

# MISTIE

Chuyi Sun

2025-02-09

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(tidyr)
library(haven)
library(randomForest)
```

```
## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##     combine
##
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library(glmnet)
```

```
## Loaded glmnet 4.1-8
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
```

```
## Attaching package: 'pROC'
```

```
##
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      cov, smooth, var
```

```
## Read data
```

```
dat=read_dta("M3_data_499_TIL.dta")
```

```
## summary(dat)
```

```
day7=dat%>%select(matches("til", ignore.case=FALSE))
```

```
## cor(day7,use = "pairwise.complete.obs")
```

```
## bp control
```

```
## score bp
```

```
newdata=read_dta("M3_TIL_added_data.dta")
```

```
newdata=newdata%>%
```

```
  mutate(site_continent=as.factor(site_continent))
```

```
table(newdata$site_continent)
```

```
##
```

```
## Australia-Asia      Europe  North America
```

```
##              14              85              400
```

```
merge
```

```
all_data=left_join(dat,newdata, by=c("patientnum_ninds"="new_id"))
```

```
select needed variables
```

```
data_1=all_data%>%
```

```
  select(patientnum_ninds, age_at_consent, gcs_randomization, nihss_randomization, stabct_ich_volume, s
```

```
  mutate(across(c(1,7:26), as.factor))
```

```
## only for more than two levels and characteristics
```

```
summary(data_1)
```

```

## patientnum_ninds age_at_consent gcs_randomization nihss_randomization
## 2001 : 1 Min. :28.00 Min. : 3.00 Min. : 1.00
## 2002 : 1 1st Qu.:52.00 1st Qu.: 8.00 1st Qu.:15.00
## 2003 : 1 Median :62.00 Median :10.00 Median :19.00
## 2004 : 1 Mean :61.11 Mean :10.57 Mean :19.32
## 2005 : 1 3rd Qu.:71.00 3rd Qu.:13.00 3rd Qu.:23.00
## 2006 : 1 Max. :90.00 Max. :15.00 Max. :40.00
## (Other):493
## stabct_ich_volume stabct_ivh_volume eot_less_15 BaselineNEWscore_BP
## Min. : 20.89 Min. : 0.000 0 :346 0: 21
## 1st Qu.: 35.60 1st Qu.: 0.000 1 :148 1:116
## Median : 45.60 Median : 0.400 NA's: 5 2:138
## Mean : 48.96 Mean : 2.698 3:142
## 3rd Qu.: 58.60 3rd Qu.: 3.200 4: 22
## Max. :127.09 Max. :61.800 5: 60
##
## Baseline_BP_control Baseline_Hypotension Baseline_Hyperpyrexia
## 0: 21 0:480 0:447
## 1:191 1: 17 1: 38
## 2:224 2: 2 2: 8
## 3: 63 3: 6
##
##
## Baseline_Hyperglycemia Baseline_ICP Baseline_herniation Baseline_INR
## 0:399 0 :487 0 :484 0:466
## 1: 49 1 : 8 1 : 8 1: 24
## 2: 35 2 : 1 2 : 6 2: 9
## 3: 16 3 : 2 NA's: 1
## NA's: 1
##
##
## Day7NEWscore_BP D7_BP_control D7_Hypotension D7_Hyperpyrexia D7_Hyperglycemia
## 0 : 14 0 : 14 0 :444 0 :323 0 :370
## 1 :130 1 :164 1 : 29 1 : 78 1 : 40
## 2 :233 2 :221 2 : 13 2 : 38 2 : 37
## 3 : 69 3 : 87 NA's: 13 3 : 47 3 : 39
## 4 : 37 NA's: 13 NA's: 13 NA's: 13
## 5 : 3
## NA's: 13
## D7_ICP D7_herniation D7_INR D7_DNR glasgow_rankin_0_3_30
## 0 :454 0 :474 0 :462 0:489 0 :436
## 1 : 15 1 : 6 1 : 13 1: 4 1 : 57
## 2 : 9 2 : 6 2 : 11 2: 2 NA's: 6
## 3 : 7 NA's: 13 NA's: 13 3: 4
## 4 : 1
## NA's: 13
##
## site_continent
## Australia-Asia: 14
## Europe : 85
## North America :400
##
##

```

```
##  
##
```

```
na_rows=data_1 %>%  
  filter(if_any(everything(), is.na))  
  
data_1=drop_na(data_1)## remove 22?  
  
