The overall structure of the program and the readme file is friendly toward reproducibility. There are however a few issues some small some critical. The authors use absolute vector length number such as 54996 in the index, which should have been a variable. User would need to copy the generated plots to a separate folder, which could have been done automatically. Also, some plots are not generated automatically (SVM plot). For the SVM job that runs on cluster, the paths was absolute at a couple places, authors could have passed in an environment variable. Also, an estimated waiting time for cluster job would have been nice.

One major issue is the table in the write up contain all copied numbers, which make it difficult to reproduce. As such, we were not able to compare the model performance between new set of images and old set. Looking at the plots, things are not too much different. For example, the variable importance according to random forest are still the same. This make sense because out of bag were chosen randomly row-wise, not block-wise. So for the new images and old images, results should be the same.