

Oral Cancer Risk Index (OCRI) Test Report

Clinical Information

```
##   DNA_Index  X
## 1  4.707715 NA
## 2  4.106588 NA
## 3  2.408386 NA
## 4  2.393531 NA
## 5  2.380806 NA
## 6  2.353944 NA
```

Name: jjj Sex: o Age: 8 Tel: 000-000-0000 Address: 111 ABC St. City ST 00000 Sample No: 0011 Sampling date: 00/00/0000 Sampling hospital:Some Hospital Sampling physician: Dr. A B OCRI test date: 00/00/0000

```
combindo <- params$rda
library(caret)
library(pROC)
library(Metrics)
library(e1071)
library(ranger)
require(compiler)
multiClassSummary <- cmpfun(function (data, lev = NULL, model = NULL)
{
  #Load Libraries
  require(Metrics)
  require(caret)
  #Check data
  if (!all(levels(data[, "pred"]) == levels(data[, "obs"])))
    stop("levels of observed and predicted data do not match")
  #Calculate custom one-vs-all stats for each class
  prob_stats <- lapply(levels(data[, "pred"]), function(class)
  {
    #Grab one-vs-all data for the class
    pred <- ifelse(data[, "pred"] == class, 1, 0)
    obs <- ifelse(data[, "obs"] == class, 1, 0)
    prob <- data[,class]
    #Calculate one-vs-all AUC and logLoss and return
    cap_prob <- pmin(pmax(prob, .000001), .999999)
    prob_stats <- c(auc(obs, prob), logLoss(obs, cap_prob))
    names(prob_stats) <- c("ROC", "logLoss")
    return(prob_stats)
  })
  prob_stats <- do.call(rbind, prob_stats)
  rownames(prob_stats) <- paste( "Class:" , levels(data[, "pred"]))

  #Calculate confusion matrix-based statistics
  CM <- confusionMatrix(data[, "pred"], data[, "obs"])
  #Aggregate and average class-wise stats
  #Todo: add weights
  class_stats <- cbind(CM$byClass, prob_stats)
  class_stats <- colMeans(class_stats)
  #Aggregate overall stats
  overall_stats <- c(CM$overall)
```

```

#Combine overall with class-wise stats and remove some stats we don't want
stats <- c(overall_stats, class_stats)
stats <- stats[! names(stats) %in% c("AccuracyNull", "Prevalence", "Detection Prevalence")]

#Clean names and return

names(stats) <- gsub('[:blank:]', '_', names(stats))
return(stats)
})

library(ROCR)
set.seed(12345)
inTrainingSet <- createDataPartition(combindo$V11, p=.7, list=FALSE)
labelTrain <- combino[ inTrainingSet,]

ctrl <- trainControl(method = "repeatedcv",
                      repeats = 5,
                      summaryFunction = multiClassSummary,
                      classProbs = TRUE)

set.seed(1024)

rfFit <- train(V11 ~ ., data = labelTrain,
              ## training model: svm >>>
              method = "ranger",
              metric = "ROC",
              tuneLength = 10,
              trControl = ctrl)

## note: only 9 unique complexity parameters in default grid. Truncating the grid to 9 .

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```

## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in cbind(CM$byClass, prob_stats): number of rows of result is not a
## multiple of vector length (arg 1)

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
## trainInfo, : There were missing values in resampled performance measures.

peakfunc <- function(x, y){
  return(x[which(diff(sign(diff(y)))== -2)])
}
library(Hmisc)

interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
interval

## [1] 0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10.5

transform1 <- function(bb){
  answers8 <- peakfunc(density(bb$DNA_Index)$x, density(bb$DNA_Index)$y)
  answers8

  interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
  interval

  matrix2 <- rep(0, 10)

