MIMIC III Lung and Respiratory Cancer Cohort

Daniel Jaso

Description

Design a dataset from MIMIC III for the use of clinical trials on lung cancer immunotherapeutic drugs. We will include cancer patients within a certain age both genders, all ethnicities, all insurance types, and all ICD-9 diagnosis codes that include variations of commonly assigned lung cancer (162.0, 162.2, 162.3, 162.4, 162.5, 162.8, 162.9, 197.0, 231.2, 23.57, 239.1) We will also include exclusions on comorbidities with Addison's Disease, Thyroiditis, or HIV-AIDS as a measure to exclude uncontrolled responses to new immunotherapies. Furthermore, we will exclude patients that have undergone radiotherapeutic treatments. In addition to this, we took into account laboratory measures with regards to 'Bicarbonate', 'Red Blood Cells', 'White Blood Cells', 'Platelet Count', 'Oxygen Saturation','pH') or LOINC codes 1963-8, 789-8, 804-5, 777-3, 20564-1, 11558-4.

Data Summary

```
dim(lung)
```

```
## [1] 8587   14
```

```
colnames(lung)
```

```
##  [1] "subject_id"      "icd9_list"       "loinc_code"      "min_value"
##  [5] "max_value"       "patient_age"     "gender"          "ethnicity"
##  [9] "insurance"       "los"             "has_chemo"       "has_radio"
## [13] "has_cancer"      "deceased_status"
```

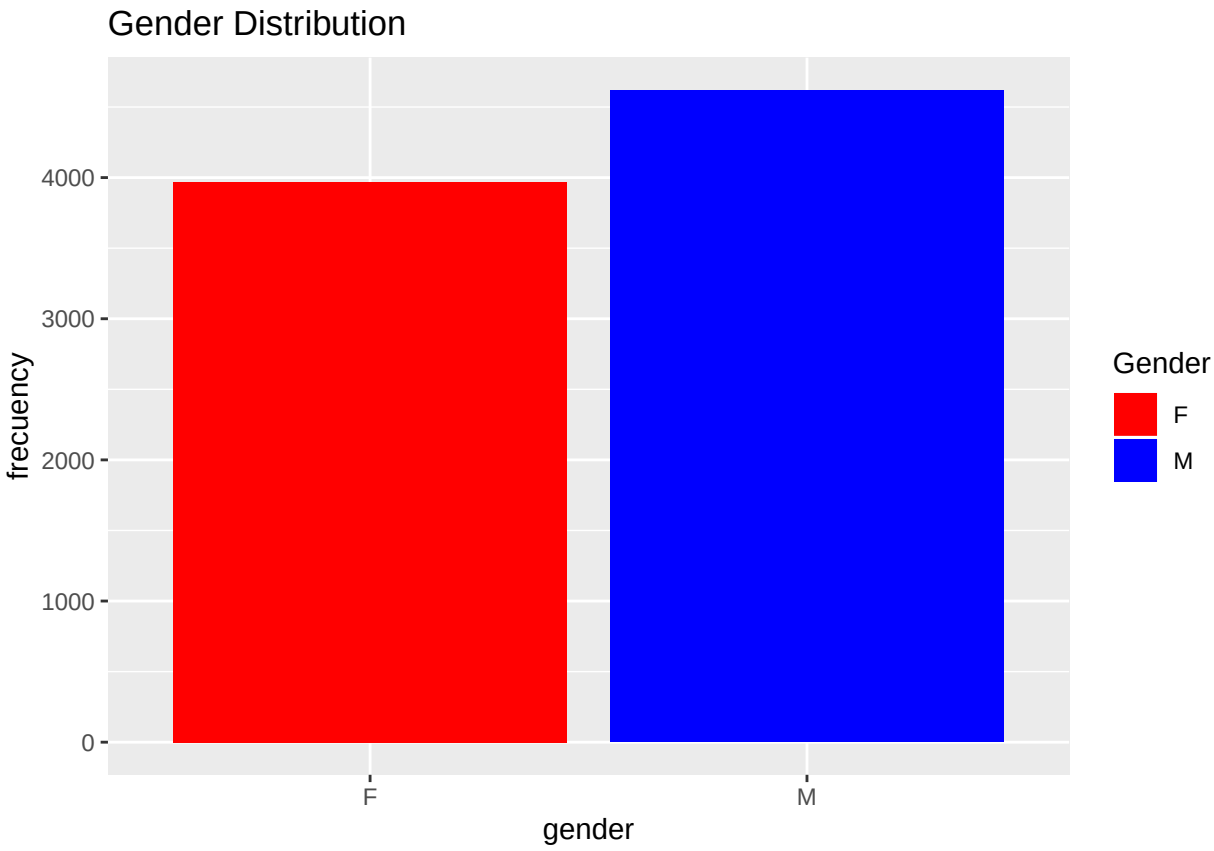
```
summary(lung)
```

```
##    subject_id
##  Min.   :   56
##  1st Qu.:13741
##  Median :28686
##  Mean   :38442
##  3rd Qu.:62802
##  Max.   :99899
##
##                                                        icd9_list
##  1623                                                        :  11
##  0389,V1082,99591,1978,1976,27651,1977,1970,00845            :   5
```

```
##  1,623,440,041,519,450,000,000                                    :   5
##  1,623,584,948,553,080,000,000,000                                :   5
##  1,628,198,541,241,400,000,000,000,000,000,000,000,000,000,000,000:   5
##  1,988,917,141,970,190,000,000                                    :   5
##  (Other)                                                          :8551
##    loinc_code     min_value      max_value     patient_age     gender
##  11558-4:1308   10     : 208   98     : 285   Min.   : 20.00   F:3968
##  1963-8 :1631   22     : 164   99     : 239   1st Qu.: 57.00   M:4619
##  20564-1: 756   21     : 163   9.9    : 227   Median : 66.00
##  777-3  :1631   23     : 151   29     : 182   Mean   : 70.72
##  789-8  :1630   24     : 143   30     : 172   3rd Qu.: 75.00
##  804-5  :1631   20     : 140   9.8    : 169   Max.   :307.00
##                 (Other):7618   (Other):7313
##                   ethnicity         insurance         los
##  WHITE                  :6338   Government: 199   Min.   :  0.000
##  BLACK/AFRICAN AMERICAN: 639   Medicaid  : 698   1st Qu.:  4.000
##  UNKNOWN/NOT SPECIFIED : 618   Medicare  :4738   Median :  7.000
##  ASIAN                  : 257   Private   :2887   Mean   :  9.426
##  OTHER                  : 164   Self Pay  :  65   3rd Qu.: 12.000
##  HISPANIC OR LATINO     : 126                     Max.   :110.000
##  (Other)                : 445
##    has_chemo        has_radio         has_cancer deceased_status
##  Min.   :0.0000   Min.   :0.00000   Min.   :1    Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:1    1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :1    Median :0.0000
##  Mean   :0.0389   Mean   :0.05532   Mean   :1    Mean   :0.2501
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:1    3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1    Max.   :1.0000
##
```

