The first thought I had when looking at the default code was that the 50th percentile gray-value feature was useful but not terribly discriminatory. We are probably more interested in the extremes of that data. So, using basic guess-and-check, I changed the percentile to the 10th. The second thing I did was evaluate the existing “greycoprops” function. I first switched the algorithm to “dissimilarity” and got an improved result. However, I noticed when looking at the documentation, that the “contrast” algorithm using the same equation but with the differences amplified and positive (because of the squared difference). Switching to this algorithm gave me an even better result. Next, I thought about exploiting the fact that we know what the false positives are. I realized later that this was sort of like “cheating” because we know which are false positives from metadata in the test data set—meta data that would not exist in the real world. However, I thought this would be a good simulation of human-aided machine learning. Verifying even some of the synapses via humans post-analysis can make the next run of the program over test data much more accurate. My hypothesis proved to be correct as feeding into the program one more feature with a false positive correlation matrix increased the accuracy visibly.

Note: In order to test the human correction simulation, you must “Run Cells Below” cell 4.