```

```

z1 <- interval[1:10]
z2 <- interval[2:11]
for (i in answers8){
  for (j in 1:10){
    if (i > z1[j] & i < z2[j]){
      matrix2[j] <- matrix2[j] + 1
    }
  }
}
matrix2 <- as.data.frame(t(matrix2))
matrix2$V11 <- "c"
matrix2$V11 <- as.factor(matrix2$V11)
matrix2 <- rbind(matrix2, matrix2)
return(matrix2)
}

plotmatrix2 <- function(bb){
  answers8 <- peakfunc(density(bb$DNA_Index)$x, density(bb$DNA_Index)$y)
  answers8

  interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
  interval

  matrix2 <- rep(0, 10)
  z1 <- interval[1:10]
  z2 <- interval[2:11]
  for (i in answers8){
    for (j in 1:10){
      if (i > z1[j] & i < z2[j]){
        matrix2[j] <- matrix2[j] + 1
      }
    }
  }
  matrix2 <- as.data.frame(t(matrix2))
  matrix2$V11 <- "c"
  matrix2$V11 <- as.factor(matrix2$V11)
  matrix2 <- rbind(matrix2, matrix2)
  jk <- as.numeric(as.matrix(matrix2[1,1:10]))
  names(jk) = c("0.5-1.5", "1.5-2.5", "2.5-3.5", "3.5-4.5", "4.5-5.5",
               "5.5-6.5", "6.5-7.5", "7.5-8.5", "8.5-9.5", "9.5-10.5")
  barplot(jk, xlab = "intervals", ylab = "peaks")
  text(labels = round(jk, digits = 2), x = seq(0.7, 12, by = 1.2),
       y = rep(1.5, 10))
}

fabaf <- function(bb){
  answers8 <- peakfunc(density(bb$DNA_Index)$x, density(bb$DNA_Index)$y)
  answers8

  interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
  interval

  matrix2 <- rep(0, 10)

```

```

z1 <- interval[1:10]
z2 <- interval[2:11]
for (i in answers8){
  for (j in 1:10){
    if (i > z1[j] & i < z2[j]){
      matrix2[j] <- matrix2[j] + 1
    }
  }
}
matrix2 <- as.data.frame(t(matrix2))
matrix2$V11 <- "c"
matrix2$V11 <- as.factor(matrix2$V11)
matrix2 <- rbind(matrix2, matrix2)
rfProbs <- predict(rfFit, matrix2, type = "prob")
return(rfProbs[1,1]*100)
}

suprema <- params$rdaz

rfProbs1 <- predict(rfFit, suprema, type = "prob")
rfProbs1

```

```

##           c           n
## 1  0.86245623 0.1375437693
## 2  0.69401372 0.3059862827
## 3  0.88252897 0.1174710295
## 4  0.95490437 0.0450956321
## 5  0.99626296 0.0037370430
## 6  0.87605167 0.1239483272
## 7  0.87473637 0.1252636341
## 8  0.87749432 0.1225056834
## 9  0.93881163 0.0611883717
## 10 0.99683439 0.0031656145
## 11 0.87833495 0.1216650466
## 12 0.99378677 0.0062132335
## 13 0.98312928 0.0168707154
## 14 0.87182769 0.1281723118
## 15 0.87825990 0.1217400969
## 16 0.86937065 0.1306293521
## 17 0.95422758 0.0457724223
## 18 0.87523913 0.1247608697
## 19 0.99478677 0.0052132335
## 20 0.86749264 0.1325073596
## 21 0.67109098 0.3289090234
## 22 0.68476935 0.3152306456
## 23 0.86651119 0.1334888115
## 24 0.72145996 0.2785400385
## 25 0.94123911 0.0587608903
## 26 0.87121133 0.1287886750
## 27 0.95264661 0.0473533860
## 28 0.88089261 0.1191073931
## 29 0.55900274 0.4409972624
## 30 0.99916800 0.0008320000