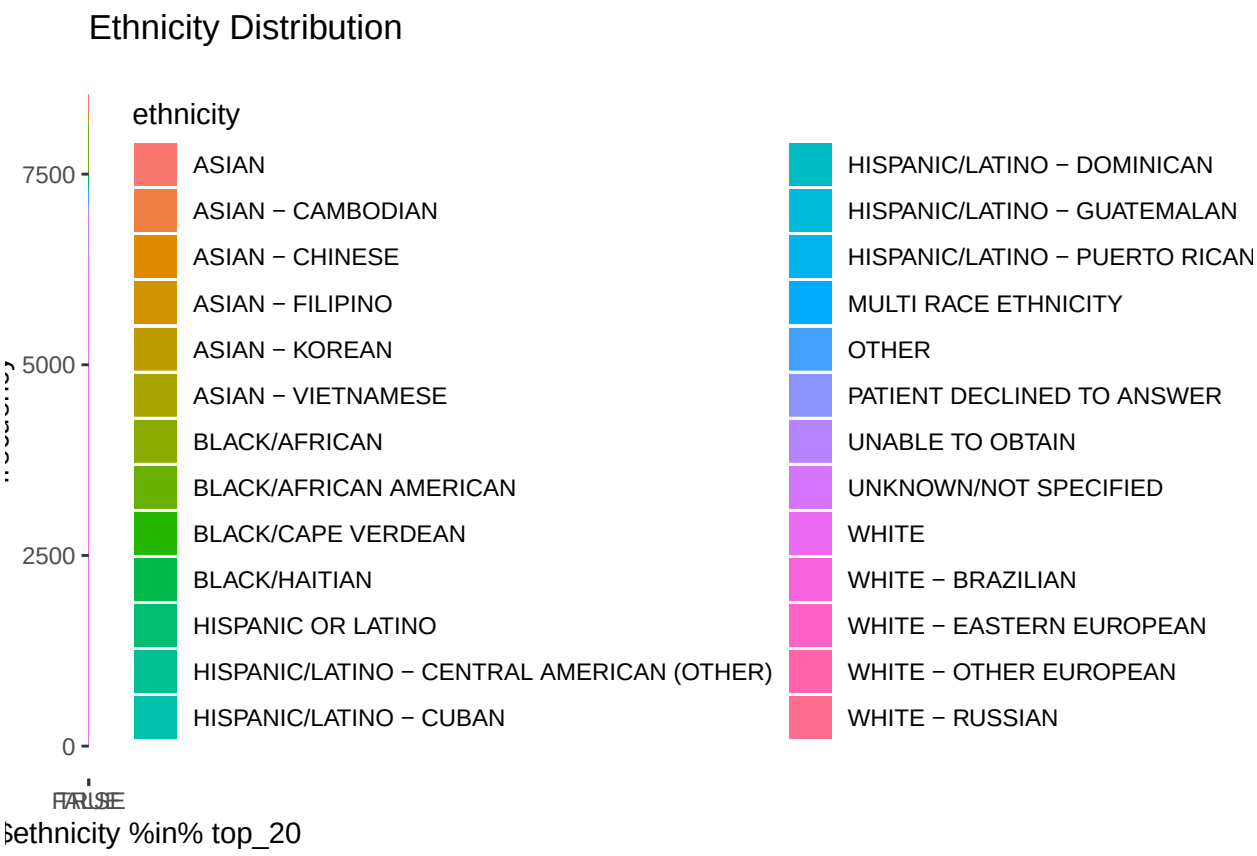
Exploratory Data Analysis (EDA)

