Deep learning for sparse and noisy experimental data

Intellegens

Tom Whitehead
Gareth Conduit

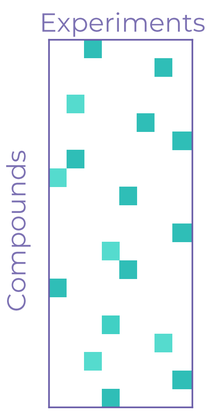
Ben Irwin
Peter Hunt
Matthew Segall

optibrium

Experimental data

Experimental data are complex. No experimentalist has ever run every experiment they could imagine, resulting in **sparse** data availability. Biological data are also **noisy**, as running the same experiment twice gives different results. Drug discovery data can also be **weakly defined**: running the same assay in two different groups can give systematically different results, so is this one assay or two?

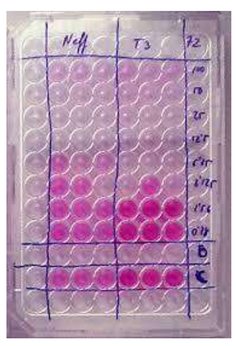
Sparse



Noisy

pIC_{50}=6.5±0.5

Weakly defined

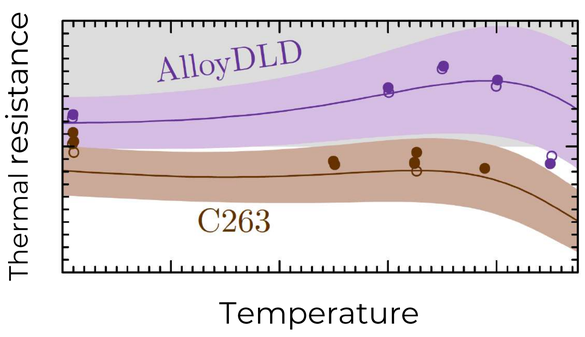


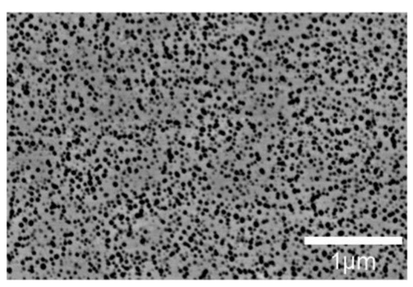
Alchemite™ engine

Intellegens is a spin-out from the University of Cambridge, and specialises in applying deep learning to sparse and noisy experimental data using our proprietary Alchemite™ engine. We combine all accessible data sources to **extend horizons**, learning between endpoints to capture all available information. By focussing on the **most confident** predictions we are able to increase the utility of the results, rejecting low-likelihood outcomes. **Flexible**, user-defined relationships between variables overcome weak definitions. Alchemite™ is available now in collaborative projects: see **https://intellegens.ai**.

Material design

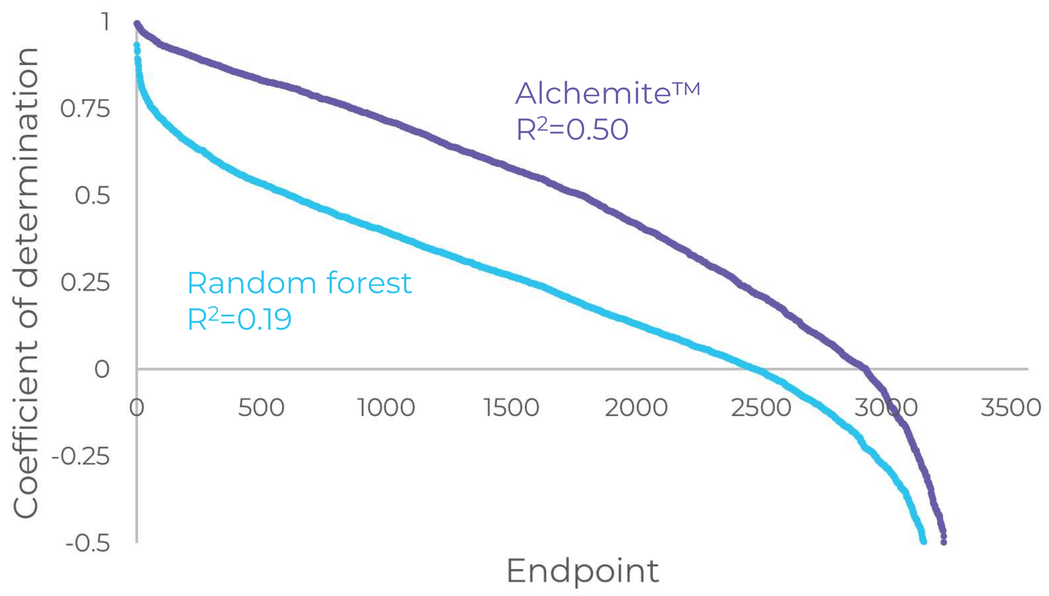
Intellegens' Alchemite™ engine has been used to design new high performance alloys [1], metals for 3D printing [2], and industrial lubricants [3]. Below are a predicted and **experimentally verified** property for an alloy designed for 3D printing, and the alloy itself.



