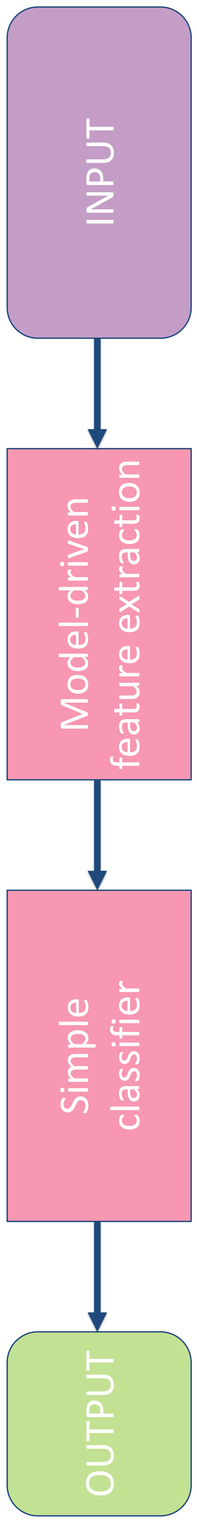
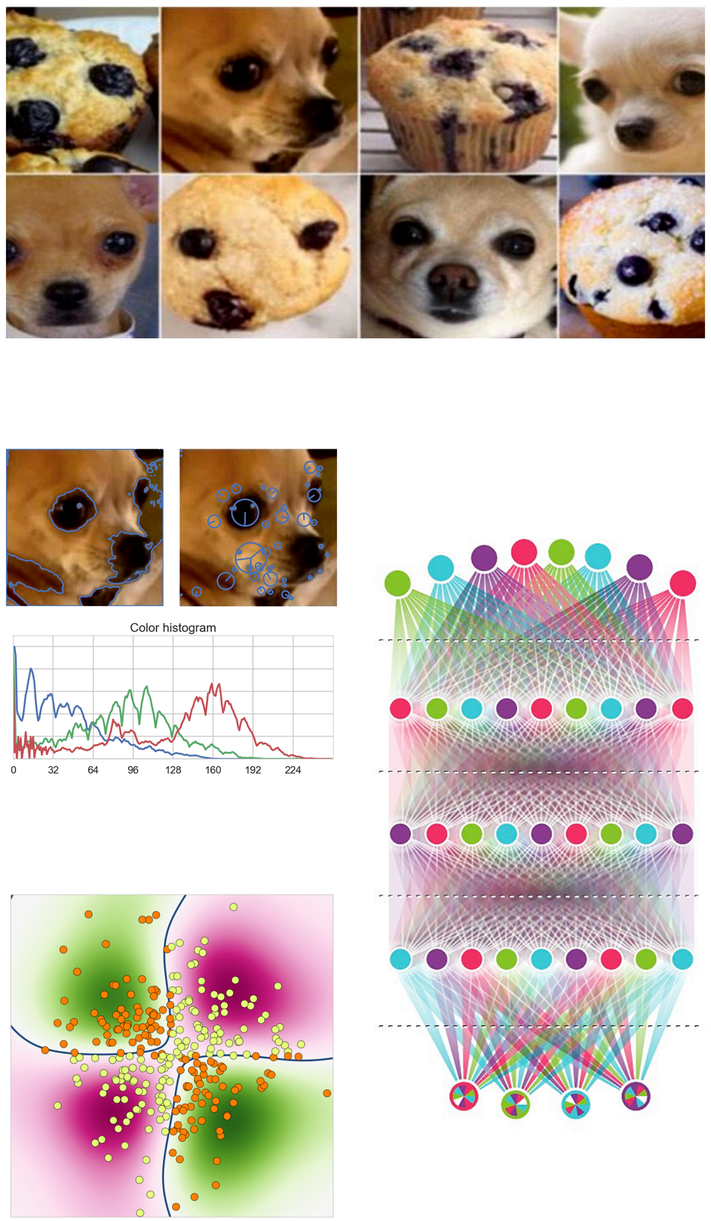
Mass spectrometry proteomics: Ready for the deep learning (r)evolution?

