Data Science Final Project

Jobelle Sorino-Simblante

December 16, 2022

Context

Irrespective of whether or not data nd imaes are stored/analyzed in a centralized manner, variability in scanner models, acquisition protocols and reconstruction settings are unavoidable in the current clinical practice. Yet radiomics are notoriously sensitive to such protocol variations. Hence, there is a clear need for the harmonization of features in order to allow consistent findings in radiomics multicenter studies.

Objective

The objective of this project is to develop different models to predic failure (endpoint) of the radiomics signature based from MRI, PET and CT scans.

Needed Packages

These are the needed packages in this activity:

```
library(readr)
library(dplyr)
library(ggplot2)
library(stringr)
library(recipes)
library(rsample)
library(xgboost)
library(gbm)
library(rpart)
library(rpart.plot)
library(ROCR)
library(pROC)
library(gridExtra)
library(tidyverse)
library(cluster)
library(factoextra)
library(caret)
library(keras)
library(tfruns)
library(tensorflow)
library(tfestimators)
library(mclust)
```

Dataset

Radiomics dataset has 431 variables with 197 observations.

```
radiomics = read.csv("radiomics_completedata.csv",header = TRUE, sep = ",")
attach(radiomics)
str(radiomics)
```

```
## 'data.frame':    197 obs. of  431 variables:
##  $ Institution                 : chr  "A" "A" "A" "A" ...
##  $ Failure.binary              : int  0 1 0 1 0 1 0 0 1 1 ...
##  $ Failure                     : num  49.3 12.6 79.8 17.9 39.6 ...
##  $ Entropy_cooc.W.ADC          : num  12.9 12.2 12.8 13.5 12.6 ...
##  $ GLNU_align.H.PET            : num  46.3 27.5 90.2 325.6 89.6 ...
##  $ Min_hist.PET                : num  6.25 11.01 2.78 6.3 3.58 ...
##  $ Max_hist.PET                : num  17.83 26.47 6.88 22.03 7.92 ...
##  $ Mean_hist.PET               : num  9.78 15.43 4.3 10.33 4.45 ...
##  $ Variance_hist.PET           : num  6.814 12.932 0.923 6.65 0.572 ...
##  $ Standard_Deviation_hist.PET: num  2.612 3.598 0.962 2.581 0.757 ...
##  $ Skewness_hist.PET           : num  0.689 0.79 0.249 0.832 1.575 ...
##  $ Kurtosis_hist.PET           : num  -0.34 -0.32 -0.944 0.856 3.25 ...
##  $ Energy_hist.PET             : num  0.00509 0.0063 0.00502 0.00329 0.00807 ...