```

31 0.99807933 0.0019206743
32 0.86471262 0.1352873779
33 0.95918987 0.0408101320
34 0.68275404 0.3172459590
35 0.66179898 0.3382010201
36 0.66196612 0.3380338789
37 0.85607748 0.1439225188
38 0.67290938 0.3270906246
39 0.94978571 0.0502142875
40 0.86049017 0.1395098343
41 0.98553849 0.0144615144
42 0.96961355 0.0303864527
43 0.87722410 0.1227759009
44 0.99000444 0.0099955602
45 0.69481499 0.3051850072
46 0.98323027 0.0167697257
47 0.95191793 0.0480820682
48 0.87986981 0.1201301940
49 0.70849467 0.2915053273
50 0.96873364 0.0312663611
51 0.99916800 0.0008320000
52 0.95874887 0.0412511299
53 0.65149624 0.3485037618
54 0.88973996 0.1102600392
55 0.87784877 0.1221512334
56 0.69489477 0.3051052289
57 0.94995162 0.0500483764
58 0.99916800 0.0008320000
59 0.64378901 0.3562109916
60 0.69645898 0.3035410187
61 0.97802428 0.0219757165
62 0.95928569 0.0407143143
63 0.96075052 0.0392494782
64 0.86245623 0.1375437693
65 0.87766035 0.1223396534
66 0.87869859 0.1213014102
67 0.87825990 0.1217400969
68 0.87210834 0.1278916614
69 0.68194554 0.3180544634
70 0.95656645 0.0434335510
71 0.87568564 0.1243143625
72 0.67675669 0.3232433095
73 0.69387157 0.3061284337
74 0.80098656 0.1990134423
75 0.87535553 0.1246444669
76 0.99905479 0.0009452075
77 0.97222833 0.0277716736
78 0.68951907 0.3104809344
79 0.99683439 0.0031656145
80 0.95918987 0.0408101320
81 0.87192811 0.1280718870
82 0.95817839 0.0418216055
83 0.95207018 0.0479298244
84 0.67643174 0.3235682585

85 0.95809253 0.0419074694
86 0.99626296 0.0037370430
87 0.69103191 0.3089680854
88 0.64981741 0.3501825864
89 0.87779603 0.1222039711
90 0.98456626 0.0154337367
91 0.87882474 0.1211752643
92 0.86193026 0.1380697417
93 0.87932474 0.1206752643
94 0.13431518 0.8656848199
95 0.19368020 0.8063198035
96 0.10167754 0.8983224649
97 0.13519336 0.8648066435
98 0.13547441 0.8645255896
99 0.10334575 0.8966542456
100 0.12931195 0.8706880467
101 0.18085775 0.8191422487
102 0.13415081 0.8658491851
103 0.18085775 0.8191422487
104 0.12393168 0.8760683210
105 0.13547441 0.8645255896
106 0.11811274 0.8818872568
107 0.12381403 0.8761859681
108 0.20078559 0.7992144129
109 0.12624425 0.8737557543
110 0.09341474 0.9065852555
111 0.12362335 0.8763766519
112 0.13960426 0.8603957396
113 0.10696260 0.8930374010
114 0.10173002 0.8982699840
115 0.11263223 0.8873677734
116 0.10792925 0.8920707536
117 0.10791393 0.8920860670
118 0.12330556 0.8766944359
119 0.19543020 0.8045698035
120 0.10505351 0.8949464919
121 0.12378446 0.8762155385
122 0.10853158 0.8914684199
123 0.09556182 0.9044381814
124 0.13314871 0.8668512937
125 0.20173967 0.7982603313
126 0.12509815 0.8749018473
127 0.12467205 0.8753279509
128 0.12358662 0.8764133820
129 0.10173002 0.8982699840
130 0.13415081 0.8658491851
131 0.12976774 0.8702322638
132 0.11503901 0.8849609937
133 0.12498051 0.8750194944
134 0.12330556 0.8766944359
135 0.11671884 0.8832811594
136 0.10737855 0.8926214550
137 0.10791393 0.8920860670
138 0.12358662 0.8764133820

