library(tidyverse)

library(GEOquery)

library(tidyverse)

library(GEOquery)

library(limma)

library(affy)

library(stringr)

gset <- getGEO('GSE55457', destdir=".",

AnnotGPL = T,

getGPL = T)

exp<-exprs(gset[[1]])

exp<-log2(exp+1)

cli<-pData(gset[[1]])

GPL<-fData(gset[[1]])

gpl<-GPL[,c(1,3)]

gpl$`Gene symbol`<-data.frame(sapply(gpl$`Gene symbol`,function(x)unlist(strsplit(x,"///"))[1]),stringsAsFactors=F)[,1]

exp<-as.data.frame(exp)

exp$ID<-rownames(exp)

exp\_symbol<-merge(exp,gpl,by="ID")

exp\_symbol<-na.omit(exp\_symbol)

table(duplicated(exp\_symbol$`Gene symbol`))

exp\_unique<-avereps(exp\_symbol[,-c(1,ncol(exp\_symbol))],ID=exp\_symbol$`Gene symbol`)

write.table(exp\_unique,"exp55457.csv",sep=",")

#nomo,dac,cal,auc analysis of trainset

library(rms)

mydata<-read.csv("exp yuan.csv", header = TRUE, sep = ",")

str(mydata)

mydata$group<-ifelse(mydata$group =="RA",1,0)

ddist <- datadist(mydata)

options(datadist='ddist')

fit1<-lrm(group~RRM2+DLGAP5+KIF11,data=mydata,x=TRUE,y=TRUE)

Nomo\_LR<-nomogram(fit1,

fun=plogis,

lp=F,

fun.at = c(0.001,0.01,0.1,0.5,0.9,0.99),

funlabel = "Risk of RA")

plot(Nomo\_LR)

cal1 <- calibrate(fit1, method='boot', B=1000)

plot(cal1,

xlim = c(0,1),

xlab = "Predicted Probability",

ylab = "Observed Probability",

legend = FALSE,

subtitles = FALSE)

abline(0,1,col = "black",lty = 2,lwd = 2)

lines(cal1[,c("predy","calibrated.orig")], type = "l",lwd = 2,col="red",pch =16)

lines(cal1[,c("predy","calibrated.corrected")], type = "l",lwd = 2,col="green",pch =16)

legend(0.55,0.35,

c("Apparent","Ideal","Bias-corrected"),

lty = c(2,1,1),

lwd = c(2,1,1),

col = c("black","red","green"),

bty = "n") # "o"为加边框

library(rmda)

library(ggplot2)

library(rms)

library(caret)

complex<-decision\_curve(group ~RRM2+DLGAP5+KIF11,

data = mydata,family = binomial(link ='logit'),

thresholds = seq(0,1, by = 0.01),

confidence.intervals= 0.95)

plot\_decision\_curve(complex,

curve.names=c("Model"),

cost.benefit.axis =FALSE,

col= c("blue","red","black"),

confidence.intervals=FALSE,#

standardize = FALSE)#

library(multipleROC)

rm(list = ls())

df <- as.data.frame(mydata)

p1 <- multipleROC(group~RRM2,data=df)

#p2 <- multipleROC(group~ERBB4,data=df)

p3 <- multipleROC(group~DLGAP5,data=df)

p4<-multipleROC(group~KIF11,data=df)

p5<-multipleROC(group~LMCD1,data=df)

#p6<-multipleROC(group~CSF3R,data=df)

plot\_ROC(list(p1,p3,p4),

show.points = T,

show.eta = F,

show.sens = F,

show.AUC = T,

facet = F )

library(pROC)

library(ggplot2)

pred\_f\_training<-predict(fit1,mydata)

#

modelroc <- roc(mydata$group,pred\_f\_training)

auc(modelroc)

ci(modelroc)

# Area under the curve: 0.7851

# 95% CI: 0.7335-0.8368 (DeLong)

plot(modelroc,col="red",

legacy.axes=T,

print.auc=TRUE,

print.thres=TRUE,

grid=c(0.2,0.2),grid.col=c("blue","yellow"))

#nomo,roc,dac,cal analysis of gse55457

library(rms)

mydata<-read.csv("exp55457.csv", header = TRUE, sep = ",")

str(mydata)

mydata$group<-ifelse(mydata$group =="RA",1,0)

ddist <- datadist(mydata)

options(datadist='ddist')

fit1<-lrm(group~RRM2+DLGAP5+KIF11,data=mydata,x=TRUE,y=TRUE)

Nomo\_LR<-nomogram(fit1,

fun=plogis,

lp=F,

fun.at = c(0.001,0.01,0.1,0.3,0.5,0.7,0.9,0.99),

funlabel = "Risk of RA")

plot(Nomo\_LR)

cal1 <- calibrate(fit1, method='boot', B=1000)

plot(cal1,

xlim = c(0,1),

xlab = "Predicted Probability",

ylab = "Observed Probability",

legend = FALSE,

subtitles = FALSE)

abline(0,1,col = "black",lty = 2,lwd = 2)

lines(cal1[,c("predy","calibrated.orig")], type = "l",lwd = 2,col="red",pch =16)

lines(cal1[,c("predy","calibrated.corrected")], type = "l",lwd = 2,col="green",pch =16)

legend(0.55,0.35,

c("Apparent","Ideal","Bias-corrected"),

lty = c(2,1,1),

lwd = c(2,1,1),

col = c("black","red","green"),

bty = "n") # "o"为加边框

library(rmda)

library(ggplot2)

library(rms)

library(caret)

complex<-decision\_curve(group ~RRM2+DLGAP5+KIF11,

data = mydata,family = binomial(link ='logit'),

thresholds = seq(0,1, by = 0.01),

confidence.intervals= 0.95)

plot\_decision\_curve(complex,

curve.names=c("Model"),#图例上每条曲线的名字

cost.benefit.axis =FALSE,# cost.benefit.axis是另外附加的一条横坐标轴，损失收益比，默认值是TRUE

col= c("blue","red","black"),#曲线颜色

confidence.intervals=FALSE,# 设置是否画出曲线的置信区间

standardize = FALSE)# 设置是否对净受益率（NB）使用患病率进行校正

library(multipleROC)

rm(list = ls())

df <- as.data.frame(mydata)

p1 <- multipleROC(group~RRM2,data=df)

#p2 <- multipleROC(group~ERBB4,data=df)

p3 <- multipleROC(group~DLGAP5,data=df)

p4<-multipleROC(group~KIF11,data=df)

p5<-multipleROC(group~LMCD1,data=df)

#p6<-multipleROC(group~CSF3R,data=df)

plot\_ROC(list(p1,p3,p4),

show.points = T,

show.eta = F,

show.sens = F,

show.AUC = T,

facet = F )

library(pROC)

library(ggplot2)

pred\_f\_training<-predict(fit1,mydata)

modelroc <- roc(mydata$group,pred\_f\_training)

auc(modelroc)#

ci(modelroc) #AUC95%CI

# Area under the curve: 0.7851

# 95% CI: 0.7335-0.8368 (DeLong)

plot(modelroc,col="red",

legacy.axes=T,

print.auc=TRUE,

print.thres=TRUE,

grid=c(0.2,0.2),grid.col=c("blue","yellow"))