

Supplementary Materials

A Machine Learning Tool to Predict the Antibacterial Capacity of Nanoparticles

Mahsa Mirzaei ¹, Irini Furxhi ^{1,2,*}, Finbarr Murphy ^{1,2} and Martin Mullins ¹

¹ Department of Accounting and Finance, Kemmy Business School, University of Limerick, V94PH93 Limerick, Ireland; mahsa.mirzaei@ul.ie (M.M.); finbarr.murphy@ul.ie (F.M.); martin.mullins@ul.ie (M.M.)

² Transgero Limited, Cullinagh, Newcastle West, Co., V42V384 Limerick, Ireland

* Correspondence: irini.furxhi@ul.ie; Tel.: +353-85-106-9771

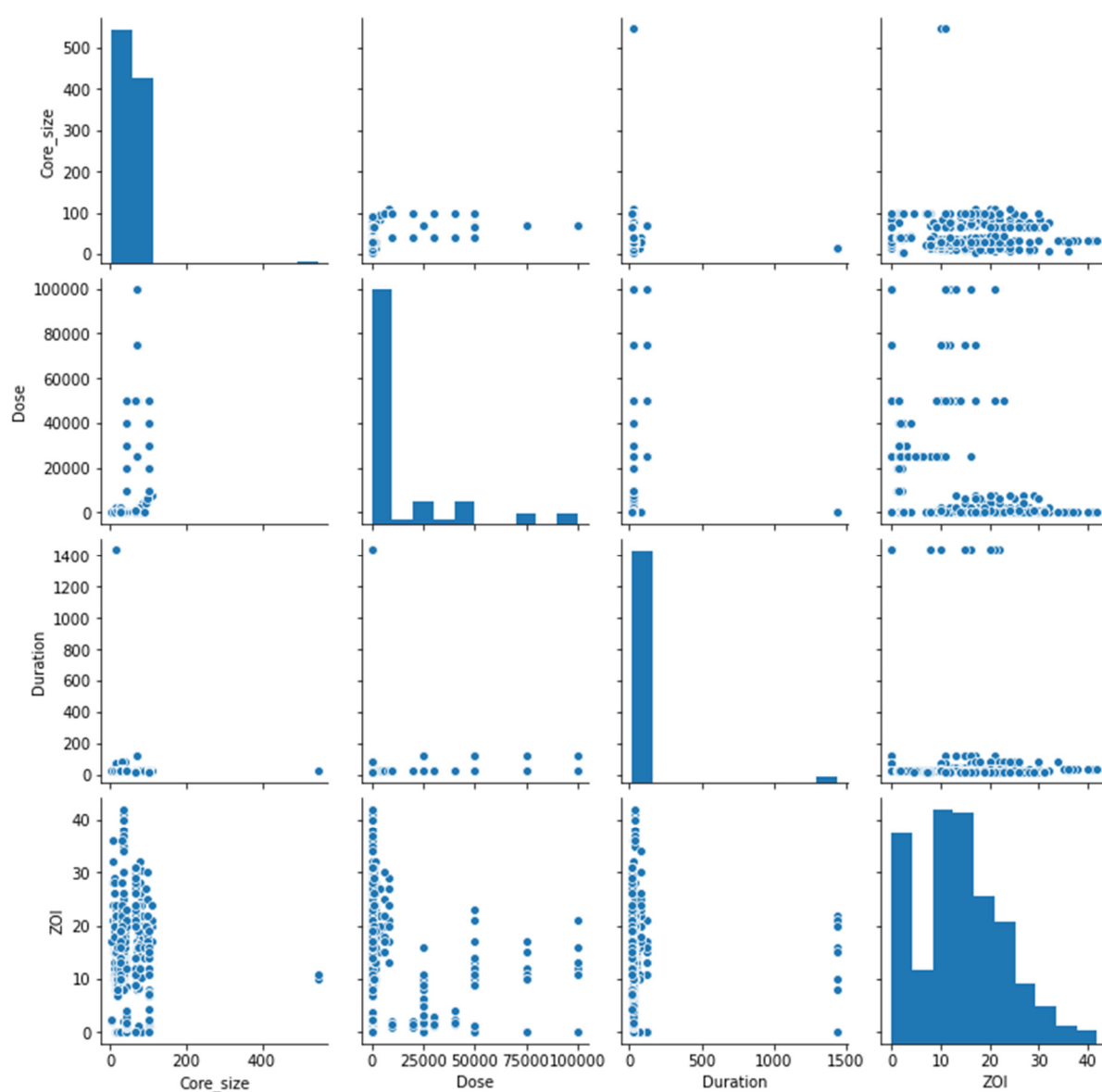


Figure S1. Seaborn pairplot correlation analysis of input variables with the outcome.

