Metagenomics

Taxonomic analysis. User takes meta-genomic read, feeds it through a NN to classify into taxa (orders).

Training data contains ~9,000 DNA sequences, with features containing all the possible up-to-5-length combinations of nucleotides.

The biggest problem is that most of the classes only have one example in training data. Current ideas include over-sampling or excising underrepresented classes to improve classification.