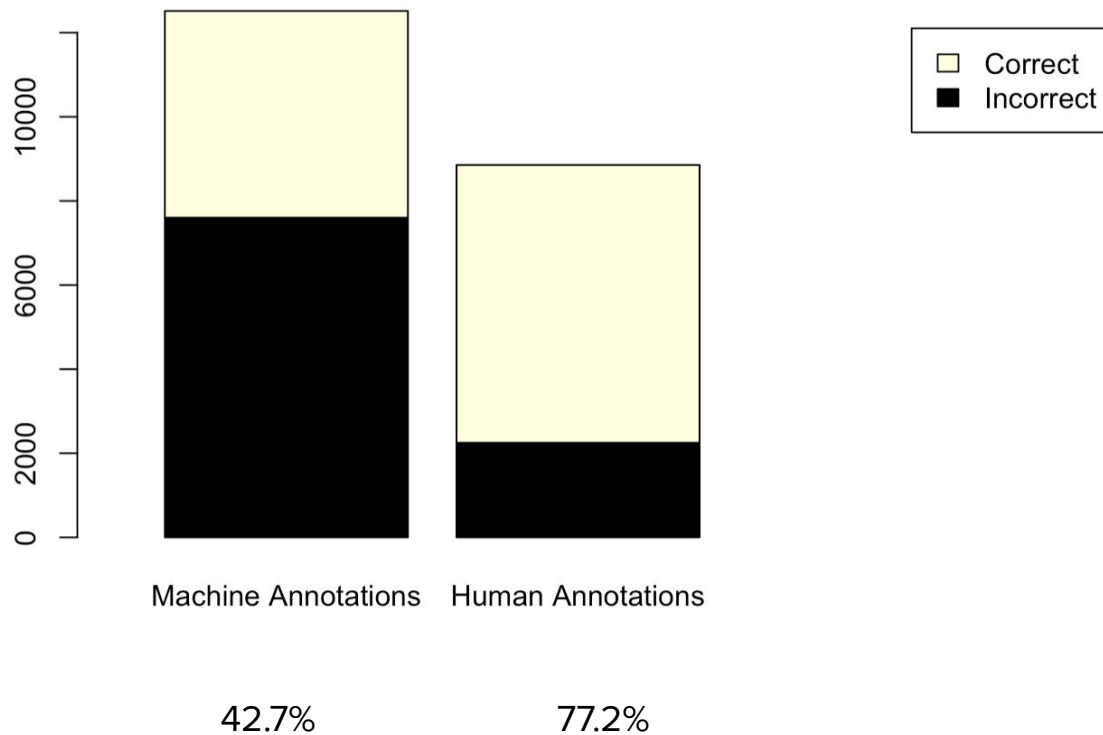# Nodule Validation Automation

Dorothy(Huiwen) Duan
10/29/2019

The Validation Model

- A random forest model
- Response variable: GT
- Predicting variables: a combination of nodule markings and series level information
- Left out observations that are annotated by both machine and human

Response Variable: GT

Predictors - Nodule Markings

1. radius_mm: numerical, radius of a marking in mm
2. user_id: categorical variable. NAs(n=11520) means machine, so replace NA with 0.

Predictors - Series Level

1. series_id: series refers to a set of 3D medical imagery that belongs to a CT scan, and each series is uniquely identified by its series_id.
   - I converted it to dummy variables in my model.
   - GT scores vary extensively across different CT series. So future direction needs to consider a mixture model.