## glasgow_rankin_0_3_30 6 NAs  
## other NA: 12 most missingness are DAY7  
sum(is.na(data_1))
```

```
## [1] 0
```

```
## check correlation for factor, cor(data_1)  
## cor(data_1[,7:24])
```

## split data

```
## split by region  
set.seed(2025)  
  
train_data = data_1 %>%  
  group_by(site_continent) %>%           # Group by region  
  slice_sample(prop = 0.8) %>%         # Sample 80% from each region  
  ungroup()  
test_data = anti_join(data_1, train_data, by = "patientnum_ninds")
```

```
## logistics regression
```

```
##
```

```
# m_log=glmer(glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization + stabct_ich_volume + stabct_ivh_volume + eot_less_15 + BaselineNEWscore_BP + Baseline_BP_control + Baseline_Hypotension + Baseline_Hyperpyrexia + Baseline_Hyperglycemia + Baseline_ICP + Baseline_herniation + Baseline_INR + Day7NEWscore_BP + D7_BP_control + D7_Hypotension + D7_Hyperpyrexia + D7_Hyperglycemia + D7_ICP + D7_herniation + D7_INR + D7_DNR + site_continent, family = binomial, data = train_data)
```

```
## day 7 bp* control
```

```
m_log=glm(glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization + stabct_ich_volume + stabct_ivh_volume + eot_less_15 + BaselineNEWscore_BP + Baseline_BP_control + Baseline_Hypotension + Baseline_Hyperpyrexia + Baseline_Hyperglycemia + Baseline_ICP + Baseline_herniation + Baseline_INR + Day7NEWscore_BP + D7_BP_control + D7_Hypotension + D7_Hyperpyrexia + D7_Hyperglycemia + D7_ICP + D7_herniation + D7_INR + D7_DNR + site_continent, family = binomial, data = train_data)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(m_log)
```

```
##
```

```
## Call:
```

```
## glm(formula = glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization +  
##   nihss_randomization + stabct_ich_volume + stabct_ivh_volume +  
##   eot_less_15 + BaselineNEWscore_BP + Baseline_BP_control +  
##   Baseline_Hypotension + Baseline_Hyperpyrexia + Baseline_Hyperglycemia +  
##   Baseline_ICP + Baseline_herniation + Baseline_INR + Day7NEWscore_BP +  
##   D7_BP_control + D7_Hypotension + D7_Hyperpyrexia + D7_Hyperglycemia +  
##   D7_ICP + D7_herniation + D7_INR + D7_DNR + site_continent,  
##   family = binomial, data = train_data)
```

```

##
## Coefficients: (2 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -9.524e+00  3.936e+03  -0.002  0.9981
## age_at_consent  -4.142e-02  2.092e-02  -1.980  0.0477 *
## gcs_randomization -9.793e-02  1.176e-01  -0.833  0.4050
## nihss_randomization -2.998e-01  6.061e-02  -4.947  7.53e-07 ***
## stabct_ich_volume  -5.180e-02  2.220e-02  -2.334  0.0196 *
## stabct_ivh_volume  -1.222e-02  7.265e-02  -0.168  0.8665
## eot_less_151      8.254e-01  4.804e-01   1.718  0.0858 .
## BaselineNEWscore_BP1 -3.680e-01  1.361e+00  -0.270  0.7868
## BaselineNEWscore_BP2 -4.822e-01  1.373e+00  -0.351  0.7253
## BaselineNEWscore_BP3 -8.835e-02  1.409e+00  -0.063  0.9500
## BaselineNEWscore_BP4 -5.879e-01  2.201e+00  -0.267  0.7894
## BaselineNEWscore_BP5 -2.921e+00  1.876e+00  -1.558  0.1193
## Baseline_BP_control1  1.208e+00  9.772e-01   1.236  0.2164
## Baseline_BP_control2  6.231e-02  1.060e+00   0.059  0.9531
## Baseline_BP_control3      NA      NA      NA      NA
## Baseline_Hypotension1 -1.862e+01  4.128e+03  -0.005  0.9964
## Baseline_Hypotension2 -1.937e+01  1.001e+04  -0.002  0.9985
## Baseline_Hyperpyrexia1  6.599e-01  1.206e+00   0.547  0.5844
## Baseline_Hyperpyrexia2 -4.758e-04  8.821e+03   0.000  1.0000
## Baseline_Hyperpyrexia3 -1.724e+01  6.490e+03  -0.003  0.9979
## Baseline_Hyperglycemia1  7.613e-01  1.200e+00   0.634  0.5258
## Baseline_Hyperglycemia2  1.783e+00  1.668e+00   1.069  0.2852
## Baseline_Hyperglycemia3  1.589e+00  2.282e+00   0.696  0.4863
## Baseline_ICP1        -1.243e+01  5.252e+03  -0.002  0.9981
## Baseline_ICP2         4.152e+01  1.792e+04   0.002  0.9982
## Baseline_ICP3         5.101e+00  1.478e+04   0.000  0.9997
## Baseline_herniation1  -1.580e+01  5.936e+03  -0.003  0.9979
## Baseline_herniation2  -1.263e+01  6.079e+03  -0.002  0.9983
## Baseline_INR1        -5.873e-01  1.308e+00  -0.449  0.6534
## Baseline_INR2        -1.667e+01  5.472e+03  -0.003  0.9976
## Day7NEWscore_BP1     -1.270e+00  1.496e+00  -0.849  0.3960
## Day7NEWscore_BP2     -2.285e+00  1.582e+00  -1.444  0.1487
## Day7NEWscore_BP3     -5.887e-01  1.658e+00  -0.355  0.7226
## Day7NEWscore_BP4     -3.043e+00  2.217e+00  -1.372  0.1700
## Day7NEWscore_BP5     -2.030e+01  7.906e+03  -0.003  0.9980
## D7_BP_control1       -7.285e-01  7.586e-01  -0.960  0.3369
## D7_BP_control2       -3.774e-01  8.075e-01  -0.467  0.6402
## D7_BP_control3        NA      NA      NA      NA
## D7_Hypotension1      1.400e+00  1.247e+00   1.123  0.2616
## D7_Hypotension2      1.539e+00  1.969e+00   0.782  0.4343
## D7_Hyperpyrexia1     -1.027e+00  7.322e-01  -1.402  0.1608
## D7_Hyperpyrexia2     -1.884e+01  2.586e+03  -0.007  0.9942
## D7_Hyperpyrexia3     -1.885e+01  2.302e+03  -0.008  0.9935
## D7_Hyperglycemia1     1.102e-01  1.471e+00   0.075  0.9403
## D7_Hyperglycemia2    -2.095e+00  1.502e+00  -1.395  0.1630
## D7_Hyperglycemia3    -1.409e+00  2.220e+00  -0.635  0.5256
## D7_ICP1              -1.460e+01  3.965e+03  -0.004  0.9971
## D7_ICP2              -1.299e+01  5.799e+03  -0.002  0.9982
## D7_ICP3              -1.371e+01  5.593e+03  -0.002  0.9980
## D7_ICP4              -1.523e+01  1.773e+04  -0.001  0.9993
## D7_herniation1       8.844e-01  1.102e+04   0.000  0.9999

```