##  $ Entropy_hist.PET            : num  9.63 8.07 9.67 10.57 7.62 ...
##  $ AUC_hist.PET                : num  0.507 0.508 0.503 0.544 0.544 ...
##  $ H_suv.PET                   : num  1.124 1.927 0.411 0.92 0.306 ...
##  $ Volume.PET                  : num  13752 9328 26624 51058 29415 ...
##  $ X3D_surface.PET             : num  5623 8357 16832 29100 7769 ...
##  $ ratio_3ds_vol.PET           : num  3.21 4.85 3.16 2.03 4.82 ...
##  $ ratio_3ds_vol_norm.PET      : num  15.9 21.1 19.5 20.1 21 ...
##  $ irregularity.PET            : num  2.21 2.35 2.12 1.86 2.22 ...
##  $ tumor_length.PET            : num  44 39.4 50.9 76.2 36.9 ...
##  $ Compactness_v1.PET          : num  0.00337 0.00308 0.00314 0.00312 0.00308 ...
##  $ Compactness_v2.PET          : num  0.00278 0.00264 0.00266 0.00265 0.00264 ...
##  $ Spherical_disproportion.PET: num  15.9 21.1 19.5 20.1 21 ...
##  $ Sphericity.PET              : num  0.0654 0.0499 0.0538 0.0522 0.0501 ...
##  $ Asphericity.PET             : num  14.9 20.1 18.5 19.1 20 ...
##  $ Center_of_mass.PET          : num  0.811 0.588 0.393 0.867 0.526 ...
##  $ Max_3D_diam.PET             : num  44 39.4 50.9 76.2 36.9 ...
##  $ Major_axis_length.PET       : num  34.6 35.1 48.1 64.1 36 ...
##  $ Minor_axis_length.PET       : num  25.9 27.3 30.4 54.5 23.8 ...
##  $ Least_axis_length.PET       : num  25 21.2 27.5 51.6 21.4 ...
##  $ Elongation.PET              : num  0.751 0.78 0.634 0.852 0.665 ...
##  $ Flatness.PET                : num  0.725 0.605 0.574 0.807 0.597 ...
##  $ Max_cooc.L.PET              : num  0.00502 0.00819 0.00503 0.00597 0.00755 ...
##  $ Average_cooc.L.PET          : num  22.9 21.9 27.3 17.8 15.4 ...
##  $ Variance_cooc.L.PET         : num  206 227 209 103 142 ...
##  $ Entropy_cooc.L.PET          : num  10.69 10.29 10.88 10.24 9.83 ...
##  $ DAVE_cooc.L.PET             : num  11.86 13.99 12.28 7.47 10.24 ...
##  $ DVAR_cooc.L.PET             : num  84.2 129.4 85.3 43.9 79.4 ...
##  $ DENT_cooc.L.PET             : num  5 5.21 5 4.38 4.8 ...
##  $ SAVE_cooc.L.PET             : num  45.8 43.8 54.5 35.6 30.7 ...
##  $ SVAR_cooc.L.PET             : num  588 581 600 311 385 ...
##  $ SENT_cooc.L.PET             : num  6.53 6.49 6.59 6.11 6.05 ...
##  $ ASM_cooc.L.PET              : num  0.0033 0.0036 0.0032 0.00368 0.004 ...
```