Gender Distribution

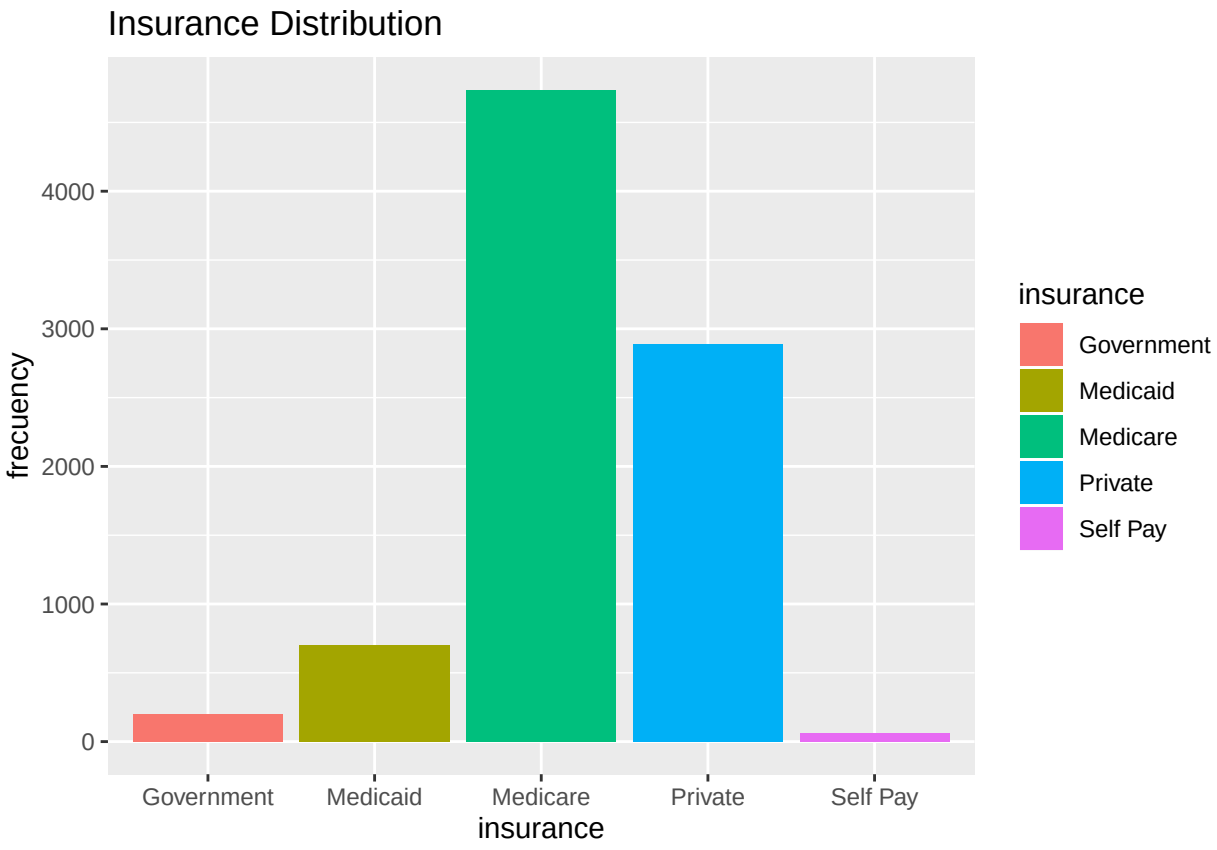


Age Distribution

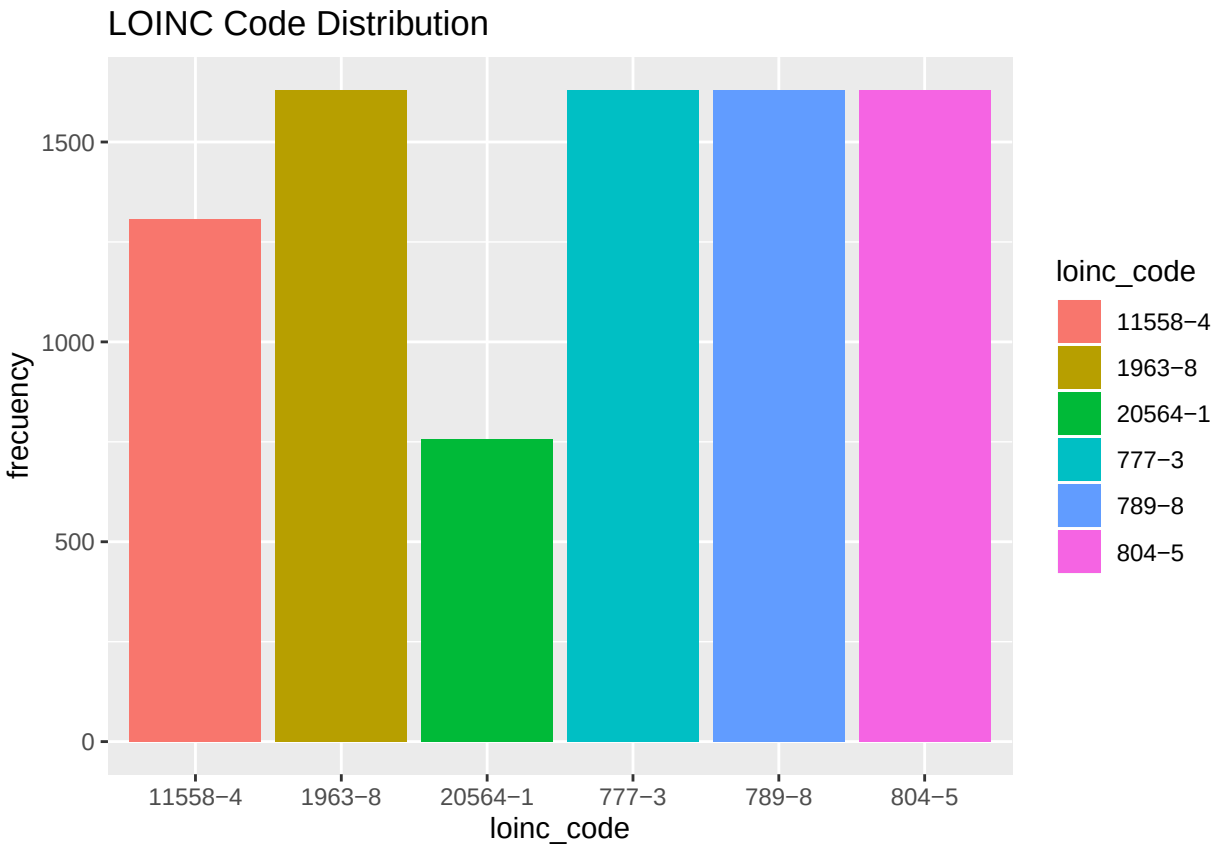
Ethnicity Distribution



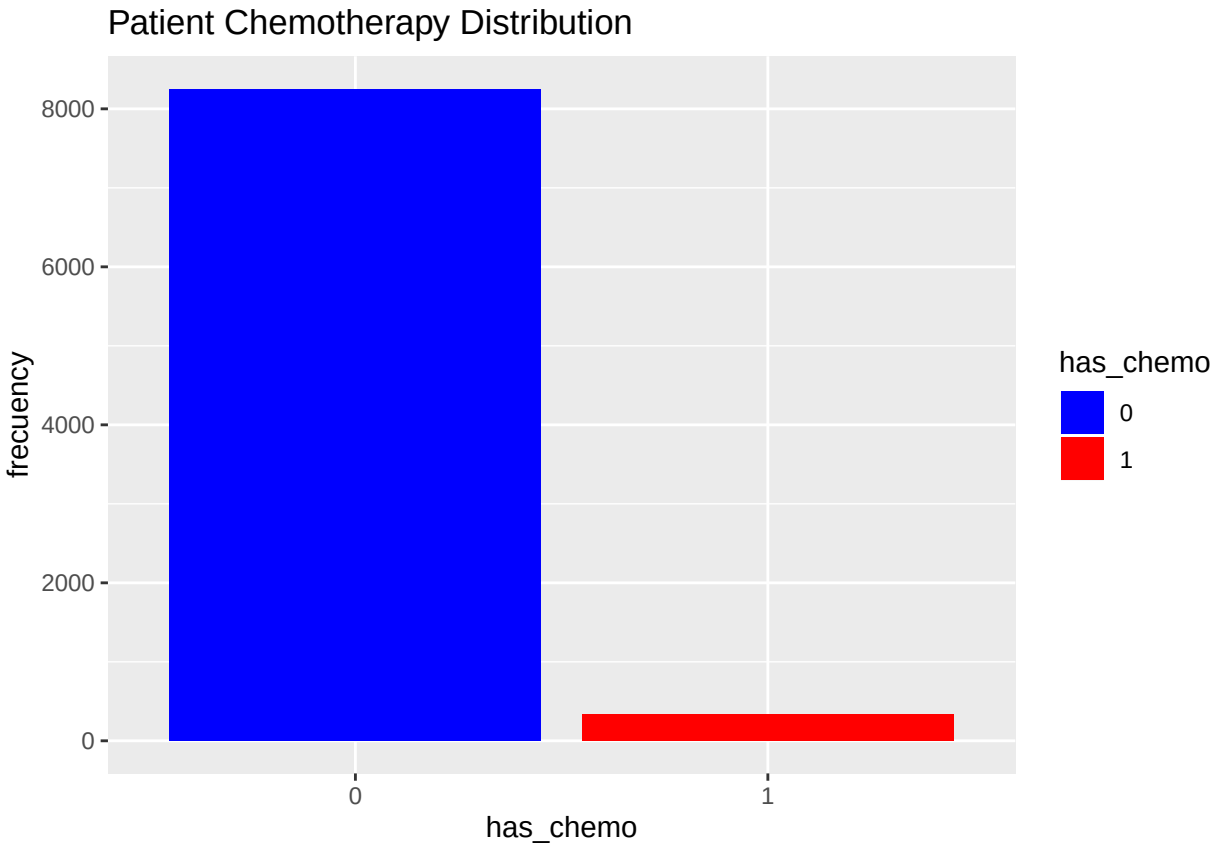
Insurance Distribution



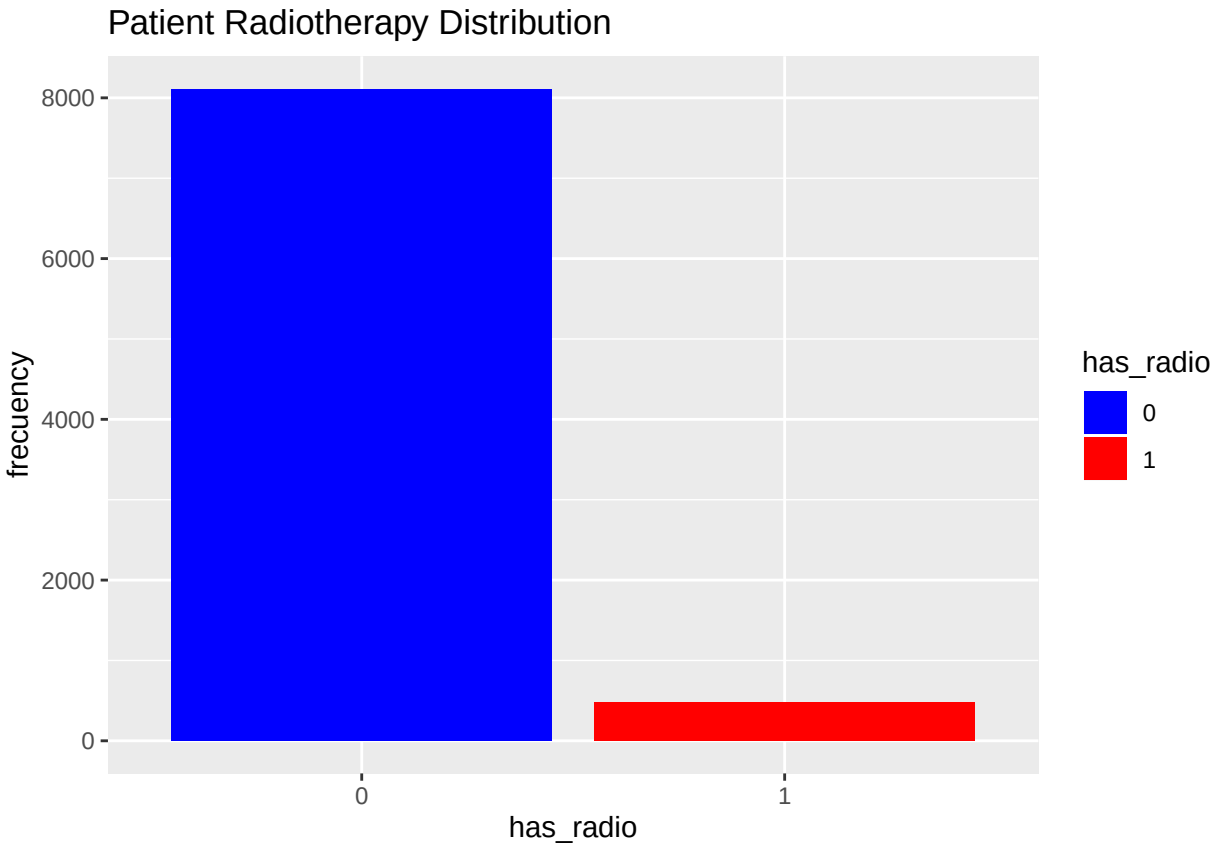
LOINC Code Distribution



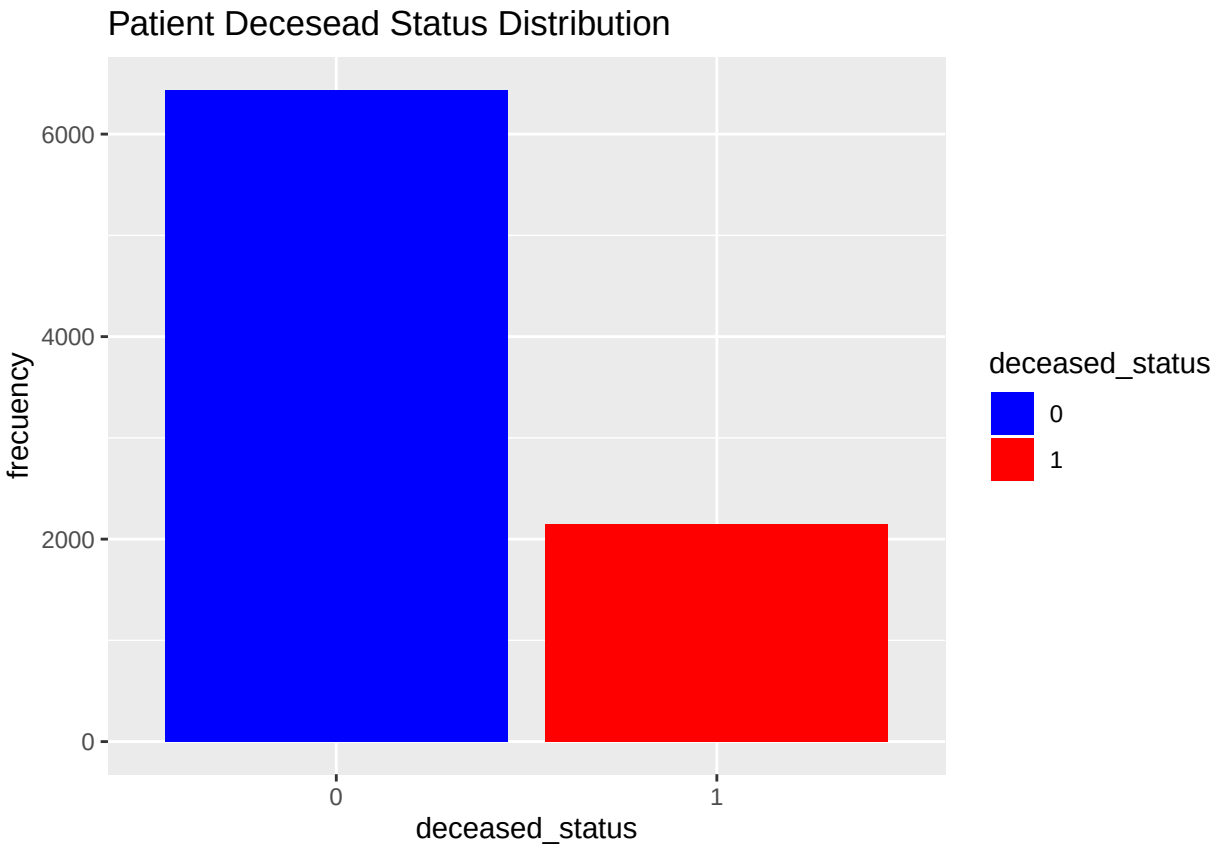
Patient Chemotherapy Distribution



Patient Radiotherapy Distribution



Patient Decesed Status Distribution



Fix imbalance data for deceased_status

```
lung$deceased_status <- as.factor(lung$deceased_status)

new <- SMOTE(form = deceased_status ~ ., data = lung, perc.over = 100)
table(new$deceased_status)
```

```
##
##    0    1
## 4296 4296
```

Train/Test Split (70/30)

```
lung2 <- new[, c(3, 4:7, 9:12,14)]
colnames(lung2)
```

```
##  [1] "loinc_code"      "min_value"       "max_value"       "patient_age"
##  [5] "gender"          "insurance"       "los"             "has_chemo"
##  [9] "has_radio"       "deceased_status"
```

```
lung2$deceased_status <- factor(lung2$deceased_status)


intrain <- createDataPartition(y = lung2$deceased_status, p= 0.7, list = FALSE)
training <- lung2[intrain,]
testing <- lung2[-intrain,]
dim(training); dim(testing)
```

```
## [1] 6016   10
```

```
## [1] 2576   10
```

Model Classifiers

Elastic Net

Elastic Net Parameter Tunning

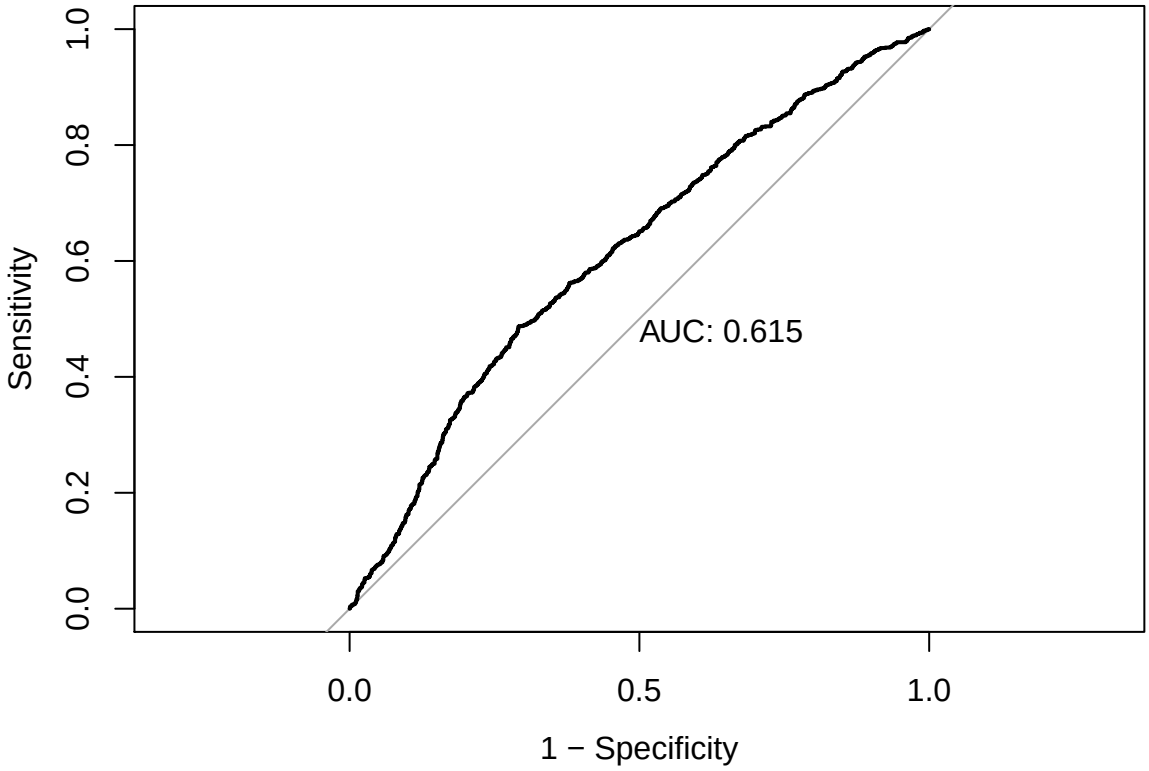
```
elastic_net_model
```

```