```

## D7_herniation2          -1.235e+01  5.384e+03 -0.002  0.9982
## D7_INR1                 -1.658e+01  4.337e+03 -0.004  0.9970
## D7_INR2                 -1.928e+01  4.023e+03 -0.005  0.9962
## D7_DNR1                 -1.398e+01  8.067e+03 -0.002  0.9986
## D7_DNR2                 -1.663e+01  1.185e+04 -0.001  0.9989
## D7_DNR3                 -1.628e+01  6.366e+03 -0.003  0.9980
## site_continentEurope    1.931e+01  3.936e+03  0.005  0.9961
## site_continentNorth America 2.030e+01  3.936e+03  0.005  0.9959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 280.45  on 379  degrees of freedom
## Residual deviance: 132.51  on 323  degrees of freedom
## AIC: 246.51
##
## Number of Fisher Scoring iterations: 19

## predict use test data
test_data$log_pred1=ifelse(predict(m_log, newdata = test_data, type = "response")<0.12, 0, 1)

conf_matrix_1 = table(Predicted = test_data$log_pred1, Actual = test_data$glasgow_rankin_0_3_30)

# Calculate PPV(Precision)
PPV_1 = conf_matrix_1["1", "1"] / sum(conf_matrix_1["1", ])
print(PPV_1)

## [1] 0.2105263

roc_curve_log = roc(test_data$glasgow_rankin_0_3_30, test_data$log_pred1)

