|  |  |
| --- | --- |
| **Criteria** | **Points (0-10)** |
| **Content** | 10 |
| **Subject Knowledge** | 9 |
|  | 19/20 **Excellent job!** |

**General Feedback**

**Content**

* Code is complete but a little messy in the data management section. There are some areas of redundancy where things like the pipe could be used to make it cleaner, and comments should be added within a number of chunks to add detail on what is going on.
* Great use of functions to remove redundancy in the model training and results formatting!
* Can’t run on my CPU due to the data not being available, but I’m guessing that is due to data access restrictions, no problem at all.
* All figures, tables, analyses referenced in documents appear to be in code and clearly marked

**Subject Knowledge**

* Great background section, nicely details the necessity for a machine learning algo for prediction. I honestly got a little lost in the specifics with the acronyms and medical terminology, but I got the main gist, and I know you are setting this up for a publication in a medical journal, so explaining all of these terms is probably redundant for the journal you are targeting.
* Excellent sample description and detail of inclusion criteria. Would like more info on the sampling mechanism done to get the sample (as I see it is an old, mostly White sample), though not sure if this was available to you all.
* Great detail on all variables, including predictors and the definition of the outcome. I would like to see a bit more breaking down of some of the terms (the “at least one … adenoma” makes it seem like multiple polyps were available for a patient?) though again this may be more just ignorance of the medical background on my part.
* Great description of training and testing, and detail of tuning (though how many folds did you use in your tuning CV? Should mentioned in the text). Also, at what stage was SMOTE used? Did you only SMOTE the training set, or did you SMOTE the whole data followed by a 60:40 split? This is **critical** to have unbiased results. Looks like you did SMOTE with the training set only from your code, which is perfect, just detail this in the text as well.
* I have a question about optimizing the tuning parameters for sensitivity. In theory, if you’re trying to just maximize sensitivity, a tuning parameter with 100% sensitivity but 0% specificity would be “best” unless you also consider specificity. Would like more detail on this.
* **GREAT** use of a customized rule for creating a probability threshold versus using something like Youden’s index which is broad.
* In the results section, it says repeated CV was used for tuning. Again, would like more specifics on the CV tuning process used. If you tried to optimize sensitivity, I’m a bit surprised your RF had such a low training set sensitivity vs specificity? Also “XXX percent” seen, I think you forgot to add in the number.
* Great conclusion/discussion section. Would like some discussion on the limitations as far as what population your results pertain to given the mostly White sample. Also, I see “XXX percent” again