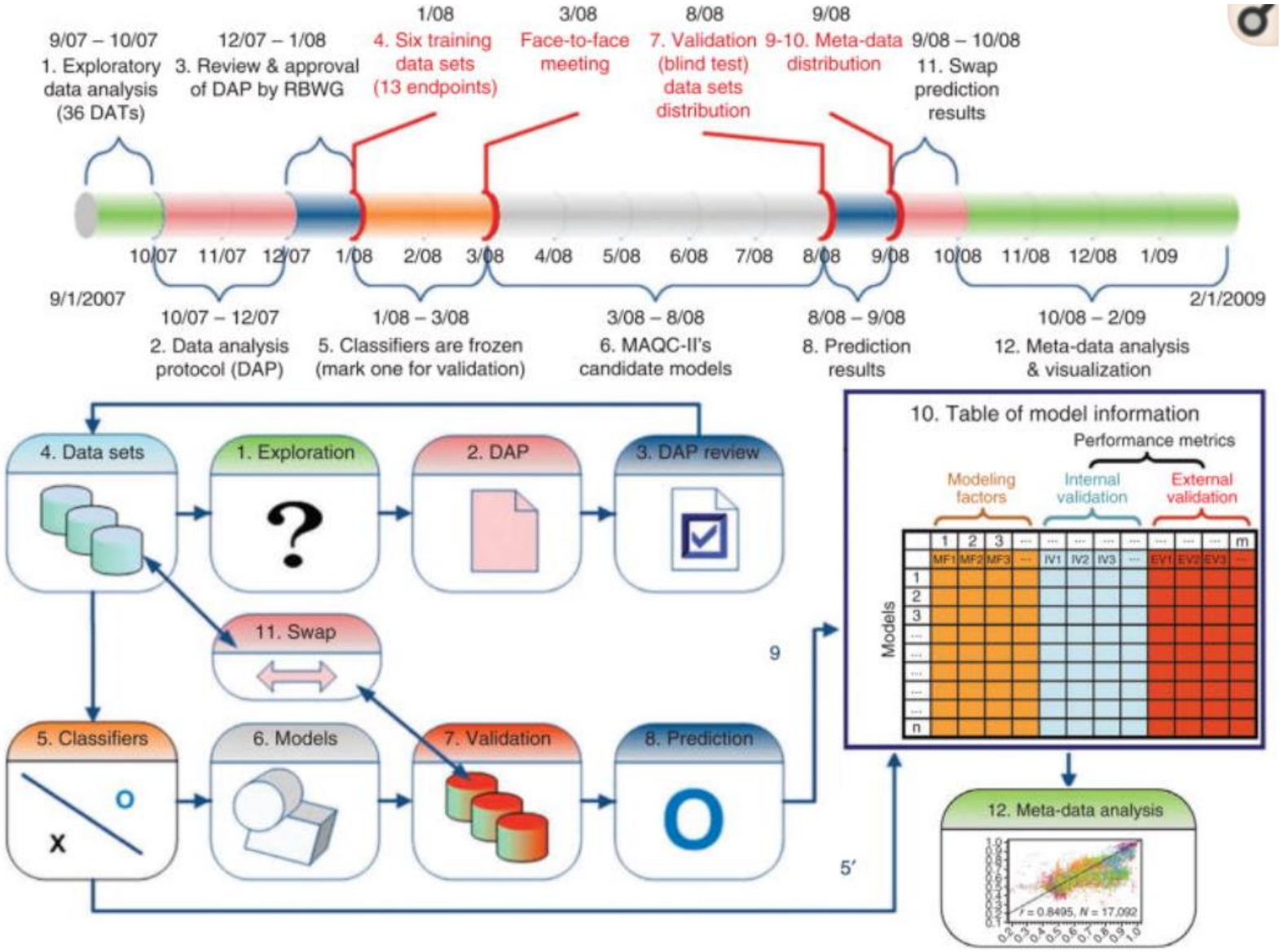
- To improve clinical success rates
  - Identify the right drug for the right patient
- Using ML on preclinical data sets for biomarker prediction
- Why is biomarker prediction not applied for clinical trials?