## Setting levels: control = 0, case = 1

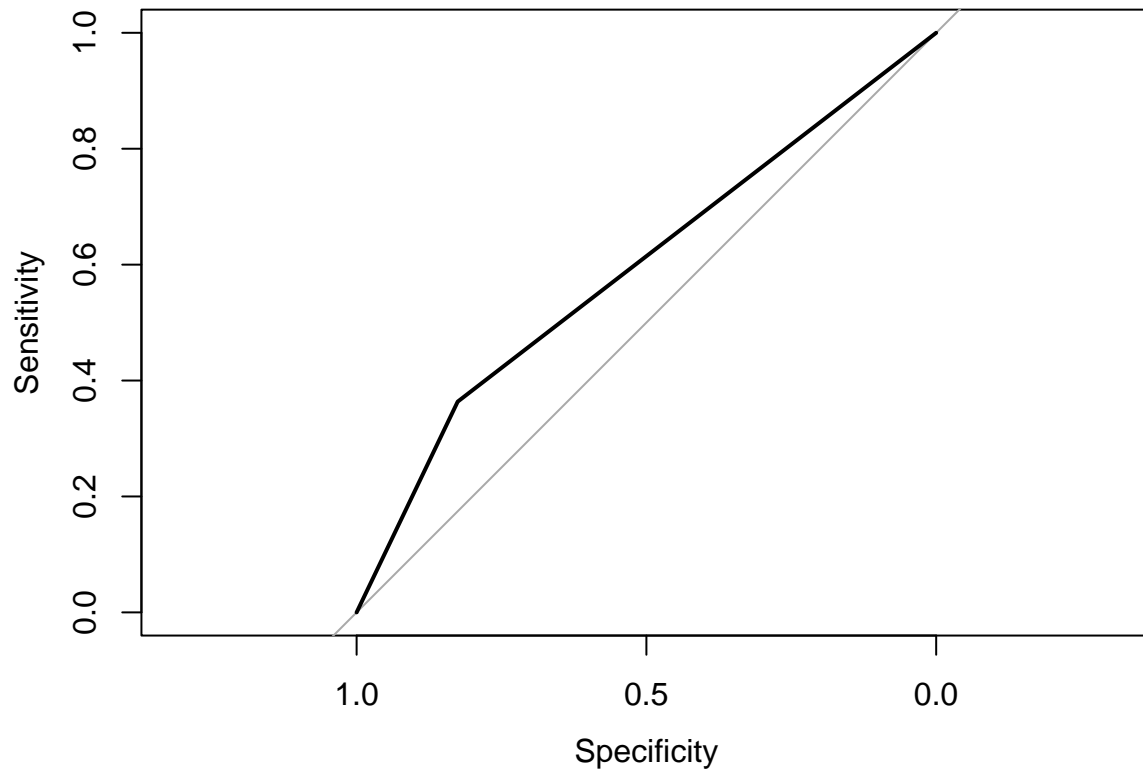
## Setting direction: controls < cases

auc(roc_curve_log)

## Area under the curve: 0.5946

plot(roc_curve_log)

```



```
## if use BaselineNEWscore * baseline_BP_control, is day7 also need to include interaction term?
## Why have NA? (baseline bp control_3,)
```

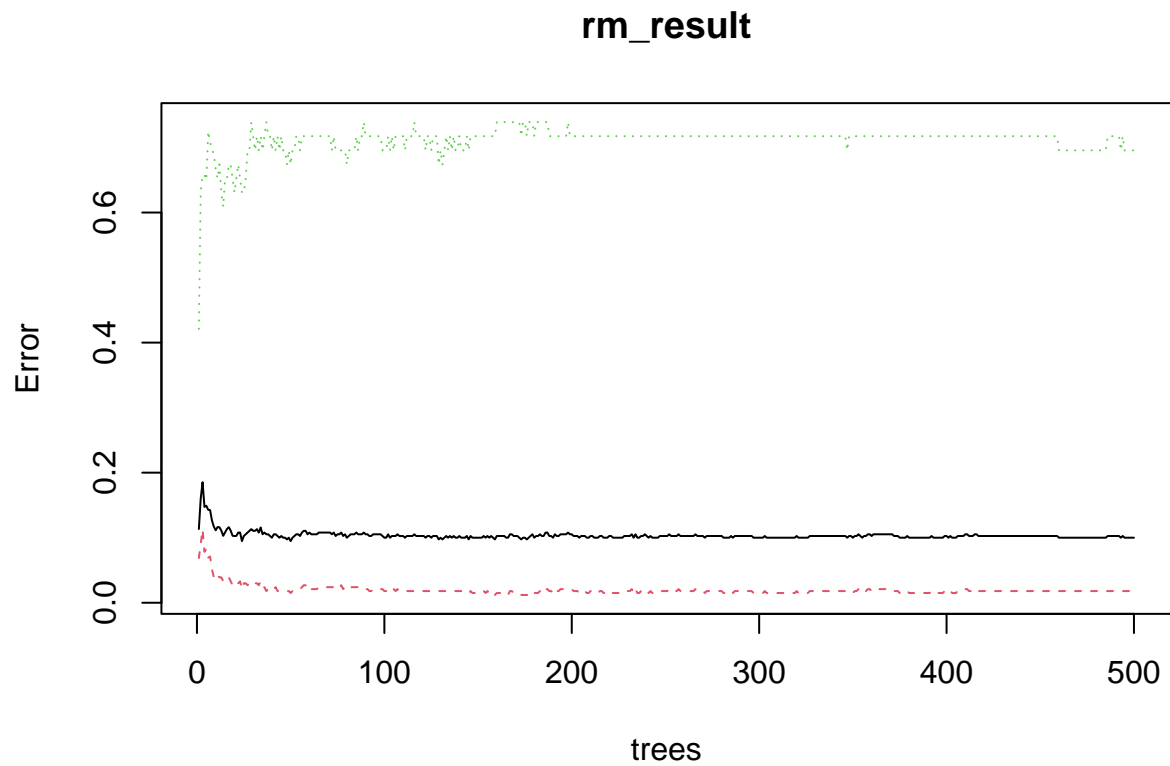
```
## random forest
set.seed(1234)
## number of tree 500, 200, 300
## include region or not
rm_result=randomForest(x=train_data[c(-1,-25)], y=train_data$glasgow_rankin_0_3_30, ntree = 500)