Predictors - Series Level
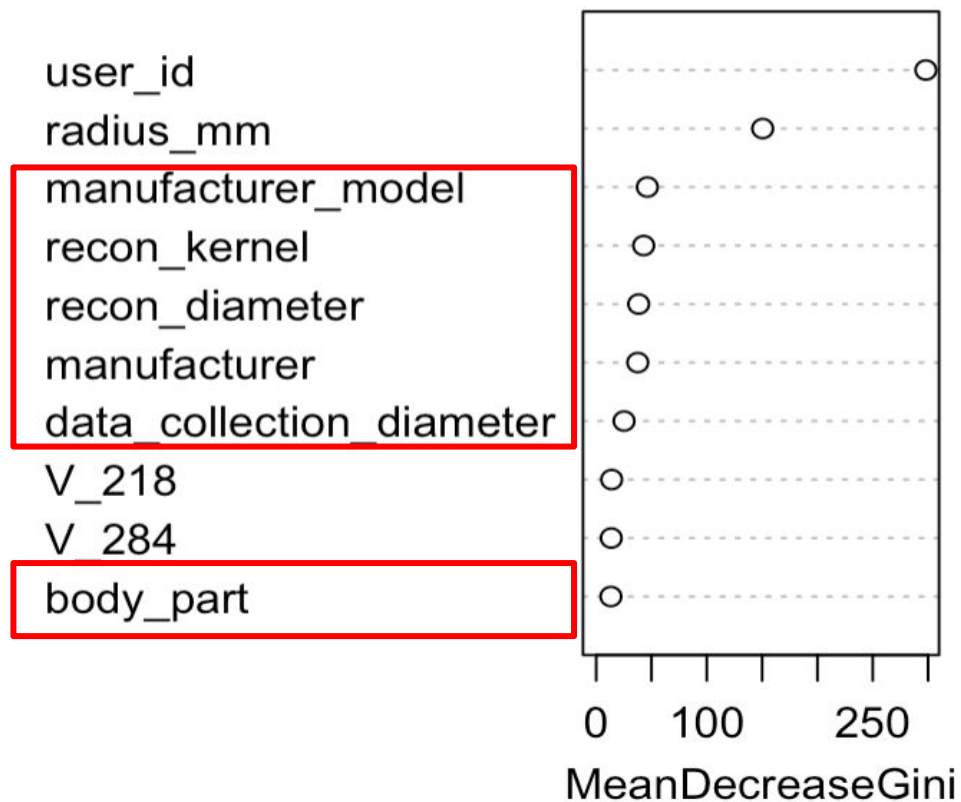
Predictors - Series Level Predictor Importance

2. data_collection_diameter: numerical with **2** levels

3. manufacturer: categorical with **4** levels

4. kvp: categorical with **2** levels, peak kilo voltage output of the x-ray generator used

5. manufacturer_model: categorical with **7** levels

6. recon_kernel: categorical with **9** levels, the algorithm used to build the 3D image

7. body_part: categorical with **2** levels, chest versus other. change nas to other

8. recon_diameter: numerical variable, diameter in mm of the region from within which data were used in creating the reconstruction of the image

9. image_type: categorical with **3** levels

Performance metric: Positive Predictive Value

- Positive predictive value: when the real validation is positive, how likely the model can make correct prediction.

- In our situation, positive validation means the nodule indeed exists. Failure to identify an existing nodule is detrimental.

91.7%

Positive predictive value: when the real validation is positive, how likely the model can make correct prediction.

Future direction: mixture model

- My validation model identify multiple series_id as important predictors, which means that observations are not totally independent. Because they come from different CT series, there exist random effects.

