For the random forest analysis, these are the arguments that were used that are different than the default values in the R random forest package (R version 4.0.5):

1. Replace = F (we sample *without* replacement)
2. Sampsize = c(n, n) where n = 0.5*sample size of the smaller group
3. Ntrees = 50,000
4. Mtry = default value = (sqrt(p))
5. These values were based on historical use and validation, so were not optimized based on this particular data set which would cause bias
6. The mean decrease accuracy measure of importance was used rather than the GINI index