importance_table=importance(rm_result)
importance_table=importance_table[order(importance_table[, 1], decreasing = TRUE), ]
importance_table
```

## nihss_randomization	stabct_ich_volume	age_at_consent
## 18.99014827	10.31308140	8.13676944
## gcs_randomization	stabct_ivh_volume	BaselineNEWscore_BP
## 6.94426703	5.07124939	4.42361627
## Day7NEWscore_BP	Baseline_BP_control	D7_BP_control
## 4.09562626	3.06960050	2.80274639
## eot_less_15	D7_Hyperpyrexia	D7_Hyperglycemia
## 2.45430389	1.60379902	1.43423120
## Baseline_Hyperglycemia	site_continent	D7_Hypotension
## 1.42457439	1.09492059	0.87581130
## Baseline_ICP	Baseline_INR	Baseline_Hyperpyrexia
## 0.85426623	0.74878399	0.47268927

##	D7_INR	Baseline_Hypotension	D7_ICP
##	0.34810488	0.17403049	0.13678183
##	D7_DNR	Baseline_herniation	D7_herniation
##	0.09796139	0.03786649	0.02665590

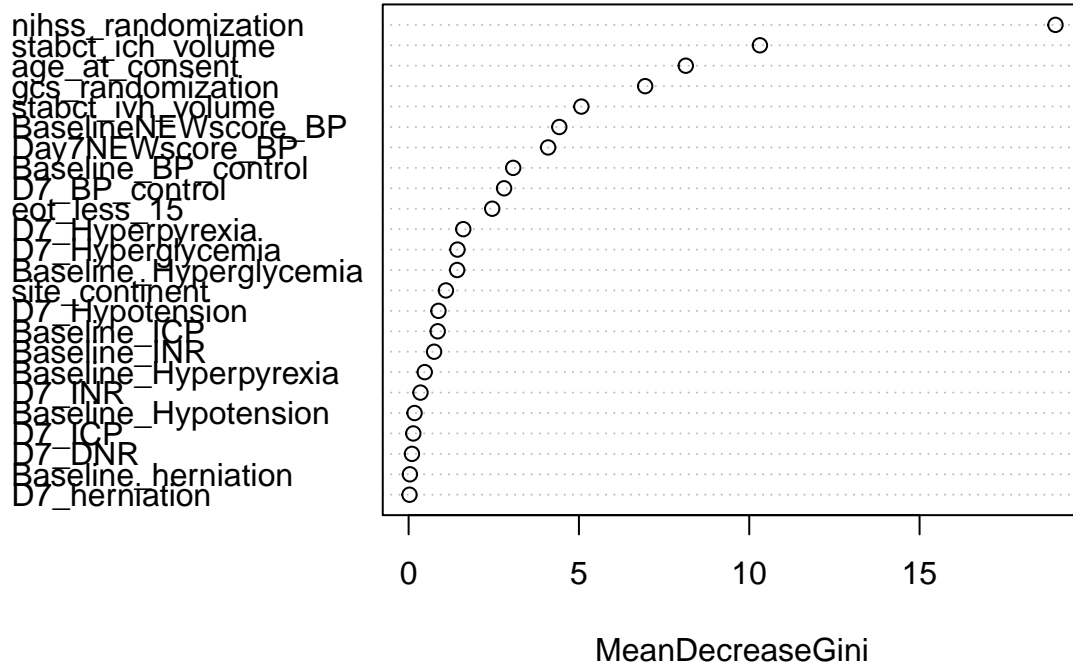
```
plot(rm_result)
```



```
varImpPlot(rm_result)
```



## rm\_result



```
## hernation doesn't seem so important OR maybe hypotension
test_data$rf_pred = predict(rm_result, newdata = test_data[c(-1,-25)])

conf_matrix_2 = table(Predicted = test_data$rf_pred, Actual = test_data$glasgow_rankin_0_3_30)