```
##  $ Contrast_cooc.L.PET        : num  234.8 325.1 236.1 99.8 184.2 ...
##  $ Dissimilarity_cooc.L.PET   : num  11.86 13.99 12.28 7.47 10.24 ...
##  $ Inv_diff_cooc.L.PET        : num  0.166 0.156 0.154 0.229 0.189 ...
##  $ Inv_diff_norm_cooc.L.PET   : num  0.859 0.839 0.853 0.905 0.876 ...
##  $ IDM_cooc.L.PET             : num  0.0889 0.0854 0.079 0.1416 0.1083 ...
##  $ IDM_norm_cooc.L.PET        : num  0.954 0.938 0.953 0.98 0.964 ...
##  $ Inv_var_cooc.L.PET         : num  0.0913 0.0875 0.0846 0.1498 0.1144 ...
##  $ Correlation_cooc.L.PET     : num  0.432 0.285 0.438 0.517 0.355 ...
##  $ Autocorrelation_cooc.L.PET : num  612 544 833 370 286 ...
##  $ Tendency_cooc.L.PET        : num  588 581 600 311 385 ...
##  $ Shade_cooc.L.PET           : num  6860 4692 403 3806 9785 ...
##  $ Prominence_cooc.L.PET      : num  869822 803735 800130 345453 743501 ...
##  $ IC1_.L.PET                 : num  -0.084 -0.0967 -0.0724 -0.0503 -0.0707 ...
##  $ IC2_.L.PET                 : num  0.79 0.814 0.758 0.655 0.728 ...
##  $ Coarseness_vdif_.L.PET     : num  0.01432 0.0142 0.01627 0.00494 0.01724 ...
##  $ Contrast_vdif_.L.PET       : num  1.021 1.51 1.014 0.306 0.854 ...
##  $ Busyness_vdif_.L.PET       : num  0.0874 0.0802 0.0575 0.3927 0.082 ...
##  $ Complexity_vdif_.L.PET     : num  17053 21289 15200 10762 16797 ...
##  $ Strength_vdif_.L.PET       : num  27.4 35.76 24.45 5.55 57.04 ...
##  $ SRE_align.L.PET            : num  0.987 0.99 0.989 0.973 0.986 ...
##  $ LRE_align.L.PET            : num  1.07 1.06 1.06 1.13 1.07 ...
##  $ GLNU_align.L.PET           : num  10.16 8.42 9.12 94.57 10.57 ...
##  $ RLNU_align.L.PET           : num  384 263 395 2941 262 ...
##  $ RP_align.L.PET             : num  0.981 0.985 0.985 0.964 0.981 ...
##  $ LGRE_align.L.PET           : num  0.0637 0.0658 0.0392 0.0481 0.0917 ...
##  $ HGRE_align.L.PET           : num  590 560 781 387 296 ...
##  $ LGSRE_align.L.PET          : num  0.0625 0.0642 0.0388 0.0466 0.0902 ...
##  $ HGSRE_align.L.PET          : num  581 555 768 377 292 ...
##  $ LGHRE_align.L.PET          : num  0.0687 0.0724 0.041 0.0544 0.0978 ...
##  $ HGLRE_align.L.PET          : num  632 584 836 428 309 ...
##  $ GLNU_norm_align.L.PET      : num  0.0279 0.0334 0.0248 0.0323 0.0411 ...
##  $ RLNU_norm_align.L.PET      : num  0.961 0.97 0.968 0.929 0.96 ...
##  $ GLVAR_align.L.PET          : num  202 215 217 108 121 ...
##  $ RLVAR_align.L.PET          : num  0.0259 0.0215 0.0208 0.0464 0.0245 ...
##  $ Entropy_align.L.PET        : num  5.59 5.39 5.7 5.48 5.05 ...
##  $ SZSE.L.PET                 : num  0.927 0.961 0.974 0.906 0.966 ...
##  $ LZSE.L.PET                 : num  1.38 1.24 1.11 1.62 1.15 ...
##  $ LGLZE.L.PET                : num  0.0623 0.0648 0.0405 0.048 0.0933 ...
##  $ HGLZE.L.PET                : num  593 567 770 394 301 ...
##  $ SZLGE.L.PET                : num  0.0561 0.0606 0.0404 0.0433 0.0911 ...
##  $ SZHGE.L.PET                : num  554 546 736 361 296 ...
##  $ LZLGE.L.PET                : num  0.09 0.0865 0.0407 0.0768 0.1018 ...
##  $ LZHGE.L.PET                : num  832 650 905 591 322 ...
##  $ GLNU_area.L.PET            : num  9.17 7.82 8.88 83.35 10.25 ...
##  $ ZSNU.L.PET                 : num  301 233 372 2206 242 ...
##  $ ZSP.L.PET                  : num  0.9 0.941 0.966 0.861 0.956 ...
##  $ GLNU_norm.L.PET            : num  0.0275 0.0326 0.0247 0.0319 0.0409 ...
##  $ ZSNU_norm.L.PET            : num  0.823 0.9 0.931 0.781 0.91 ...
##  $ GLVAR_area.L.PET           : num  202 214 216 110 124 ...
##  $ ZSVAR.L.PET                : num  0.142 0.1098 0.0385 0.2592 0.0488 ...
##  $ Entropy_area.L.PET         : num  5.89 5.55 5.78 5.9 5.16 ...
##  $ Max_cooc.H.PET             : num  0.0312 0.0436 0.1694 0.0402 0.4235 ...
##  $ Average_cooc.H.PET         : num  39.9 39.2 44.9 38.2 49.5 ...
##  $ Variance_cooc.H.PET        : num  255.3 259.2 226.9 276.5 65.5 ...
```