## glmnet
##
## 6016 samples
##    9 predictor
##    2 classes: 'F', 'T'
##
## Pre-processing: centered (9), scaled (9)
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 4812, 4814, 4813, 4813, 4812
## Resampling results across tuning parameters:
##
##   alpha      lambda       ROC        Sens       Spec
##   0.1001746  0.029447170  0.6207990  0.5651577  0.5983975
##   0.1602459  0.419952176  0.5784045  0.4136058  0.6799155
##   0.2132285  0.020336795  0.6211282  0.5668205  0.5987297
##   0.3083945  1.040188860  0.5000000  0.4000000  0.6000000
##   0.3463324  0.003976269  0.6203551  0.5608366  0.6017236
##   0.4583346  0.004757222  0.6205724  0.5641611  0.5997286
##   0.4760670  0.015962237  0.6211571  0.5718105  0.5927452
##   0.5107257  0.076005698  0.6134266  0.6040503  0.5611567
##   0.6028125  0.018892947  0.6209453  0.5801234  0.5910841
##   0.6799698  0.001960255  0.6203800  0.5608366  0.6013908
##   0.7053881  0.005317259  0.6209263  0.5664894  0.5980658
##   0.7309996  0.086132647  0.6099786  0.5518322  0.5930581
##   0.7527179  1.907926191  0.5000000  0.4000000  0.6000000
##   0.8807979  0.002958175  0.6207760  0.5654911  0.6000619
##   0.9317473  1.142867676  0.5000000  0.4000000  0.6000000
##
## ROC was used to select the optimal model using the largest value.
## The final values used for the model were alpha = 0.476067 and lambda
##  = 0.01596224.
```