- In the future, I will develop mixture models to further investigate the random effects.
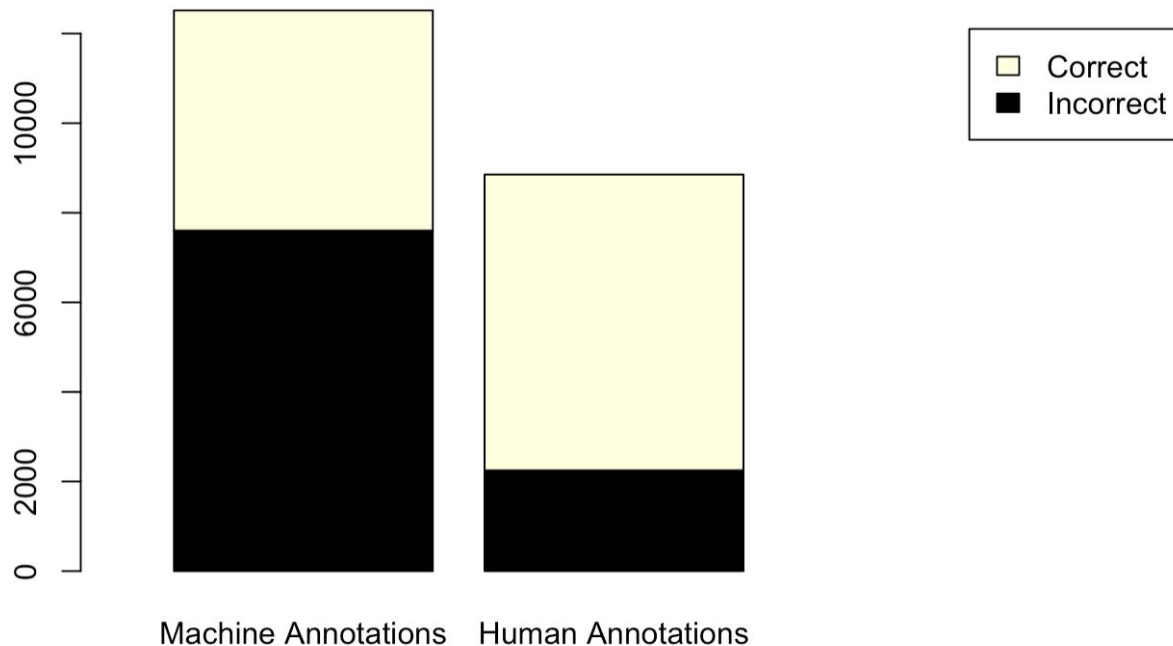
Future direction: CT values

- Incorporate average brightness within and surrounding the nodule annotation area

- Deep learning models

Investigate the Annotation Model

851 observations have both human and machine annotated

Compared to human annotators, the annotation model has much fewer true positive(GT)s.

Testing the validation model on the 851 common annotations between human and machine

Machine Test Data:

Accuracy : 81.9%

True Positive Rate : 91.7%

Pos Pred Value : 87.9%

Human Test Data:

Accuracy :90.4%

True Positive Rate : 92.2%

Pos Pred Value : 97.8%

What is the effect of model prediction score?
What action do you recommend to the model team?

Prob_fpr:
true: 1
False: 11517
NAs: 9852

Prob_rpn
FALSE:11508
TRUE :10
NA's :9852

Nodule_score
FALSE:11511
TRUE :7
 NA's :9852
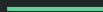
Recommendation: investigate the huge imbalance of those model prediction scores.

Follow-ups & others

1. Coordinates
2. Radius
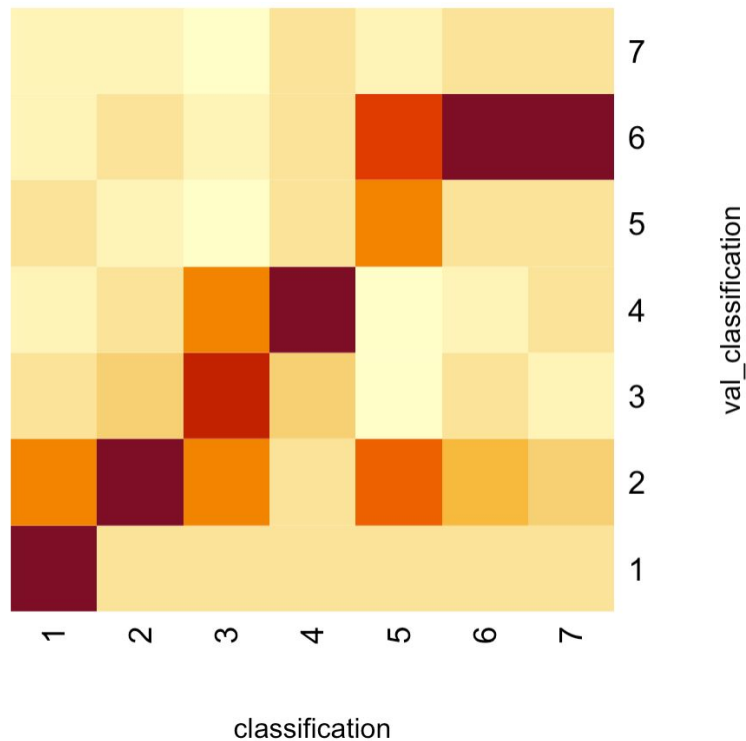3. classification-Val_classification

Coordinates:

- Coordinates: x,y,z numerical, not consistent across CT series, so not included in the current validation model
- Recommendation: scale them to have the same coordinate system
- Coordinates information can also provide insights regarding which part of the lung is more vulnerable to nodules
- Interactive plots:

Radius: machines annotate larger radius than human
Human annotations have a larger range in radius
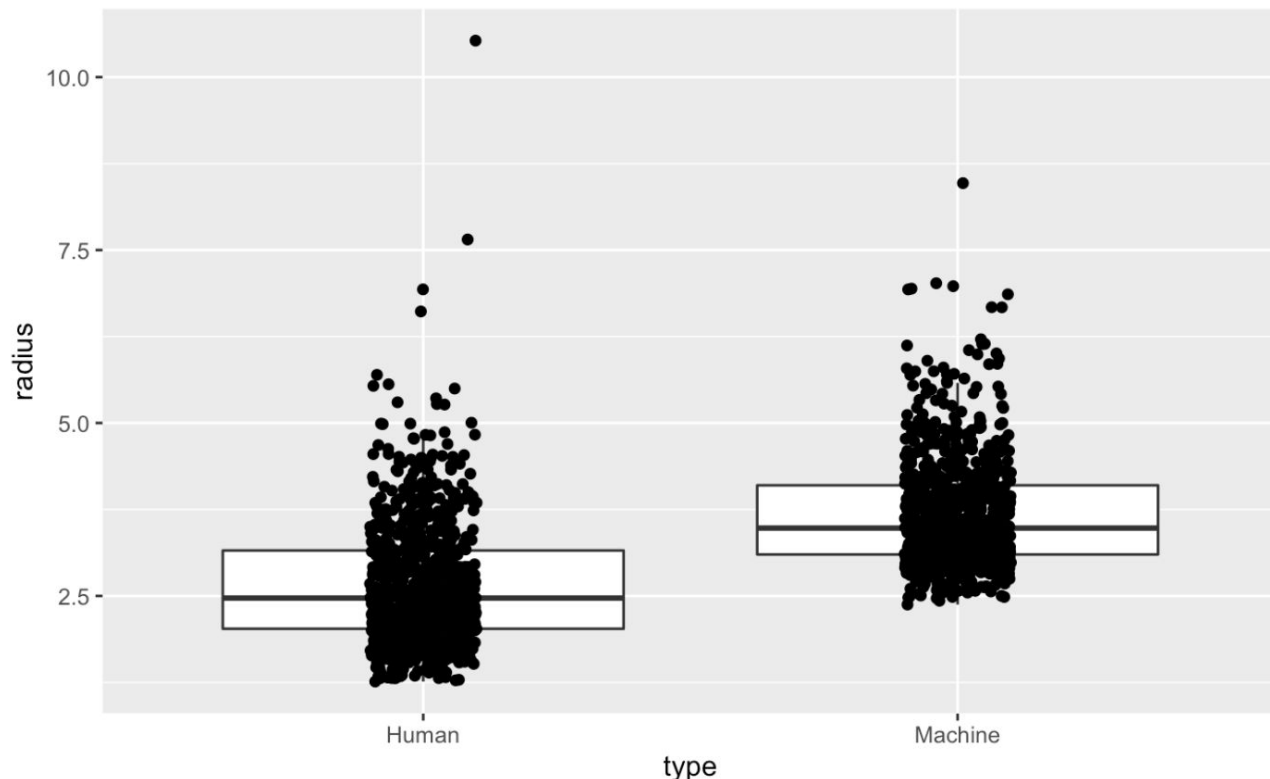
Exploring classification and val_classification

Does consensus on 'classification' and 'val_classification' affect the probability of a nodule marking being a True Positive (GT= True)?

- Val_classification only has value when the GT is true, so it is not viable to explore its relationship to the probability of GT being true. However, I want to explore the relationship between those two classification variables.

- The consensus rate is 61%

- In the future, I would like to make a model to predict the val_classification