139 0.11133337 0.8886666259
140 0.10749619 0.8925038079
141 0.13547441 0.8645255896
142 0.10321070 0.8967893002
143 0.11430112 0.8856988840
144 0.13519336 0.8648066435
145 0.10211097 0.8978890316
146 0.12976774 0.8702322638
147 0.13431518 0.8656848199
148 0.11951706 0.8804829443
149 0.13431518 0.8656848199
150 0.12372099 0.8762790080
151 0.13572797 0.8642720346
152 0.12976774 0.8702322638
153 0.09303379 0.9069662079
154 0.11811274 0.8818872568
155 0.10739386 0.8926061416
156 0.12854959 0.8714504072
157 0.12393168 0.8760683210
158 0.12455440 0.8754455979
159 0.11753687 0.8824631283
160 0.11739512 0.8826048818
161 0.10942428 0.8905757168
162 0.12598862 0.8740113848
163 0.19580771 0.8041922868
164 0.11016662 0.8898333793
165 0.12902304 0.8709769609
166 0.10739386 0.8926061416
167 0.13110741 0.8688925875
168 0.19720317 0.8027968275
169 0.12770559 0.8722944123
170 0.12498051 0.8750194944
171 0.12393168 0.8760683210
172 0.59374292 0.4062570831
173 0.12467205 0.8753279509
174 0.12322904 0.8767709588
175 0.12498051 0.8750194944
176 0.13110741 0.8688925875
177 0.11739512 0.8826048818
178 0.11379664 0.8862033642
179 0.11739512 0.8826048818
180 0.12931195 0.8706880467
181 0.12800949 0.8719905142
182 0.10522392 0.8947760842
183 0.12482483 0.8751751698
184 0.12726296 0.8727370377
185 0.10696260 0.8930374010
186 0.10696260 0.8930374010
187 0.10769915 0.8923008480
188 0.12482483 0.8751751698
189 0.10545290 0.8945470987
190 0.12329080 0.8767091998
191 0.12498051 0.8750194944
192 0.11670712 0.8832928830

