I know this isn’t the most reproducible way to record this, but my Python and Jupyter notebook game is weak compared to R.

* Original, peer-reviewed research
  + Accuracy: 83.9%
  + Training set size: 1204
  + Testing set size: 301

Non-human classifier: 93.4%

Accuracy scores: [0.9375 0.9375 0.94270833 0.92708333 0.91906005 0.94270833

0.93229167 0.93229167 0.94010417 0.92950392]

Training set size: 1535

Testing set size: 384