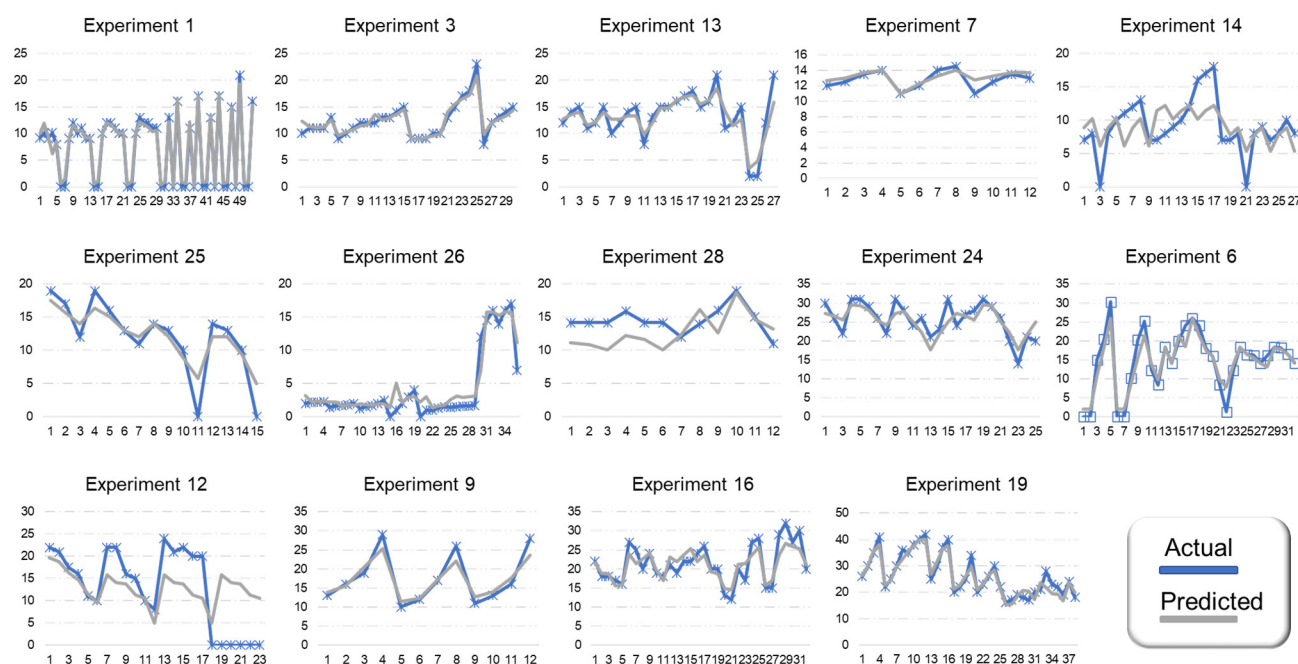


Figure S2. Batch/experimental effect in modelling training. Each experiment signifies one study. The x axis demonstrates the data points that each study provided. The y axis demonstrates the actual and predicted zone of inhibition values. Experiment with Data points >10.

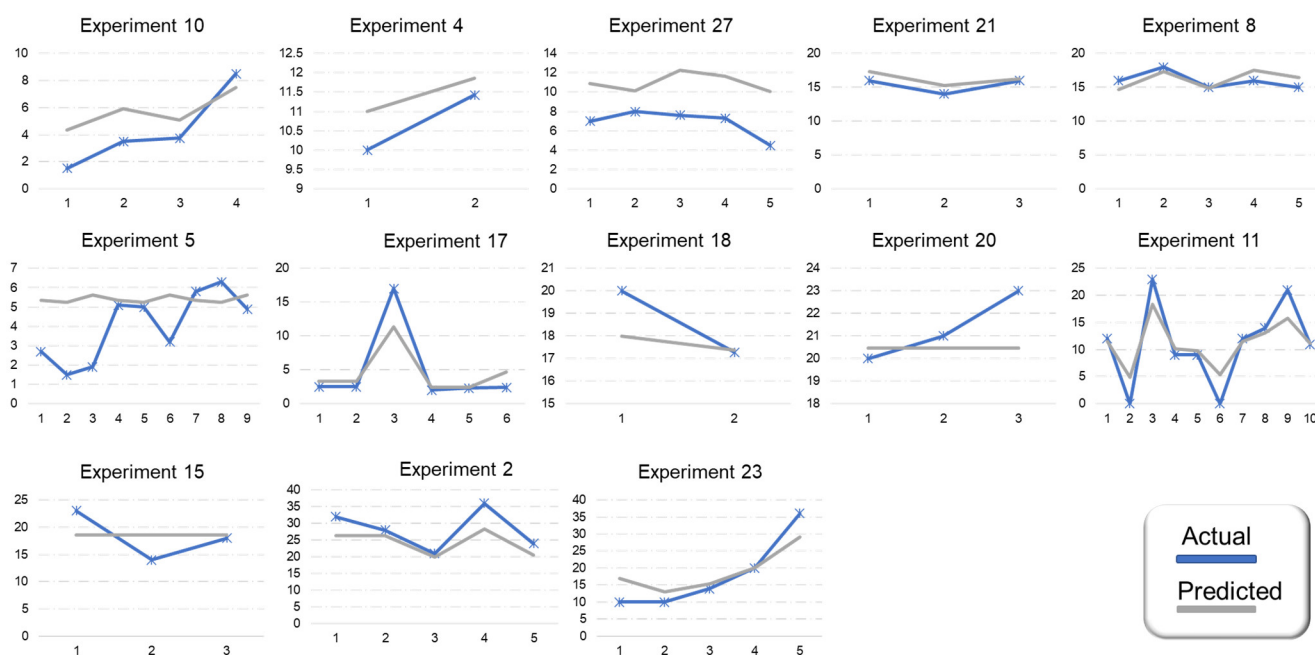


Figure S3. Batch/experimental effect in modelling training. Each experiment signifies one study. The x axis demonstrates the data points that each study provided. The y axis demonstrates the actual and predicted zone of inhibition values. Experiment with Data points <10.

It is evident from Figure S2 and Figure S3 that experiments (studies) that provided few instances (data points <10) are not captured well by the random forest model. However, this demonstrates a realistic scenario where instances are taken as in the case of integrated databases. We did not stratified the dataset artificially in order to demonstrate the importance of the data volume provided by individual studies. However, we could have include datapoints from each experiment into the training model to generalize it.