library(caret)

#read in data

MLpihdata<-read.csv("H:\\Personal\\predindhypo\\MLpihdummy.csv")

#feature selection using recursive feature elimination

control <- rfeControl(functions=rfFuncs, method="cv", number=10)

subsets<-c(5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,78)

results <- rfe(MLpihdata[,c("timeofday","prop","fent","midaz","rocur","etom","succ","firstMAP","maxsev","maxdes","meanpip","DM","HTN","CHF","CAD","AS","OSAp","AFIB","COPD","CKD","asthma","GERD","BMI.x","ASA\_RATING","Sex","Age..Years.","ACEorNOT","ALPHAorNOT","ANTIANGorNOT","ANTIANXorNOT","ANTIARRorNOT","ANTICONVorNOT","ANTIDEPorNOT","ANTINEOorNOT","ANTIPARKorNOT","ANTIPLATorNOT","ANTIRETROorNOT","ARBorNOT","ASTHMAANTICHOLorNOT","ASTHMABETALONGorNOT","ASTHMABETASHORTorNOT","ASTHMAGLUCorNOT","BETAorNOT","BIGUANIDEorNOT","BISPHOSorNOT","CALCIMIMEorNOT","DIGorNOT","DMARDorNOT","DPP4orNOT","EPOorNOT","GLUCOCORTorNOT","GLUCOSIDASEorNOT","HYDRALorNOT","IMMUNOSUPPorNOT","INCRETINorNOT","INSULONGorNOT","INSUSHORTorNOT","KSPARINGorNOT","LEUKOTRIorNOT","LITHIUMorNOT","MEGLITorNOT","NARCorNOT","PHOSBINDorNOT","PROSTATEorNOT","RIFAMYCINorNOT","SULFONYLorNOT","THIAZIDEorNOT","THROMBINHorNOT","THYROIDorNOT","TZDorNOT","WARForNOT","LOOPorNOT","KSPARINGorNOT","GERDorNOT","STATINorNOT","ASPIRINorNOT","CCBorNOT","XAINHIBorNOT")], MLpihdata$hypo, sizes=subsets, rfeControl=control)

print(results)

#prepare output names

MLpihdata$hypo<-as.factor(MLpihdata$hypo)

levels(MLpihdata$hypo) <- make.names(levels(factor(MLpihdata$hypo)))

#randomly split data into train and test sets

trainIndex = sample(1:nrow(MLpihdata), size = round(0.7\*nrow(MLpihdata)), replace=FALSE)

train = MLpihdata[trainIndex ,]

test = MLpihdata[-trainIndex ,]

#set training parameters

train\_control<-trainControl(method="repeatedcv", number=10, repeats=3,savePredictions=T,classProbs=T,summaryFunction=twoClassSummary,verboseIter=TRUE)

#select algorithm (random forest = rf; logistic regression = glm; naive bayes = nb; support vector machines = svmRadial; k nearest neighbor = knn; linear discriminant analysis = lda; neural net = nnet; gradient boosting machine = gbm)

whichmethod="glm"

#train model

whichmodel<-train(hypo~firstMAP+Age..Years.+prop+ARBorNOT+meanpip+BMI.x+ASA\_RATING+rocur+fent+Sex+BISPHOSorNOT+maxsev+DM+BETAorNOT+midaz+ANTIPLATorNOT+THYROIDorNOT+etom+GERDorNOT+HTN+timeofday+maxdes+STATINorNOT+succ+ASPIRINorNOT+PROSTATEorNOT+THIAZIDEorNOT+LOOPorNOT+PHOSBINDorNOT+ACEorNOT+CALCIMIMEorNOT+DMARDorNOT+IMMUNOSUPPorNOT+ANTIARRorNOT+CCBorNOT+KSPARINGorNOT+SULFONYLorNOT+ANTICONVorNOT+GLUCOCORTorNOT+BIGUANIDEorNOT+OSAp+CHF+COPD+ASTHMABETASHORTorNOT+ANTIANXorNOT+XAINHIBorNOT+AFIB+EPOorNOT+GERD+ANTINEOorNOT+WARForNOT+ANTIDEPorNOT+CAD+ASTHMAGLUCorNOT+DPP4orNOT+CKD+asthma+ASTHMAANTICHOLorNOT, data=train, trControl=train\_control, method=whichmethod,metric="ROC")

#view model performance metrics

whichmodel