# Appendix

## Supplementary Section 1

In the main text, Figure 7 shows only algorithms that were used more than once. The remaining algorithms that were used only once in the literature were spectral clustering, point completion network, generative adversarial networks, linear discriminant quadratic discriminant classifier, uncorrelated normal density-based classifier, polynomial classifier, logistic classifier, Parzen classifier, conditional random fields, hidden Markov model (manually constructed), hidden Markov model (automatic), hybrid clustering, symbolic regression, grammatical evolution, generalized regression, ensemble method, non-linear regression, deep neural network, and deep belief network.

## Supplementary Section 2

Of the 18 occasions where multiple algorithms were used and could be compared, RF performed best nine times. This is not surprising since most problems were classification tasks, however RF is particularly strong in these tasks, and the fact that it performed best on 50% of these occasions is reflective of this. However, it must be borne in mind that the number of shallow ML usage greatly outweighed DL. The same is true for the higher proportion of classification tasks, leading to an overrepresentation of classification algorithm. Comparisons across papers are also not standardized and are also situation dependent. Nonetheless, the dominance of RF algorithms, when compared against other algorithms is in line with RF’s reputation for performing well in tasks of classification. Similarly, SVM is also a higher performer in these tasks.

Supplementary Figure 1. The number of times an algorithm performs best in any situation where comparison with another algorithm or algorithms was possible.

## Supplementary Section 3

In the main text, Figure 13 shows all evaluation metrics that used more than once. the remaining evaluation metrics that were used only one were misclassification rate, model loss, combined cluster purity, Gini coefficient, standard error, mean squared error, Calinski-Harabasz score, median absolute error.