Elastic Net Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   F   T
##          F 719 569
##          T 512 776
##
##                Accuracy : 0.5804
##                  95% CI : (0.561, 0.5995)
##     No Information Rate : 0.5221
##     P-Value [Acc > NIR] : 1.676e-09
##
##                   Kappa : 0.1607
##
##  Mcnemar's Test P-Value : 0.08852
##
##             Sensitivity : 0.5770
##             Specificity : 0.5841
##          Pos Pred Value : 0.6025
##          Neg Pred Value : 0.5582
##              Prevalence : 0.5221
##          Detection Rate : 0.3012
##    Detection Prevalence : 0.5000
##       Balanced Accuracy : 0.5805
##
##        'Positive' Class : T
##
```

Elastic Net ROC Curve



Support Vector Machine (SVM)

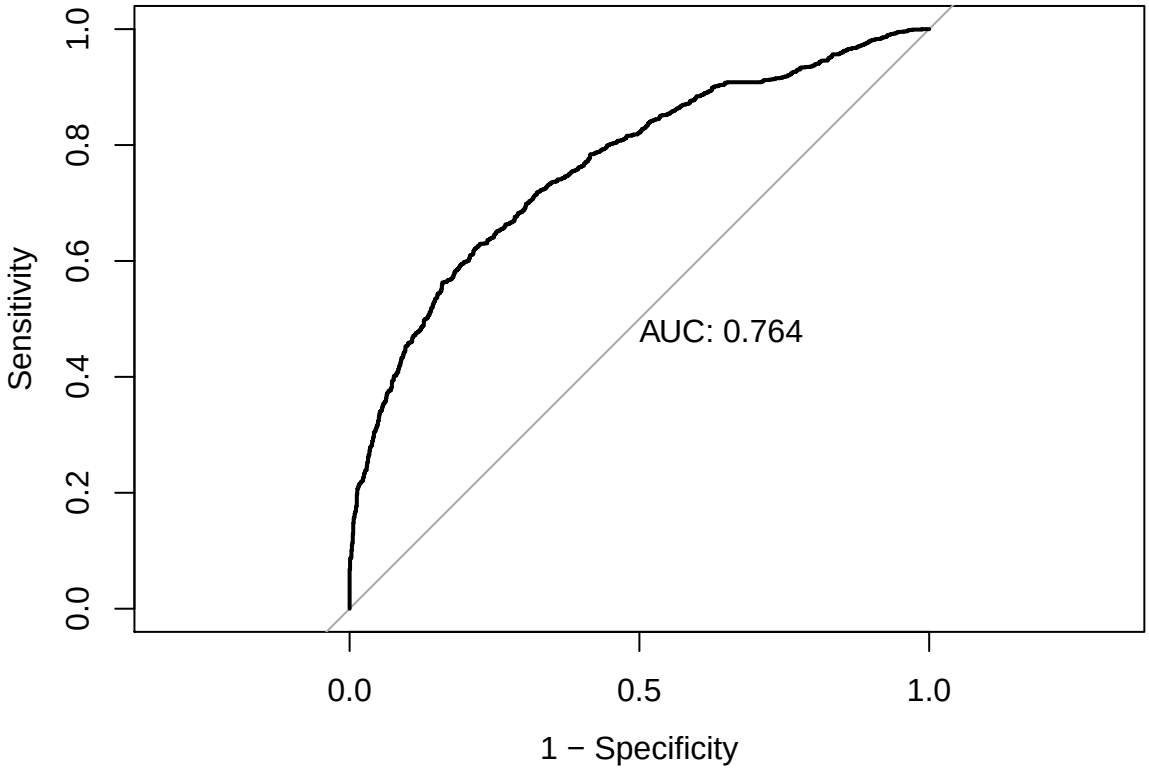
SVM Parameter Tunning

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 6016 samples
##    9 predictor
##    2 classes: 'F', 'T'
##
## Pre-processing: centered (9), scaled (9)
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 4812, 4814, 4813, 4813, 4812
## Resampling results across tuning parameters:
##
##   C       ROC        Sens       Spec
##     0.25  0.6810290  0.6745644  0.5822929
##     0.50  0.6983399  0.6835087  0.5984013
##     1.00  0.7095045  0.6911537  0.6190126
##     2.00  0.7179435  0.7007905  0.6176865
##     4.00  0.7247504  0.7087733  0.6156881
##     8.00  0.7296621  0.7180756  0.6110226
##    16.00  0.7320952  0.7277141  0.6083615
##    32.00  0.7338856  0.7343641  0.6040409
##    64.00  0.7459870  0.7274135  0.6321294
##   128.00  0.7316948  0.7244939  0.6032876
##
## Tuning parameter 'sigma' was held constant at a value of 0.105182
## ROC was used to select the optimal model using the largest value.
## The final values used for the model were sigma = 0.105182 and C = 64.
```

SVM Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   F   T
##          F 989 299
##          T 477 811
##
##                Accuracy : 0.6988
##                  95% CI : (0.6806, 0.7164)
##     No Information Rate : 0.5691
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3975
##
##  Mcnemar's Test P-Value : 2.099e-10
##
##             Sensitivity : 0.7306
##             Specificity : 0.6746
##          Pos Pred Value : 0.6297
##          Neg Pred Value : 0.7679
##              Prevalence : 0.4309
##          Detection Rate : 0.3148
##    Detection Prevalence : 0.5000
##       Balanced Accuracy : 0.7026
##
##        'Positive' Class : T
##
```

SVM ROC Curve

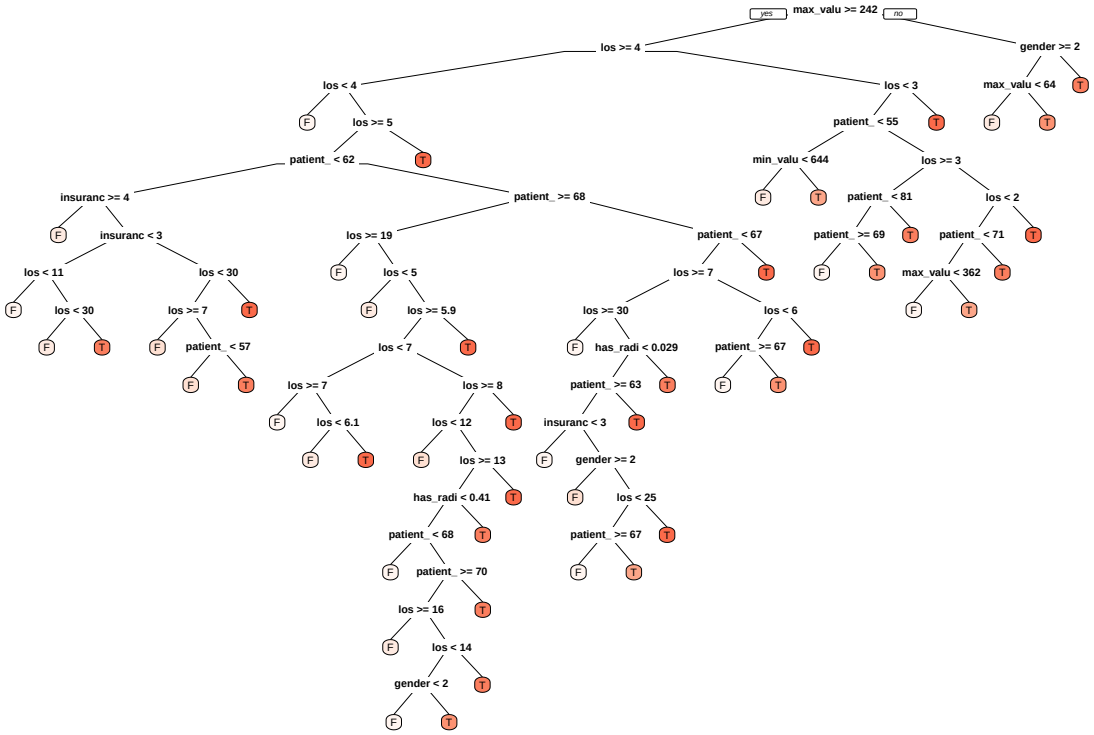


Decision Tree

Decision Tree Parameter Tunning