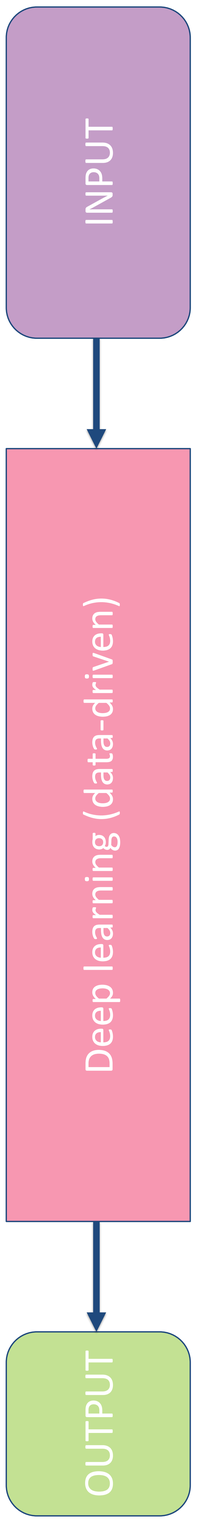
Wout Bittremieux[1,2*] **and Kris Laukens**[1,2]

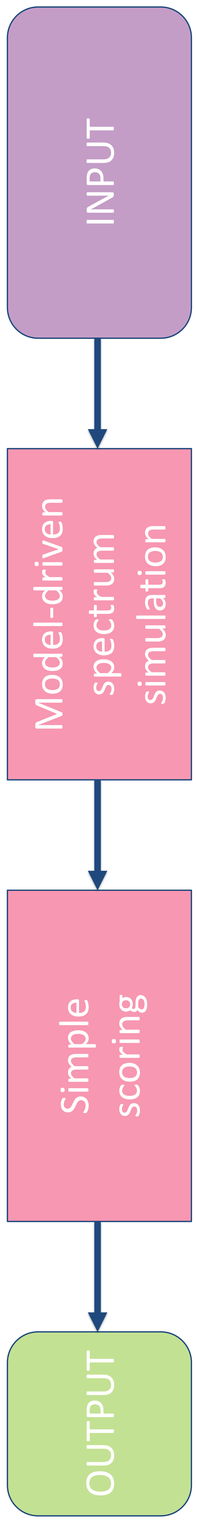
[1]Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium;
[2]Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp / Antwerp University Hospital, Antwerp, Belgium.
*wout.bittremieux@uantwerpen.be

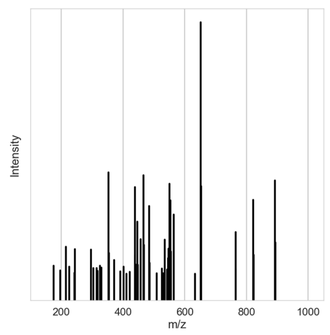


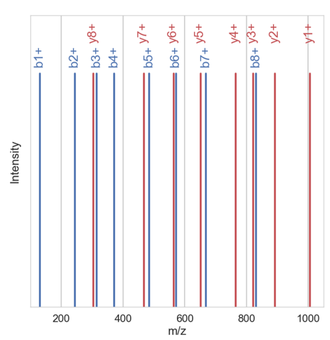


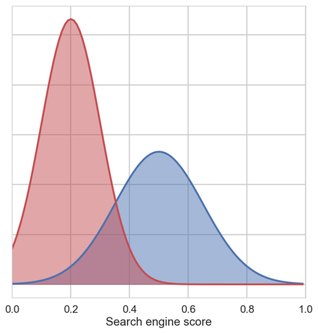
“CHIHUAHUA” – 0.97
“MUFFIN” – 0.15











“EIAGLSPYER”

Model-driven vs. data-driven

In recent years **deep learning**[1] models have vastly outperformed previous state-of-the-art solutions in a variety of domains, such as:

- Image recognition[2], including biomedical image processing[3,4]
- Speech recognition[5]
- Natural language processing[6]
- Artificial intelligence[7,8]
- …

Whereas deep learning is becoming more and more widely used to solve various biomedical challenges, so far it has not been applied to the analysis of mass spectrometry data yet.

Nevertheless, there seems a real **potential** for the use of deep learning in mass spectrometry proteomics:

- Current computational techniques to analyze mass spectrometry data are largely **model-driven** based on carefully established expert knowledge
- Only a **minority of the spectral data** can typically be **correctly interpreted** in this fashion, indicating important room for improvement
- There is a large amount of potential **training data** available in public data repositories

Striking **analogies** can be drawn between the current state of mass spectrometry data analysis and other scientific domains where deep learning has already had a major impact.
In a similar fashion as has happened in these other domains, can **data-driven** deep learning techniques radically transform mass spectrometry data analysis?

References

1. LeCun, Y. *et al. Nature* **521,** 436–444 (2015).
2. Krizhevsky, A., *et al. Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097–1105 (2012).
3. Esteva, A. *et al. Nature* **542,** 115–118 (2017).
4. Gulshan, V. *et al. JAMA* **316,** 2402 (2016).
5. Graves, A., *et al. arXiv:1303.5778 [cs]* (2013).
6. Wu, Y. *et al. arXiv:1609.08144 [cs]* (2016).
7. Mnih, V. *et al. Nature* **518,** 529–533 (2015).
8. Silver, D. *et al. Nature* **529,** 484–489 (2016).