193 0.09206909 0.9079309077
194 0.10579585 0.8942041544
195 0.13519336 0.8648066435
196 0.10545290 0.8945470987
197 0.13076739 0.8692326136
198 0.10497424 0.8950257595
199 0.67420507 0.3257949273
200 0.11263223 0.8873677734
201 0.10853158 0.8914684199
202 0.11811274 0.8818872568
203 0.11379664 0.8862033642
204 0.12582428 0.8741757193
205 0.10791393 0.8920860670
206 0.11958419 0.8804158079
207 0.11133337 0.8886666259
208 0.10792925 0.8920707536
209 0.18085775 0.8191422487
210 0.12901739 0.8709826136
211 0.13547441 0.8645255896
212 0.87932474 0.1206752643
213 0.11753687 0.8824631283
214 0.11263223 0.8873677734
215 0.10792925 0.8920707536
216 0.09303379 0.9069662079
217 0.12598862 0.8740113848
218 0.11133337 0.8886666259
219 0.10853158 0.8914684199
220 0.10696260 0.8930374010
221 0.19736948 0.8026305178
222 0.11263223 0.8873677734
223 0.12378446 0.8762155385
224 0.11739512 0.8826048818
225 0.81160278 0.1883972226
226 0.95422758 0.0457724223
227 0.12361866 0.8763813416
228 0.87844859 0.1215514102
229 0.12480952 0.8751904832
230 0.12612065 0.8738793521
231 0.61165884 0.3883411630
232 0.87182769 0.1281723118
233 0.82869098 0.1713090228
234 0.10993034 0.8900696590
235 0.11586865 0.8841313476
236 0.86382230 0.1361777004
237 0.13297510 0.8670249015
238 0.13285838 0.8671416208
239 0.12376915 0.8762308519
240 0.13541537 0.8645846285
241 0.12610533 0.8738946655
242 0.10505351 0.8949464919
243 0.12370269 0.8762973066
244 0.67960288 0.3203971193
245 0.87722410 0.1227759009
246 0.12358505 0.8764149536

```
## 247 0.12482483 0.8751751698
## 248 0.12381403 0.8761859681
## 249 0.12378446 0.8762155385
## 250 0.11753687 0.8824631283
## 251 0.12381403 0.8761859681
## 252 0.95888569 0.0411143143
## 253 0.12329080 0.8767091998
## 254 0.11811274 0.8818872568
## 255 0.09556182 0.9044381814
## 256 0.12455440 0.8754455979
## 257 0.79743165 0.2025683496
## 258 0.18085775 0.8191422487
## 259 0.59552917 0.4044708341
## 260 0.61360431 0.3863956883
## 261 0.68674325 0.3132567492
## 262 0.12467205 0.8753279509
## 263 0.12598862 0.8740113848
## 264 0.10792925 0.8920707536
## 265 0.20173967 0.7982603313
## 266 0.11133337 0.8886666259
## 267 0.66429331 0.3357066880
## 268 0.20078559 0.7992144129
## 269 0.12446990 0.8755301014
## 270 0.13415081 0.8658491851
## 271 0.12509815 0.8749018473
## 272 0.11739512 0.8826048818
## 273 0.69323784 0.3067621577
## 274 0.12916459 0.8708354107
## 275 0.10791393 0.8920860670
## 276 0.12467205 0.8753279509
## 277 0.94995162 0.0500483764
```

```
ploty1 <- function(bb){

  answers8 <- peakfunc(density(bb$DNA_Index)$x, density(bb$DNA_Index)$y)
  answers8

  interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
  interval

  matrix2 <- rep(0, 10)
  z1 <- interval[1:10]
  z2 <- interval[2:11]
  for (i in answers8){
    for (j in 1:10){
      if (i > z1[j] & i < z2[j]){
        matrix2[j] <- matrix2[j] + 1
      }
    }
  }
  matrix2 <- as.data.frame(t(matrix2))
  matrix2$V11 <- "c"
  matrix2$V11 <- as.factor(matrix2$V11)
  matrix2 <- rbind(matrix2, matrix2)
```

```

rfProbs <- predict(rfFit, matrix2, type = "prob")

par(mfrow=c(1,2), pin = c(2,2))
labels <- c(rep("OSCC", 93), rep("Normal", 102), rep("OLK", 82))
dts <- c(rfProbs1$c)
dt2plot <- as.data.frame (list (lab = labels, dt = dts))
boxplot(dt ~ lab, data = dt2plot, ylab = "OCRI", outline = FALSE,ylim = c(0,1))
stripchart(dt ~ lab, vertical = TRUE, data = dt2plot
, add = TRUE, pch = 20, col = "green")
if(rfProbs[1,1] > 0.5){
  plot(rfProbs[1,1], ylim = c(0,1), xaxt='n', xlab = "Patient", ylab = "OCRI", col = "red", pch = 20)
}
else{
  plot(rfProbs[1,1], ylim = c(0,1), xaxt='n', xlab = "Patient", ylab = "OCRI", col = "blue", pch = 20)
}
}

ploty2 <- function(bb){

  answers8 <- peakfunc(density(bb$DNA_Index)$x, density(bb$DNA_Index)$y)
  #answers8

  interval <- c(0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5)
  #interval

  matrix2 <- rep(0, 10)
  z1 <- interval[1:10]
  z2 <- interval[2:11]
  for (i in answers8){
    for (j in 1:10){
      if (i > z1[j] & i < z2[j]){
        matrix2[j] <- matrix2[j] + 1
      }
    }
  }
  matrix2 <- as.data.frame(t(matrix2))
  matrix2$V11 <- "c"
  matrix2$V11 <- as.factor(matrix2$V11)
  matrix2 <- rbind(matrix2, matrix2)
  rfProbs <- predict(rfFit, matrix2, type = "prob")

  if(rfProbs[1,1] > 0.5){
    plot(rfProbs[1,1], ylim = c(0,1), xaxt='n', xlab = "Patient", ylab = "OCRI", type='n')
    text(1,rfProbs[1,1],label=rfProbs[1,1],col='red')
  }
  else{
    plot(rfProbs[1,1], ylim = c(0,1), xaxt='n', xlab = "Patient", ylab = "OCRI", type='n')
    text(1,rfProbs[1,1],label=rfProbs[1,1],col='blue')
  }
}

```