```
## CART
##
## 6016 samples
##    9 predictor
##    2 classes: 'F', 'T'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 4812, 4814, 4813, 4813, 4812
## Resampling results across tuning parameters:
##
##   cp           ROC        Sens       Spec
##   0.002105496  0.7424393  0.7766005  0.6030685
##   0.002160904  0.7424393  0.7766005  0.6030685
##   0.002244016  0.7414784  0.7742749  0.6030746
##   0.002659574  0.7361983  0.7782516  0.5891106
##   0.002992021  0.7252100  0.7676160  0.5847712
##   0.003158245  0.7208004  0.7636249  0.5827812
##   0.003324468  0.7148692  0.7765960  0.5618360
##   0.003878546  0.7058727  0.7845695  0.5408837
##   0.004875887  0.6987856  0.7769249  0.5352370
##   0.005817819  0.6973688  0.7669515  0.5379036
##   0.007313830  0.6882164  0.7383724  0.5561821
##   0.008311170  0.6856139  0.7310634  0.5591771
##   0.017121011  0.6452804  0.6885202  0.5491788
##   0.085771277  0.5765350  0.8477432  0.3011962
##   0.118351064  0.5342060  0.7315349  0.3368771
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.002160904.
```

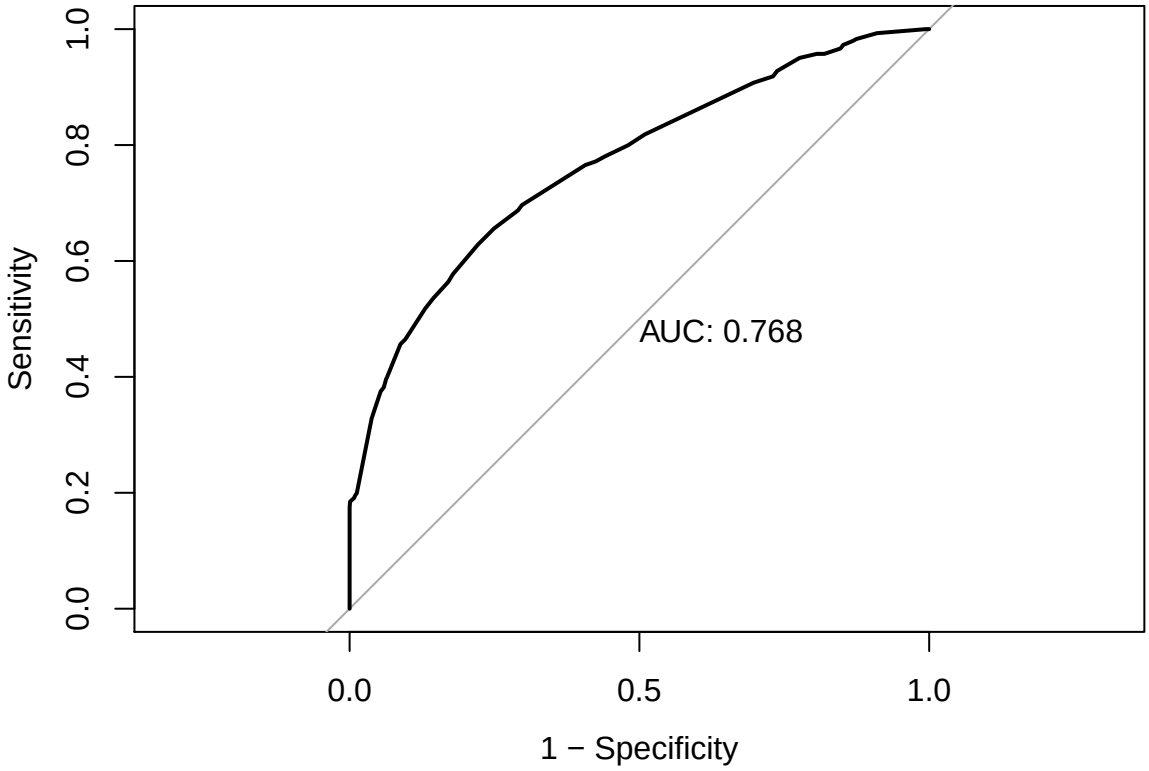
Decision Tree Leafs



Decision Tree Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    F    T
##          F 1003  285
##          T  479  809
##
##                Accuracy : 0.7034
##                  95% CI : (0.6854, 0.721)
##     No Information Rate : 0.5753
##     P-Value [Acc > NIR] : < 2e-16
##
##                   Kappa : 0.4068
##
##  Mcnemar's Test P-Value : 2.9e-12
##
##             Sensitivity : 0.7395
##             Specificity : 0.6768
##          Pos Pred Value : 0.6281
##          Neg Pred Value : 0.7787
##              Prevalence : 0.4247
##          Detection Rate : 0.3141
##    Detection Prevalence : 0.5000
##       Balanced Accuracy : 0.7081
##
##        'Positive' Class : T
##
```

Decision Tree ROC Curve



Random Forest

Random Forest Parameter Tunning

```
## Random Forest
##
## 6016 samples
##    9 predictor
##    2 classes: 'F', 'T'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold, repeated 3 times)
## Summary of sample sizes: 4812, 4814, 4813, 4813, 4812, 4812, ...
## Resampling results across tuning parameters:
##
##   mtry  ROC        Sens       Spec
##   2     0.8579879  0.8541727  0.6842772
##   5     0.8878289  0.8379941  0.7574159
##   9     0.8978671  0.8502947  0.7701605
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 9.
```

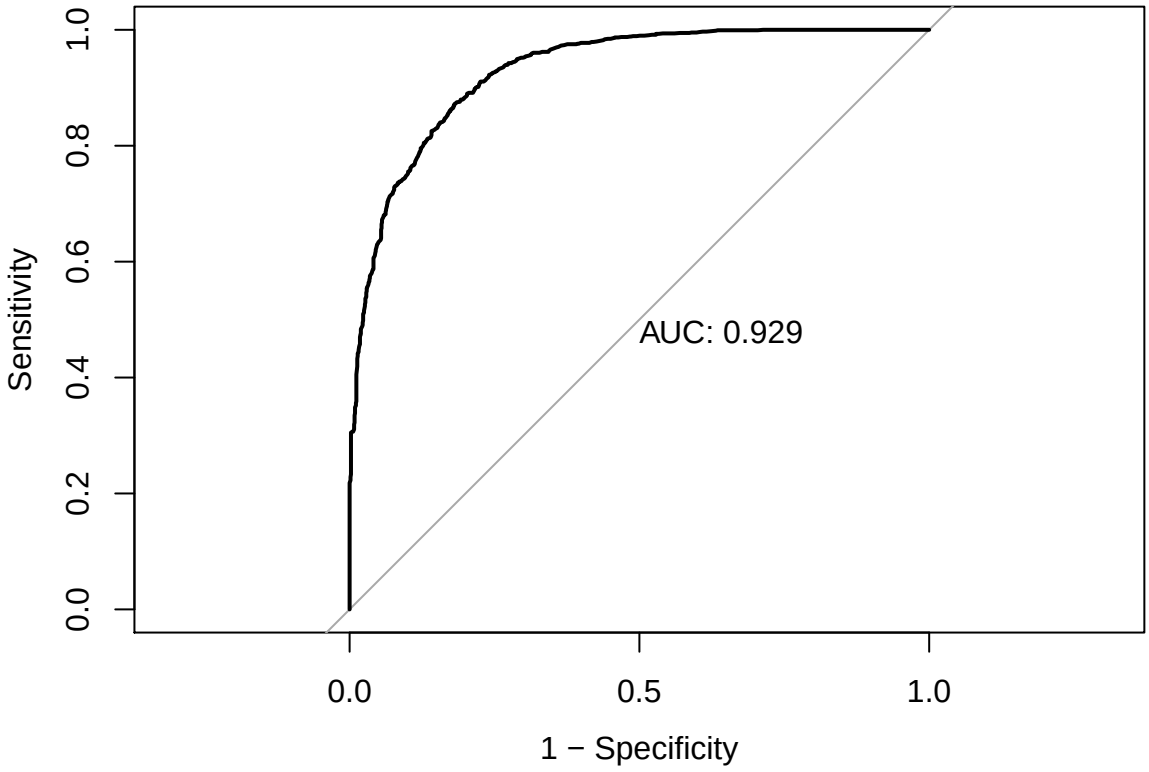
Random Forest Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    F    T
##          F 1130  158
##          T  266 1022
##
##                Accuracy : 0.8354
##                  95% CI : (0.8205, 0.8495)
##     No Information Rate : 0.5419
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6708
##
##  Mcnemar's Test P-Value : 2.032e-07
##
##             Sensitivity : 0.8661
##             Specificity : 0.8095
##          Pos Pred Value : 0.7935
##          Neg Pred Value : 0.8773
##              Prevalence : 0.4581
##          Detection Rate : 0.3967
##    Detection Prevalence : 0.5000
##       Balanced Accuracy : 0.8378
##
##        'Positive' Class : T
##
```

Random Forest ROC Curve

```
## Setting levels: control = F, case = T

## Setting direction: controls < cases
```



Extreme Gradient Boosting (XGBoost) Tree

```
## Warning: `repeats` has no meaning for this resampling method.
```

XGB Parameter Tunning