```
##   [list output truncated]
```

```r
head(radiomics[1:5])
```

```
##   Institution Failure.binary  Failure Entropy_cooc.W.ADC GLNU_align.H.PET
## 1           A              0 49.30000           12.85352         46.25635
## 2           A              1 12.56667           12.21115         27.45454
## 3           A              0 79.80000           12.75682         90.19570
## 4           A              1 17.86667           13.46730        325.64333
## 5           A              0 39.56667           12.63733         89.57904
## 6           A              1  4.76667           13.16159        101.71345
```

Preprocess the data

```r
# Check for null and missing values
any(is.na(radiomics))
```

```
## [1] FALSE
```

```r
# Check for normality, if not, normalized the data
shapiro.test(Entropy_cooc.W.ADC) # Entropy_cooc.W.ADC is normally distributed
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Entropy_cooc.W.ADC
## W = 0.98903, p-value = 0.135
```

```r
shapiro.test(GLNU_align.H.PET)   # GLNU_align.H.PET is not normally distributed
```

```
##
##  Shapiro-Wilk normality test
##
## data:  GLNU_align.H.PET
## W = 0.76271, p-value < 2.2e-16
```

```r
shapiro.test(Min_hist.PET)       # Min_hist.PET is not normally distributed
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Min_hist.PET
## W = 0.91623, p-value = 3.821e-09
```

```r
# Since some of the variables are not normally distributed, then we will
# normalized it by using scale() function
radiomics_df <- as.data.frame(scale(select(radiomics, -c("Institution",
                                                         "Failure.binary" ))))
head(radiomics_df[1:5])
```

```
##      Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET Max_hist.PET
## 1  1.1985789         0.55290547      -0.57063689   -0.4541408   -0.4361311
## 2 -0.7212472        -0.06486729      -0.78903636    0.4998369    0.1486951
## 3  2.7926271         0.45990825      -0.06024275   -1.1504338   -1.1768823
## 4 -0.4442487         1.14318298       2.67468822   -0.4446190   -0.1516658
## 5  0.6898772         0.34499368      -0.06740573   -0.9887407   -1.1061760
## 6 -1.1289054         0.84917904       0.07354603   -1.1864923   -1.2223057
```

```
# Get the correlation of the whole data except the categorical variables
cor.radiomics_df= cor(radiomics_df)
corr = round(cor.radiomics_df,2) # 2 decimals
head(corr[1:4 ,1:3])
```

```
##                    Failure Entropy_cooc.W.ADC GLNU_align.H.PET
## Failure               1.00              -0.35            -0.23
## Entropy_cooc.W.ADC   -0.35               1.00             0.39
## GLNU_align.H.PET     -0.23               0.39             1.00
## Min_hist.PET         -0.12               0.02            -0.03
```

```
corMatrix =  cor(radiomics_df, y = NULL, use = "ev")
highly_correlated_columns = findCorrelation(
  corMatrix,
  cutoff = 0.95, # correlation coefficient
  verbose = FALSE,
  names = FALSE,
  exact = TRUE
)
df <- radiomics_df[, -highly_correlated_columns]

# Final Radiomics Data
final_radiomics <- cbind(radiomics['Failure.binary'], df)
head(final_radiomics[1:4])
```

```
##   Failure.binary    Failure Entropy_cooc.W.ADC GLNU_align.H.PET
## 1              0  1.1985789         0.55290547      -0.57063689
## 2              1 -0.7212472        -0.06486729      -0.78903636
## 3              0  2.7926271         0.45990825      -0.06024275
## 4              1 -0.4442487         1.14318298       2.67468822
## 5              0  0.6898772         0.34499368      -0.06740573
## 6              1 -1.1289054         0.84917904       0.07354603
```

```
final_radiomics$Failure.binary <- as.factor(final_radiomics$Failure.binary)
str(final_radiomics)
```