# Calculate PPV(Precision)
PPV_2 = conf_matrix_2["1", "1"] / sum(conf_matrix_2["1", ])
print(PPV_2)
```

```
## [1] 0.6666667
```

```
roc_curve_rf = roc(test_data$glasgow_rankin_0_3_30, as.numeric(test_data$rf_pred))
```

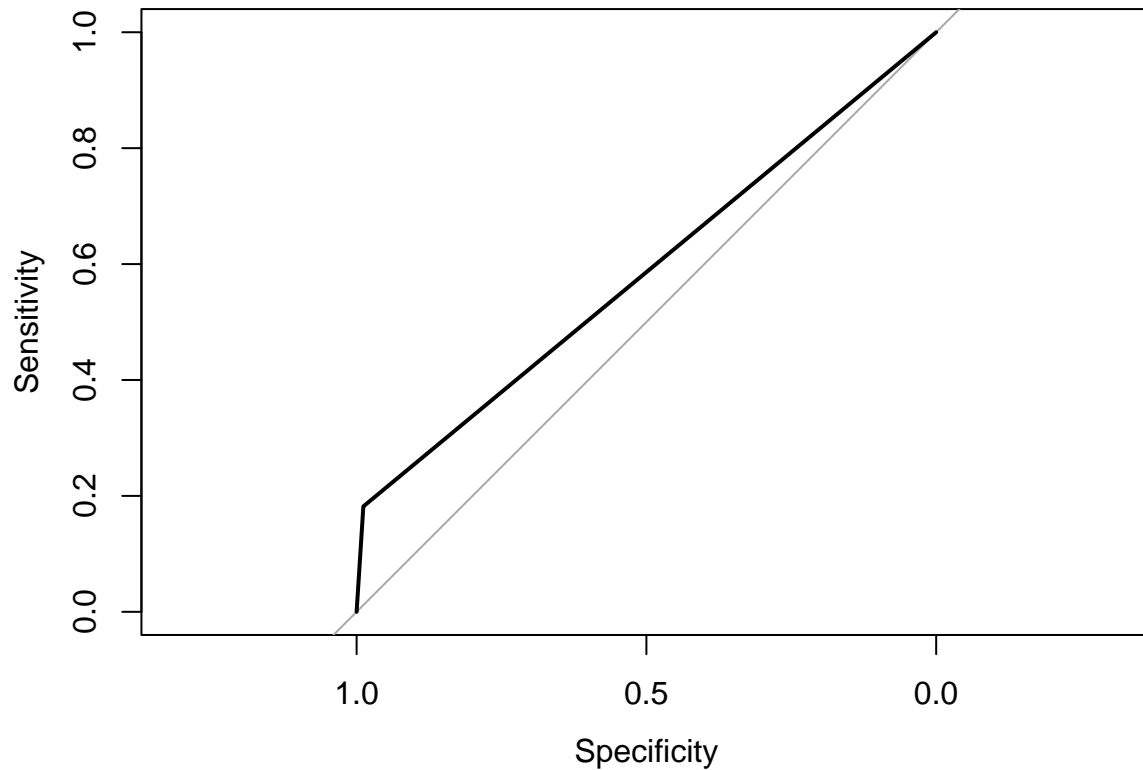
```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
auc(roc_curve_rf)
```

```
## Area under the curve: 0.5851
```

```
plot(roc_curve_rf)
```



```
## logistics regression
```

```
# m_log2=glmer(glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization + stabct_ich_volume + stabct_ivh_volume + eot_less_15 + BaselineNEWScore_BP + Baseline_BP_control + Baseline_Hyperglycemia + Day7NEWScore_BP + D7_BP_control + D7_Hyperpyrexia + D7_Hyperglycemia + site_continent, family = binomial, data = train_data)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(m_log2)
```

```
##
```

```
## Call:
```

```
## glm(formula = glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization + stabct_ich_volume + stabct_ivh_volume + eot_less_15 + BaselineNEWScore_BP + Baseline_BP_control + Baseline_Hyperglycemia + Day7NEWScore_BP + D7_BP_control + D7_Hyperpyrexia + D7_Hyperglycemia + site_continent, family = binomial, data = train_data)
```

```
##
```

```
## Coefficients: (2 not defined because of singularities)
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-10.56809	2461.71195	-0.004	0.9966
age_at_consent	-0.03082	0.01839	-1.676	0.0937
gcs_randomization	-0.04961	0.10733	-0.462	0.6439

```
## nihss_randomization      -0.27719    0.05223   -5.307 1.12e-07 ***
## stabct_ich_volume        -0.04364    0.01829   -2.386 0.0170 *
## stabct_ivh_volume        -0.02472    0.05947   -0.416 0.6777
## eot_less_151             0.86485    0.45045    1.920 0.0549 .
## BaselineNEWscore_BP1     -0.78519    1.28617   -0.610 0.5415
## BaselineNEWscore_BP2     -0.89746    1.30337   -0.689 0.4911
## BaselineNEWscore_BP3     -0.34782    1.32929   -0.262 0.7936
## BaselineNEWscore_BP4      0.07630    1.61158    0.047 0.9622
## BaselineNEWscore_BP5     -2.85575    1.76944   -1.614 0.1065
## Baseline_BP_control1      1.59692    0.85586    1.866 0.0621 .
## Baseline_BP_control2      0.74671    0.91031    0.820 0.4121
## Baseline_BP_control3      NA          NA          NA      NA
## Baseline_Hyperglycemia1   0.89039    1.04699    0.850 0.3951
## Baseline_Hyperglycemia2   1.00191    1.40081    0.715 0.4745
## Baseline_Hyperglycemia3   1.66643    1.97685    0.843 0.3992
## Day7NEWscore_BP1         -1.19784    1.43045   -0.837 0.4024
## Day7NEWscore_BP2         -2.32538    1.50100   -1.549 0.1213
## Day7NEWscore_BP3         -1.36661    1.55845   -0.877 0.3805
## Day7NEWscore_BP4         -3.88237    2.12664   -1.826 0.0679 .
## Day7NEWscore_BP5        -18.85351  4868.69451  -0.004 0.9969
## D7_BP_control1           -0.72487    0.65707   -1.103 0.2699
## D7_BP_control2           -0.46412    0.71129   -0.653 0.5141
## D7_BP_control3           NA          NA          NA      NA
## D7_Hyperpyrexia1         -0.84561    0.65379   -1.293 0.1959
## D7_Hyperpyrexia2         -1.28023    1.26617   -1.011 0.3120
## D7_Hyperpyrexia3        -17.02350  1458.42807  -0.012 0.9907
## D7_Hyperglycemia1         0.18769    1.45702    0.129 0.8975
## D7_Hyperglycemia2        -1.78570    1.27929   -1.396 0.1628
## D7_Hyperglycemia3        -1.72936    1.90331   -0.909 0.3636
## site_continentEurope      18.23605  2461.71072    0.007 0.9941
## site_continentNorth America 19.41400  2461.71067    0.008 0.9937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 280.45  on 379  degrees of freedom
## Residual deviance: 153.60  on 348  degrees of freedom
## AIC: 217.6
##
## Number of Fisher Scoring iterations: 18
```

```
test_data$log_pred2=ifelse(predict(m_log2, newdata = test_data, type = "response")<0.12, 0, 1)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
conf_matrix_3 = table(Predicted = test_data$log_pred2, Actual = test_data$glasgow_rankin_0_3_30)
```

```
# Calculate PPV(Precision)
```

```
PPV_3 = conf_matrix_3["1", "1"] / sum(conf_matrix_3["1", ])
print(PPV_3)
```

```
## [1] 0.2173913
```

