

# Imputing Compound Activities Based on Sparse and Noisy Data

Written by Matt Segall

Monday, 08 April 2019 10:35 - Last Updated Monday, 29 April 2019 11:59

---

Presented by Matt Segall at ACS 2019, Orlando, Florida

Thomas Whitehead†, Matthew Segall\*, Benedict Irwin\*, Peter Hunt\*, Gareth Conduit†  
(\*Optibrium Ltd., †Intellegens)

## Presentation

New results show the increase in accuracy by focussing on the most confident results as a reduction in RMSE, instead of increase in  $R^2$ , following feedback from earlier presentations; and we also illustrate the application of the Alchemite™ model to virtual compounds, i.e. based only on molecular descriptors. This shows it is equivalent in performance to a conventional multi-target DNN, but also retains the ability to focus the most accurate results based on the confidence in the model predictions.

Learn more about Alchemite, a novel deep learning algorithm. Unlike many deep learning methods, this approach is capable of being trained using sparse and variable input data, typical of those available in drug discovery. This enables Alchemite to learn from correlations between experimental endpoints, as well as between molecular descriptors and protein activities, to more accurately impute the missing activities.



You can download the presentation slides as a [PDF](#)