```
## 'data.frame':    197 obs. of  148 variables:
##  $ Failure.binary            : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 1 1 2 2 ...
##  $ Failure                   : num  1.199 -0.721 2.793 -0.444 0.69 ...
##  $ Entropy_cooc.W.ADC        : num  0.5529 -0.0649 0.4599 1.1432 0.345 ...
##  $ GLNU_align.H.PET          : num  -0.5706 -0.789 -0.0602 2.6747 -0.0674 ...
##  $ Min_hist.PET              : num  -0.454 0.5 -1.15 -0.445 -0.989 ...
##  $ Skewness_hist.PET         : num  -0.323 -0.177 -0.959 -0.116 0.958 ...
##  $ Kurtosis_hist.PET         : num  -0.273 -0.266 -0.472 0.12 0.907 ...
```

```
##  $ Entropy_hist.PET           : num  -0.38 -0.747 -0.37 -0.157 -0.853 ...
##  $ Volume.PET                 : num  -0.7713 -0.8698 -0.4849 0.0587 -0.4229 ...
##  $ X3D_surface.PET            : num  -0.52 -0.431 -0.155 0.244 -0.45 ...
##  $ ratio_3ds_vol.PET          : num  -0.228 0.422 -0.248 -0.701 0.409 ...
##  $ tumor_length.PET           : num  -0.499 -0.625 -0.314 0.368 -0.691 ...
##  $ Compactness_v1.PET         : num  -0.072 -0.0845 -0.0816 -0.0828 -0.0844 ...
##  $ Sphericity.PET             : num  -0.443 -0.505 -0.49 -0.496 -0.504 ...
##  $ Asphericity.PET            : num  -0.3646 0.0201 -0.0967 -0.0517 0.0143 ...
##  $ Center_of_mass.PET         : num  -0.0305 -0.3264 -0.5841 0.0433 -0.4082 ...
##  $ Max_3D_diam.PET            : num  -0.6641 -0.7524 -0.5337 -0.0528 -0.7991 ...
##  $ Minor_axis_length.PET      : num  -0.81 -0.749 -0.616 0.43 -0.899 ...
##  $ Least_axis_length.PET      : num  -0.553 -0.74 -0.43 0.74 -0.728 ...
##  $ Elongation.PET             : num  -0.377 -0.3 -0.683 -0.111 -0.601 ...
##  $ Flatness.PET               : num  0.0389 -0.3472 -0.4444 0.3031 -0.3724 ...
##  $ Variance_cooc.L.PET        : num  -0.1075 0.0906 -0.0764 -1.0807 -0.7069 ...
##  $ DVAR_cooc.L.PET            : num  -0.438 0.284 -0.42 -1.081 -0.515 ...
##  $ DENT_cooc.L.PET            : num  -0.489 -0.392 -0.485 -0.774 -0.58 ...
##  $ Contrast_cooc.L.PET        : num  -0.221 0.302 -0.214 -1.004 -0.515 ...
##  $ IDM_cooc.L.PET             : num  -0.53 -0.577 -0.66 0.158 -0.277 ...
##  $ Shade_cooc.L.PET           : num  0.167 -0.248 -1.069 -0.418 0.727 ...
##  $ Prominence_cooc.L.PET      : num  0.031 -0.0979 -0.1049 -0.9915 -0.2153 ...
##  $ IC1_.L.PET                 : num  0.2871 0.0714 0.4831 0.8565 0.5117 ...
##  $ IC2_.L.PET                 : num  -0.339 -0.27 -0.427 -0.716 -0.512 ...