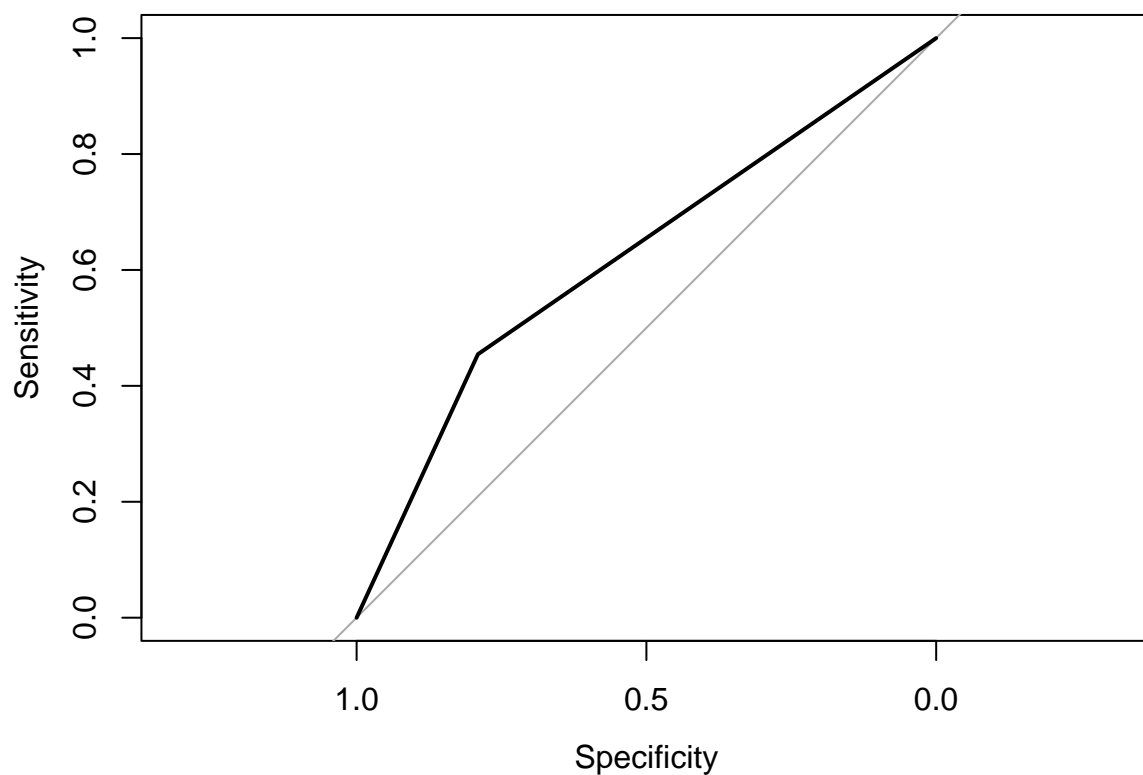
```
roc_curve_log2 = roc(test_data$glasgow_rankin_0_3_30, test_data$log_pred2)
```

```
## Setting levels: control = 0, case = 1  
## Setting direction: controls < cases
```

```
auc(roc_curve_log2)
```

```
## Area under the curve: 0.6226
```

```
plot(roc_curve_log2)
```



```
## LASSO
```

```
lasso_x = model.matrix(glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization)
```

```
m_lasso=cv.glmnet(lasso_x,train_data$glasgow_rankin_0_3_30,family="binomial", alpha=1)
```

```
## select best
```

```
best_lambda = m_lasso$lambda.1se
```

```
coef(m_lasso)
```

```
## 74 x 1 sparse Matrix of class "dgCMatrix"
```