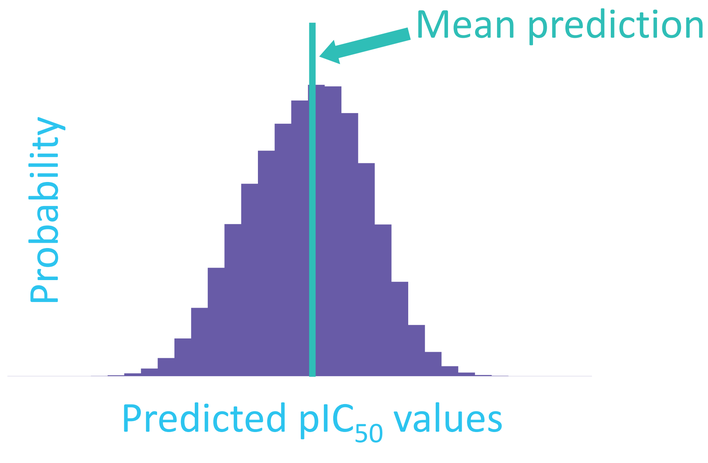


Drug discovery

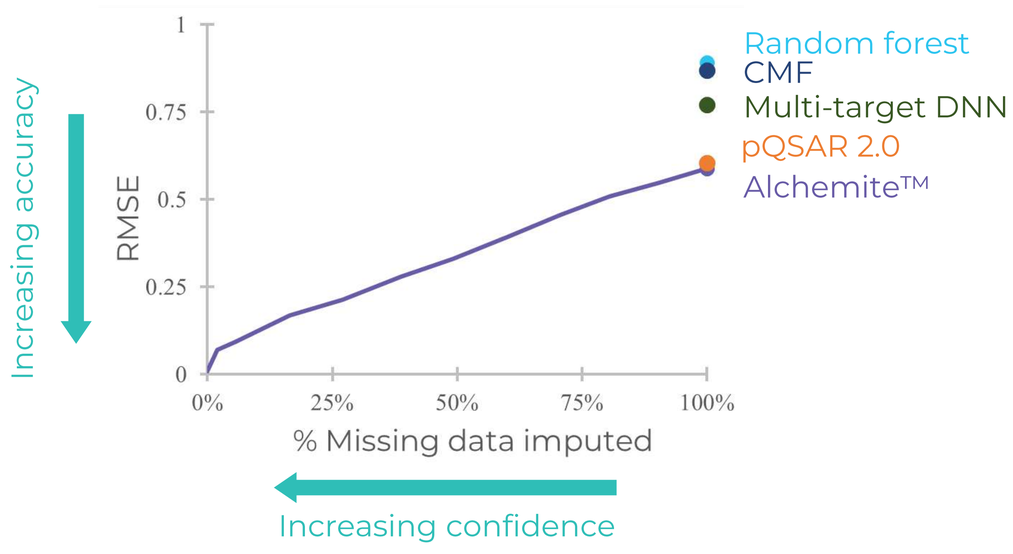
Intellegens' Alchemite™ engine has been used to impute missing experimental measurements [4] in active drug discovery projects. In application to a big pharma company's corporate collection, Alchemite™ significantly **outperformed** a leading quantitative structure-activity relationship model.



Each prediction made by Alchemite™ comes with an associated probability distribution to gauge **confidence** in the results.



By focussing on just the most confident predictions, the accuracy and **utility** of the resulting predictions increases.



References

[1] Conduit *et al.*, Materials & Design **131**, 358 (2017).
[2] Conduit *et al.*, Materials & Design **147**, 107644 (2019).
[3] Santak *et al.*, accepted by Fluid Ph. Equilibria.
[4] Whitehead *et al.*, JCIM **59**, 1197 (2019).