##  $ Contrast_vdif_.L.PET       : num  -0.2003 0.0485 -0.204 -0.5642 -0.2854 ...
##  $ Complexity_vdif_.L.PET     : num  -0.266 0.166 -0.455 -0.908 -0.292 ...
##  $ Strength_vdif_.L.PET       : num  -0.2699 -0.0894 -0.3336 -0.7416 0.3698 ...
##  $ LGHRE_align.L.PET          : num  -0.159 -0.104 -0.566 -0.37 0.268 ...
##  $ GLNU_norm_align.L.PET      : num  -0.2387 -0.0911 -0.321 -0.121 0.114 ...
##  $ RLVAR_align.L.PET          : num  -0.261 -0.377 -0.393 0.272 -0.298 ...
##  $ LZSE.L.PET                 : num  -0.448 -0.615 -0.77 -0.168 -0.73 ...
##  $ HGLZE.L.PET                : num  -0.298 -0.373 0.214 -0.874 -1.142 ...
##  $ LZLGE.L.PET                : num  -0.154 -0.1898 -0.6707 -0.2921 -0.0298 ...
##  $ LZHGE.L.PET                : num  -0.1861 -0.5424 -0.0428 -0.6587 -1.1883 ...
##  $ ZSVAR.L.PET                : num  -0.223 -0.414 -0.836 0.472 -0.775 ...
##  $ Max_cooc.H.PET             : num  -0.562 -0.464 0.534 -0.491 2.549 ...
##  $ Entropy_cooc.H.PET         : num  -0.441 -0.198 -1.23 -0.482 -1.474 ...
##  $ DENT_cooc.H.PET            : num  0.0819 -0.8326 -0.015 -0.0686 -0.2714 ...
##  $ SVAR_cooc.H.PET            : num  -0.2112 -0.5177 -0.049 -0.0605 -0.2237 ...
##  $ SENT_cooc.H.PET            : num  0.0703 0.2185 -0.7391 0.0341 -0.9923 ...
##  $ Contrast_cooc.H.PET        : num  -0.415 -0.106 -0.561 -0.55 -1.55 ...
##  $ Inv_var_cooc_.H.PET        : num  0.125 0.163 -0.42 0.183 -0.152 ...
##  $ Autocorrelation_cooc.H.PET : num  -0.636 -0.73 -0.128 -0.759 0.316 ...
##  $ Shade_cooc.H.PET           : num  0.5612 -0.0321 -0.0644 -0.3905 1.5498 ...
##  $ Prominence_cooc.H.PET      : num  -0.277 -0.383 -0.722 0.327 -1.726 ...
##  $ IC1_d.H.PET                : num  0.4584 0.841 0.0806 -0.0258 0.4427 ...
##  $ Contrast_vdif.H.PET        : num  -0.427 -0.567 0.723 -0.484 -0.542 ...
##  $ Busyness_vdif.H.PET        : num  -0.364 -0.37 -0.348 -0.247 -0.367 ...
##  $ Complexity_vdif.H.PET      : num  -0.1093 0.0616 -0.1995 -0.2352 -0.7292 ...
##  $ Strength_vdif.H.PET        : num  -0.1303 -0.0926 -0.1141 -0.2392 0.0872 ...
##  $ LRE_align.H.PET            : num  -0.72 -0.907 0.382 -0.465 0.602 ...
##  $ RLNU_align.H.PET           : num  -0.497 -0.542 -0.585 0.719 -0.632 ...
##  $ HGLRE_align.H.PET          : num  -0.68 -0.857 0.687 -0.498 1.027 ...
##  $ RLVAR_align.H.PET          : num  -0.583 -0.804 0.744 -0.262 0.96 ...
##  $ HGLZE.H.PET                : num  -0.29 -0.783 -0.382 0.527 0.726 ...
```