##	s1
## (Intercept)	-0.38768735
## age_at_consent	.
## gcs_randomization	.
## nihss_randomization	-0.09029707
## stabct_ich_volume	.
## stabct_ivh_volume	.
## eot_less_151	.
## BaselineNEWscore_BP1	.
## BaselineNEWscore_BP2	.
## BaselineNEWscore_BP3	.
## BaselineNEWscore_BP4	.
## BaselineNEWscore_BP5	.
## Baseline_BP_control1	.
## Baseline_BP_control2	.
## Baseline_BP_control3	.
## Baseline_Hypotension1	.
## Baseline_Hypotension2	.
## Baseline_Hyperpyrexia1	.
## Baseline_Hyperpyrexia2	.
## Baseline_Hyperpyrexia3	.
## Baseline_Hyperglycemia1	.
## Baseline_Hyperglycemia2	.
## Baseline_Hyperglycemia3	.
## Baseline_ICP1	.
## Baseline_ICP2	.
## Baseline_ICP3	.
## Baseline_herniation1	.
## Baseline_herniation2	.
## Baseline_INR1	.
## Baseline_INR2	.
## Day7NEWscore_BP1	.
## Day7NEWscore_BP2	.
## Day7NEWscore_BP3	.
## Day7NEWscore_BP4	.
## Day7NEWscore_BP5	.
## D7_BP_control1	.
## D7_BP_control2	.
## D7_BP_control3	.
## D7_Hypotension1	.
## D7_Hypotension2	.
## D7_Hyperpyrexia1	.
## D7_Hyperpyrexia2	.
## D7_Hyperpyrexia3	.
## D7_Hyperglycemia1	.
## D7_Hyperglycemia2	.
## D7_Hyperglycemia3	.
## D7_ICP1	.
## D7_ICP2	.
## D7_ICP3	.
## D7_ICP4	.
## D7_herniation1	.
## D7_herniation2	.
## D7_INR1	.

```
## D7_INR2 .
## D7_DNR1 .
## D7_DNR2 .
## D7_DNR3 .
## site_continentEurope .
## site_continentNorth America .
## BaselineNEWScore_BP1:Baseline_BP_control1 .
## BaselineNEWScore_BP2:Baseline_BP_control1 .
## BaselineNEWScore_BP3:Baseline_BP_control1 .
## BaselineNEWScore_BP4:Baseline_BP_control1 .
## BaselineNEWScore_BP5:Baseline_BP_control1 .
## BaselineNEWScore_BP1:Baseline_BP_control2 .
## BaselineNEWScore_BP2:Baseline_BP_control2 .
## BaselineNEWScore_BP3:Baseline_BP_control2 .
## BaselineNEWScore_BP4:Baseline_BP_control2 .
## BaselineNEWScore_BP5:Baseline_BP_control2 .
## BaselineNEWScore_BP1:Baseline_BP_control3 .
## BaselineNEWScore_BP2:Baseline_BP_control3 .
## BaselineNEWScore_BP3:Baseline_BP_control3 .
## BaselineNEWScore_BP4:Baseline_BP_control3 .
## BaselineNEWScore_BP5:Baseline_BP_control3 .
```

```
summary(m_lasso)
```

```
##          Length Class  Mode
## lambda      100    -none- numeric
## cvm          100    -none- numeric
## cvsd         100    -none- numeric
## cvup         100    -none- numeric
## cvlo         100    -none- numeric
## nzero        100    -none- numeric
## call         5     -none- call
## name         1     -none- character
## glmnet.fit   13     lognet list
## lambda.min   1     -none- numeric
## lambda.1se   1     -none- numeric
## index        2     -none- numeric
```

```
lasso_test=model.matrix(glasgow_rankin_0_3_30 ~ age_at_consent + gcs_randomization + nihss_randomization)
test_data$lasso_pred = ifelse(predict(m_lasso, lasso_test, lambda=best_lambda, type = "response")<0.12, 1, 0)
conf_matrix_4 = table(Predicted = test_data$lasso_pred, Actual = test_data$glasgow_rankin_0_3_30)

# Calculate PPV(Precision)
PPV_4 = conf_matrix_4["1", "1"] / sum(conf_matrix_4["1", ])
print(PPV_4)
```

```
## [1] 0.2777778
```

```
roc_curve_lasso = roc(test_data$glasgow_rankin_0_3_30, test_data$lasso_pred)
```

```
## Setting levels: control = 0, case = 1
```

```
## Warning in roc.default(test_data$glasgow_rankin_0_3_30, test_data$lasso_pred):  
## Deprecated use a matrix as predictor. Unexpected results may be produced,  
## please pass a numeric vector.
```

```
## Setting direction: controls < cases
```

```
auc(roc_curve_lasso)
```

```
## Area under the curve: 0.8034
```

```
plot(roc_curve_lasso)
```

