

SPECIAL REPORT

This article was published in Future Drug Discovery : Vol.2, NO.2

Imputation Versus Prediction: Applications in Machine Learning for Drug Discovery Benedict W J Irwin, Samar Mahmoud, Thomas M Whitehead, Gareth J Conduit & Matthew D Segall

Synopsis

Imputation is a powerful statistical method that is distinct from the predictive modelling techniques more commonly used in drug discovery. Imputation uses sparse experimental data in an incomplete dataset to predict missing values by leveraging correlations between experimental assays. This contrasts with quantitative structure–activity relationship methods that use only descriptor – assay correlations. We summarize three recent imputation strategies – heterogeneous deep imputation, assay profile methods and matrix factorization – and compare these with quantitative structure–activity relationship methods, including deep learning, in drug discovery settings. We comment on the value added by imputation methods when used in an ongoing project and find that imputation produces stronger models, earlier in the project, over activity and absorption, distribution, metabolism and elimination end points.



You can link to Future Drug Discovery [here](#)