```
## eXtreme Gradient Boosting
##
## 6016 samples
##    9 predictor
##    2 classes: 'F', 'T'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 4812, 4814, 4813, 4813, 4812
## Resampling results across tuning parameters:
##
##   eta  max_depth  colsample_bytree  subsample  nrounds  ROC        Sens
##   0.3  1          0.6               0.50        50      0.7081424  0.6772013
##   0.3  1          0.6               0.50       100      0.7312862  0.6895070
##   0.3  1          0.6               0.50       150      0.7461146  0.7087711
##   0.3  1          0.6               0.75        50      0.7073871  0.6848453
##   0.3  1          0.6               0.75       100      0.7251849  0.6911598
##   0.3  1          0.6               0.75       150      0.7399006  0.7107716
##   0.3  1          0.6               1.00        50      0.7054782  0.6841797
##   0.3  1          0.6               1.00       100      0.7233767  0.6934915
##   0.3  1          0.6               1.00       150      0.7364151  0.7094449
##   0.3  1          0.8               0.50        50      0.7101318  0.6858486
##   0.3  1          0.8               0.50       100      0.7315958  0.7034665
##   0.3  1          0.8               0.50       150      0.7451238  0.7174173
##   0.3  1          0.8               0.75        50      0.7066513  0.6861731
##   0.3  1          0.8               0.75       100      0.7268964  0.6998049
##   0.3  1          0.8               0.75       150      0.7395877  0.7117639
##   0.3  1          0.8               1.00        50      0.7059034  0.6878364
##   0.3  1          0.8               1.00       100      0.7237312  0.6914937
##   0.3  1          0.8               1.00       150      0.7354353  0.7074477
##   0.3  2          0.6               0.50        50      0.7908946  0.7769310
##   0.3  2          0.6               0.50       100      0.8166638  0.8194797
##   0.3  2          0.6               0.50       150      0.8229954  0.8267981
##   0.3  2          0.6               0.75        50      0.7979771  0.8015323
##   0.3  2          0.6               0.75       100      0.8198172  0.8284603
##   0.3  2          0.6               0.75       150      0.8289154  0.8394326
##   0.3  2          0.6               1.00        50      0.7980003  0.8018623
##   0.3  2          0.6               1.00       100      0.8241551  0.8377709
##   0.3  2          0.6               1.00       150      0.8361721  0.8543872
##   0.3  2          0.8               0.50        50      0.7947905  0.7895556
##   0.3  2          0.8               0.50       100      0.8175696  0.8251336
##   0.3  2          0.8               0.50       150      0.8243277  0.8188175
##   0.3  2          0.8               0.75        50      0.7995904  0.7988734
##   0.3  2          0.8               0.75       100      0.8244555  0.8351026
##   0.3  2          0.8               0.75       150      0.8332666  0.8454094
##   0.3  2          0.8               1.00        50      0.8024538  0.8088435
##   0.3  2          0.8               1.00       100      0.8261277  0.8434193
```

```
##   0.3  2          0.8              1.00        150      0.8364701  0.8620411
##   0.3  3          0.6              0.50         50      0.8175288  0.8098418
##   0.3  3          0.6              0.50        100      0.8333452  0.8281281
##   0.3  3          0.6              0.50        150      0.8439586  0.8294504
##   0.3  3          0.6              0.75         50      0.8297580  0.8414298
##   0.3  3          0.6              0.75        100      0.8477656  0.8497360
##   0.3  3          0.6              0.75        150      0.8558674  0.8507360
##   0.3  3          0.6              1.00         50      0.8301620  0.8444132
##   0.3  3          0.6              1.00        100      0.8474194  0.8563778
##   0.3  3          0.6              1.00        150      0.8591233  0.8626962
##   0.3  3          0.8              0.50         50      0.8242833  0.8231353
##   0.3  3          0.8              0.50        100      0.8393571  0.8297787
##   0.3  3          0.8              0.50        150      0.8471736  0.8271253
##   0.3  3          0.8              0.75         50      0.8328777  0.8377715
##   0.3  3          0.8              0.75        100      0.8500021  0.8543905
##   0.3  3          0.8              0.75        150      0.8569397  0.8523966
##   0.3  3          0.8              1.00         50      0.8384391  0.8500766
##   0.3  3          0.8              1.00        100      0.8602229  0.8637089
##   0.3  3          0.8              1.00        150      0.8684700  0.8666923
##   0.4  1          0.6              0.50         50      0.7160854  0.6818542
##   0.4  1          0.6              0.50        100      0.7429470  0.7067855
##   0.4  1          0.6              0.50        150      0.7547612  0.7154156
##   0.4  1          0.6              0.75         50      0.7151606  0.6881648
##   0.4  1          0.6              0.75        100      0.7370665  0.7054494
##   0.4  1          0.6              0.75        150      0.7511436  0.7244040
##   0.4  1          0.6              1.00         50      0.7143139  0.6858431
##   0.4  1          0.6              1.00        100      0.7328208  0.7047888
##   0.4  1          0.6              1.00        150      0.7461030  0.7144311
##   0.4  1          0.8              0.50         50      0.7148311  0.6878458
##   0.4  1          0.8              0.50        100      0.7421936  0.7044511
##   0.4  1          0.8              0.50        150      0.7541342  0.7154140
##   0.4  1          0.8              0.75         50      0.7154781  0.6871631
##   0.4  1          0.8              0.75        100      0.7369014  0.7007993
##   0.4  1          0.8              0.75        150      0.7515243  0.7140922
##   0.4  1          0.8              1.00         50      0.7139559  0.6898309
##   0.4  1          0.8              1.00        100      0.7328752  0.7021299
##   0.4  1          0.8              1.00        150      0.7460129  0.7184211
##   0.4  2          0.6              0.50         50      0.7965097  0.7915644
##   0.4  2          0.6              0.50        100      0.8151181  0.8022029
##   0.4  2          0.6              0.50        150      0.8253412  0.8191591
##   0.4  2          0.6              0.75         50      0.8067674  0.7965412
##   0.4  2          0.6              0.75        100      0.8253791  0.8317887
##   0.4  2          0.6              0.75        150      0.8317940  0.8407610
##   0.4  2          0.6              1.00         50      0.8102493  0.8198147
##   0.4  2          0.6              1.00        100      0.8326766  0.8567194
##   0.4  2          0.6              1.00        150      0.8403104  0.8573839
##   0.4  2          0.8              0.50         50      0.7990198  0.7835744
##   0.4  2          0.8              0.50        100      0.8218886  0.8115024
##   0.4  2          0.8              0.50        150      0.8282784  0.8238047
##   0.4  2          0.8              0.75         50      0.8065504  0.8134969
##   0.4  2          0.8              0.75        100      0.8288731  0.8374365
##   0.4  2          0.8              0.75        150      0.8362806  0.8450876
##   0.4  2          0.8              1.00         50      0.8099049  0.8191536
##   0.4  2          0.8              1.00        100      0.8291549  0.8530600
```

```
##   0.4  2          0.8              1.00       150      0.8407416  0.8607072
##   0.4  3          0.6              0.50        50      0.8257197  0.8248069
##   0.4  3          0.6              0.50       100      0.8407828  0.8214659
##   0.4  3          0.6              0.50       150      0.8466531  0.8224725
##   0.4  3          0.6              0.75        50      0.8333105  0.8344354
##   0.4  3          0.6              0.75       100      0.8479856  0.8437482
##   0.4  3          0.6              0.75       150      0.8553257  0.8464005
##   0.4  3          0.6              1.00        50      0.8393194  0.8517371
##   0.4  3          0.6              1.00       100      0.8561374  0.8627050
##   0.4  3          0.6              1.00       150      0.8644049  0.8617039
##   0.4  3          0.8              0.50        50      0.8260230  0.8191641
##   0.4  3          0.8              0.50       100      0.8404833  0.8231436
##   0.4  3          0.8              0.50       150      0.8451976  0.8164919
##   0.4  3          0.8              0.75        50      0.8393187  0.8341026
##   0.4  3          0.8              0.75       100      0.8538128  0.8414199
##   0.4  3          0.8              0.75       150      0.8612462  0.8487255
##   0.4  3          0.8              1.00        50      0.8419683  0.8583794
##   0.4  3          0.8              1.00       100      0.8587150  0.8607017
##   0.4  3          0.8              1.00       150      0.8676424  0.8537244
##   Spec
##   0.6329888
##   0.6499284
##   0.6565729
##   0.6193470
##   0.6436167
##   0.6499306
##   0.6166964
##   0.6373005
##   0.6426156
##   0.6279949
##   0.6419528
##   0.6499301
##   0.6146970
##   0.6382928
##   0.6472678
##   0.6156992
##   0.6389616
##   0.6432817
##   0.6539190
##   0.6688819
##   0.6775242
##   0.6522557
##   0.6712047
##   0.6755220
##   0.6512546
##   0.6715358
##   0.6811742
##   0.6625624
##   0.6752008
##   0.6831720
##   0.6672180
##   0.6715413
##   0.6738653
##   0.6495951
```