Predictive Biomarkers



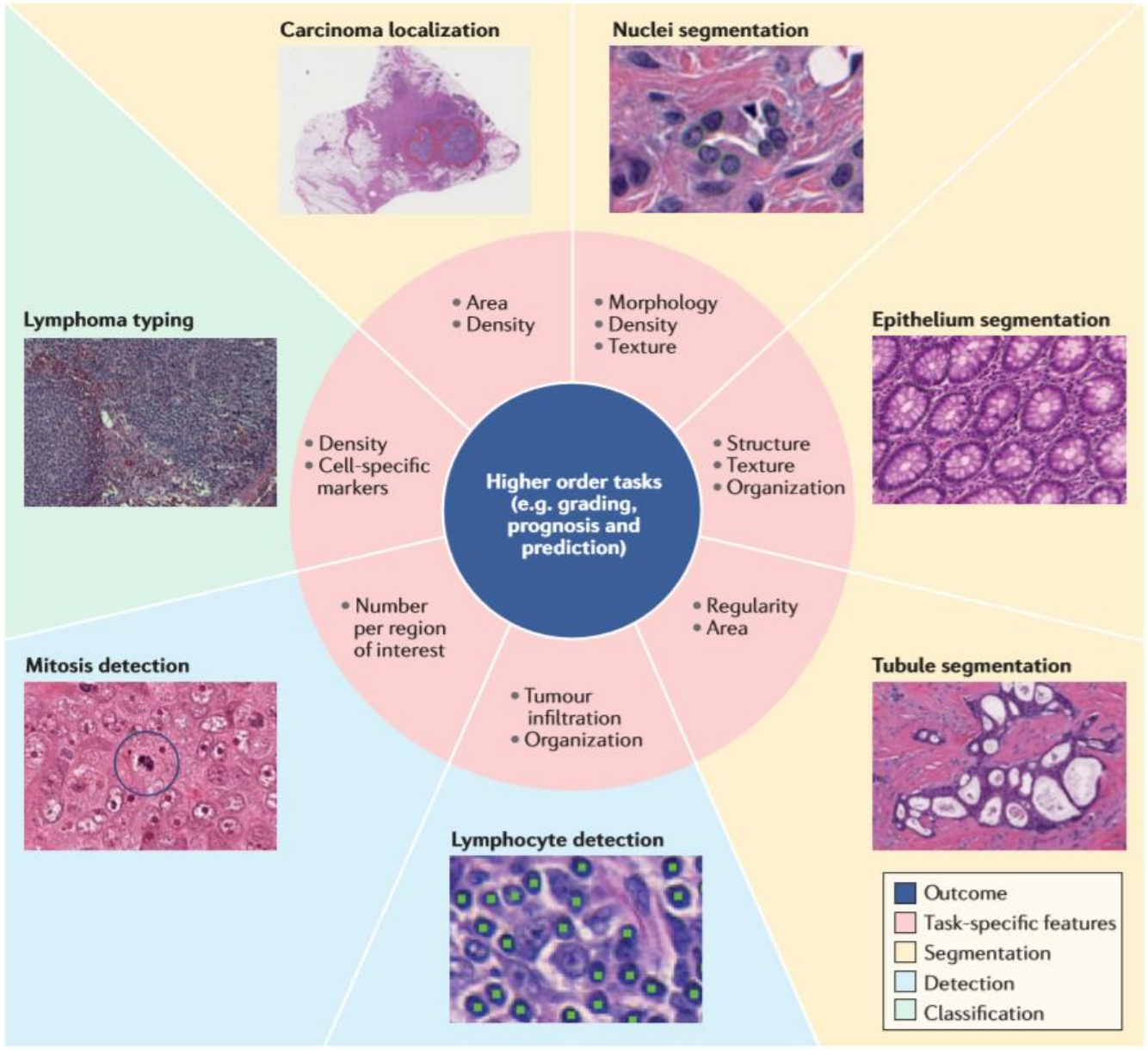
Source of figure: doi.org/10.1038/s41573-019-0024-5

MicroArray Quality Control II initiative



Source of figure: Nat Biotechnol. 2010 Aug;28(8):827-38. doi: 10.1038/nbt.1665.

Computational Pathology



Source of figure: doi.org/10.1038/s41573-019-0024-5

CONCLUSION

Positive

- ML and DL can increase efficiency across drug discovery and development
- ML applications: target identification & validation, drug design, biomarkers, pathology and prognosis in the clinic
- ML already applied in pharmaceutical companies

Negative

- Lack of interpretability
- Repeatability is not ensured
- Patent application problems
- Underestimation of ML results?
- Availability of high-quality, accurate data