Five points to consider when reading a translational machine learning paper **Supplementary Materials**

Supplementary Box: A example to outline the steps in a machine learning study

Here we outline an example of a simple machine learning study that uses supervised learning. For this, let us assume we have a dataset of people who have undergone CBT, of who, half the people have responded. We have the baseline HAMD scores, collected prior to them receiving the intervention. The aim of the study is to see if we can find an algorithm that will correctly predict who will respond (output/label) to the intervention, purely based on baseline individual items scores on HAMD (features).  We decide to use a regularized form of logistic regression (supervised learning algorithm) to find the weights (parameters) using iterative processes (i.e., learning) that best maps the relationship between the features (HAMD) and output (Responder vs. non-responder) based on a criterion chosen by the experimenter (e.g., accuracy) - this is the model.  The set of features and the associated parameters obtained from the above exercise are then used to test the performance of the above model and any additional steps on completely unseen/independent dataset (testing / validating).  The accuracy of the predicted output can then be tested against the actual output, using several measures including accuracy, predictive values, etc. The initial model can be further refined using calibration techniques, and decision curve analysis can be performed to see if implementation is feasible in the real world (3).



Figure 1: A simplified representation of a basic supervised learning algorithm. Usual pipelines are more complex and may involve a number of additional steps, including feature extraction and selection. In addition, features extracted from unsupervised and reinforcement learning methods can be used along with supervised learning techniques to assist prediction/classification. Blue arrows constitute the training phase, and green arrows represent the testing/validation phase. The yellow arrow represents how the model developed from the training phase is utilized in the testing set that involves unseen data. Briefly, the goal of this computation would be to classify responders vs. Non responders based on HAMD scores. We implemented the computation using a logistic regression. We may physically realise this, by implementing the algorithm within a primary care clinic – using data collected from corenet (similar to a QRISK).