```
##   0.6688758
##   0.6735292
##   0.6728636
##   0.6855020
##   0.6994632
##   0.6728620
##   0.6904821
##   0.7001211
##   0.6665491
##   0.6805054
##   0.6967949
##   0.6745247
##   0.6934694
##   0.7041144
##   0.6738631
##   0.6944682
##   0.7031105
##   0.6741898
##   0.6954655
##   0.7117589
##   0.6349700
##   0.6396206
##   0.6469351
##   0.6289927
##   0.6466017
##   0.6535763
##   0.6323193
##   0.6442784
##   0.6479356
##   0.6336361
##   0.6525862
##   0.6525851
##   0.6283288
##   0.6469323
##   0.6532468
##   0.6306538
##   0.6439500
##   0.6472662
##   0.6678708
##   0.6781848
##   0.6834937
##   0.6685336
##   0.6805148
##   0.6848348
##   0.6559068
##   0.6725341
##   0.6818431
##   0.6678786
##   0.6898187
##   0.6874948
##   0.6545790
##   0.6801770
##   0.6788437
##   0.6635502
```

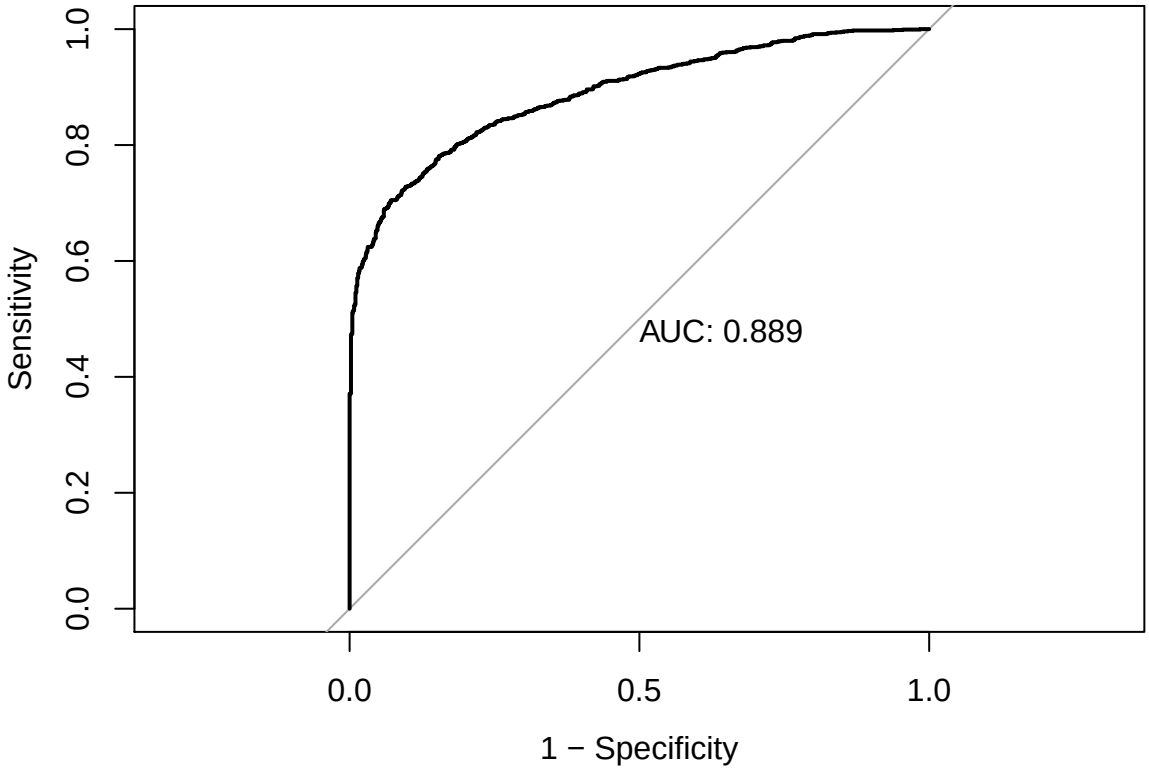
```
##   0.6711970
##   0.6795120
##   0.6791803
##   0.7031194
##   0.7177506
##   0.6801737
##   0.6991283
##   0.7110961
##   0.6771875
##   0.6984577
##   0.7104273
##   0.6864849
##   0.7021227
##   0.7120989
##   0.6924743
##   0.7037678
##   0.7104184
##   0.6844965
##   0.7027822
##   0.7137584
##
## Tuning parameter 'gamma' was held constant at a value of 0
## Tuning
##  parameter 'min_child_weight' was held constant at a value of 1
## ROC was used to select the optimal model using the largest value.
## The final values used for the model were nrounds = 150, max_depth = 3, eta
##  = 0.3, gamma = 0, colsample_bytree = 0.8, min_child_weight = 1 and subsample
##  = 1.
```

XGB Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    F    T
##          F 1164  124
##          T  354  934
##
##                Accuracy : 0.8144
##                  95% CI : (0.7989, 0.8293)
##     No Information Rate : 0.5893
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6289
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.8828
##             Specificity : 0.7668
##          Pos Pred Value : 0.7252
##          Neg Pred Value : 0.9037
##              Prevalence : 0.4107
##          Detection Rate : 0.3626
```

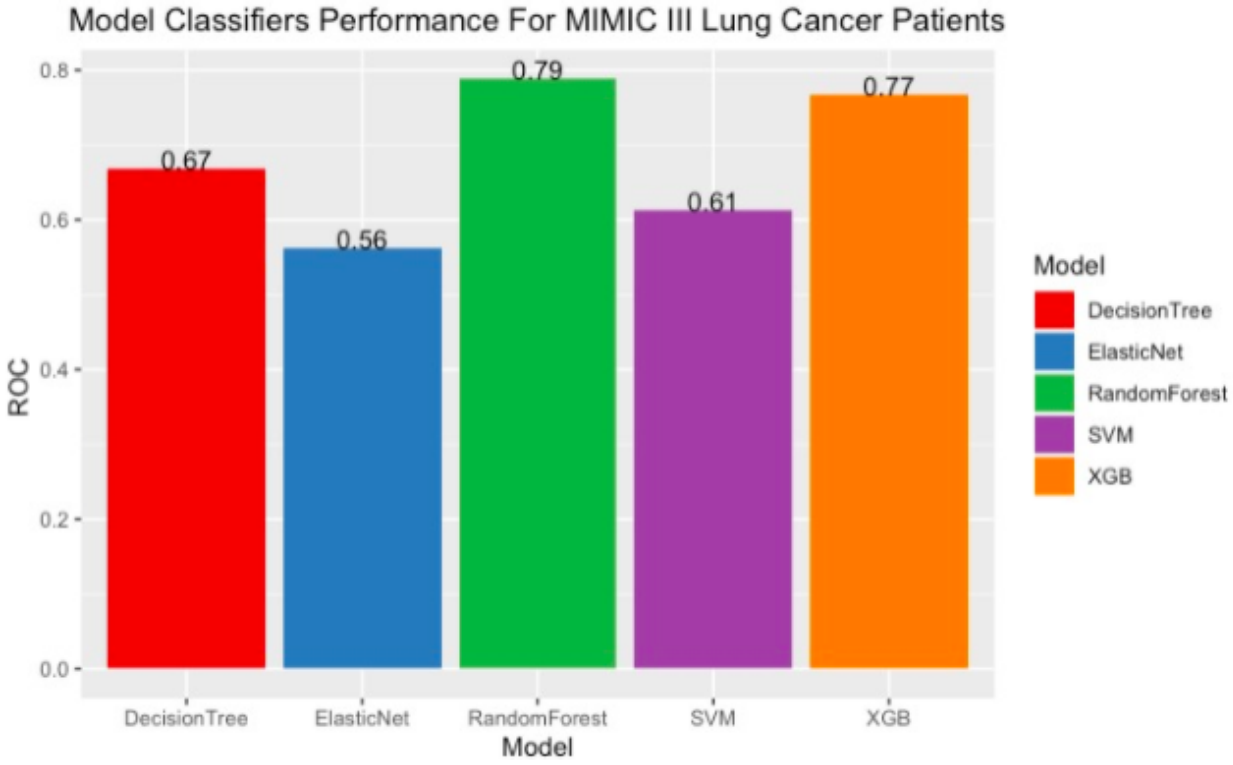
```
##    Detection Prevalence : 0.5000
##       Balanced Accuracy : 0.8248
##
##        'Positive' Class : T
##
```

XGB ROC Curve



Summary

Results with Imbalanced Original Data



Results with Balanced Data