```
##  $ SZHGE.H.PET                  : num  -3.66e-01 -9.23e-02 -9.77e-01 -5.58e-01 -2.87e-05 ...
##  $ LZLGE.H.PET                  : num  -0.254 -0.287 -0.201 -0.038 -0.12 ...
##  $ LZHGE.H.PET                  : num  -0.23385 -0.24389 -0.09548 -0.18688 0.00664 ...
##  $ GLNU_area.H.PET              : num  -0.544 -0.58 -0.429 0.539 -0.581 ...
##  $ ZSP.H.PET                    : num  -0.225 0.513 -0.929 -0.613 -1.116 ...
##  $ GLVAR_area.H.PET             : num  -0.422 -0.46 -0.732 -0.101 -1.748 ...
##  $ ASM_cooc.W.PET               : num  -0.201 -0.233 0.332 -0.189 1.229 ...
##  $ Contrast_cooc.W.PET          : num  -0.308 0.774 -0.958 -0.47 -0.971 ...
##  $ Dissimilarity_cooc.W.PET     : num  -0.254 0.536 -1.134 -0.455 -1.203 ...
##  $ Correlation_cooc.W.PET       : num  -0.24 -0.827 -0.225 0.117 -0.601 ...
##  $ Shade_cooc.W.PET             : num  -0.1939 -0.0771 -0.3808 -0.1221 -0.3673 ...
##  $ Coarseness_vdif.W.PET        : num  -0.055 -0.0353 0.0154 -0.311 0.0258 ...
##  $ Contrast_vdif.W.PET          : num  -0.185 0.981 -0.88 -0.8 -1.009 ...
##  $ Busyness_vdif.W.PET          : num  -0.698 -0.841 0.336 -0.297 0.717 ...
##  $ Complexity_vdif.W.PET        : num  -0.395 0.0832 -0.6695 -0.2371 -0.6679 ...
##  $ Strength_vdif.W.PET          : num  -0.149 0.434 -0.598 -0.483 -0.519 ...
##  $ LRE_align.W.PET              : num  -0.7391 -0.8573 -0.0674 -0.5816 0.0205 ...
##  $ GLNU_align.W.PET             : num  -0.656 -0.753 -0.379 0.831 -0.321 ...
##  $ LGRE_align.W.PET             : num  -0.402 -0.54 0.346 -0.752 1.528 ...
##  $ LGHRE_align.W.PET            : num  -0.463 -0.557 0.328 -0.667 1.732 ...
##  $ HGLRE_align.W.PET            : num  -0.3725 -0.0206 -0.8543 -0.1871 -0.9693 ...
##  $ LZSE.W.PET                   : num  -0.4602 -0.5497 0.0317 -0.2898 0.0801 ...
##  $ LZLGE.W.PET                  : num  -0.281 -0.323 -0.14 -0.303 0.362 ...
##  $ LZHGE.W.PET                  : num  -0.5234 -0.3964 -0.7161 0.0809 -1.0935 ...
##  $ GLVAR_area.W.PET             : num  -0.277 0.33 -0.896 -0.238 -0.924 ...
##  $ Min_hist.ADC                 : num  0.411 -0.866 0.609 -0.866 -0.866 ...
##  $ Max_hist.ADC                 : num  -0.5414 -0.5918 -0.0183 -0.0104 -0.4345 ...
##  $ Mean_hist.ADC                : num  -0.387 -0.519 -0.364 -0.458 -0.745 ...
##  $ Standard_Deviation_hist.ADC: num  -0.1324 -0.4277 0.5113 0.0383 -0.1624 ...
##  $ Skewness_hist.ADC            : num  0.76 -1.313 1.401 -0.334 -0.228 ...
##  $ Kurtosis_hist.ADC            : num  -0.365 0.356 0.884 -0.483 -0.293 ...
##  $ X3D_surface.ADC              : num  -0.8336 -0.7264 -0.5623 -0.0772 -0.5594 ...
##  $ ratio_3ds_vol.ADC            : num  0.407 -0.204 -0.515 -0.528 -0.479 ...
##  $ Compactness_v2.ADC           : num  -0.5654 0.0169 0.2181 -0.8744 0.0688 ...
##  $ Spherical_disproportion.ADC: num  -0.51 -0.731 -0.789 -0.34 -0.747 ...
##  $ Sphericity.ADC               : num  -0.576 -0.338 -0.266 -0.729 -0.319 ...
##  $ Asphericity.ADC              : num  -0.3281 -0.796 -0.9186 0.0325 -0.829 ...
##  $ Center_of_mass.ADC           : num  -0.16 -0.135 0.312 0.165 -0.522 ...
##   [list output truncated]
```

```
attach(final_radiomics)
```

Create training (80%) and testing (20%) data.

```
set.seed(123)
radio <- final_radiomics %>% mutate_if(is.ordered, factor, ordered = FALSE)
splitdata = initial_split(radio ,prop = 0.8 ,strata = "Failure.binary")
splitdata
```

```
## <Training/Testing/Total>
## <157/40/197>
```

```r
final_radiomics_train <- training(splitdata)
head(final_radiomics_train[1:5])
```

```
##    Failure.binary    Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET
## 5               0  0.6898772         0.34499368      -0.06740573  -0.98874071
## 7               0 -0.0714332        -0.07231092      -0.68049871   0.02814910
## 13              0  0.9198383        -0.47508410      -0.04700109   0.25645562
## 16              0  1.1811577         0.30131288      -0.60733786  -1.00812837
## 22              0  1.9233047        -1.86279117      -0.95781836  -0.49187889
## 23              0  1.1619942        -0.94025294      -0.29104352   0.04032431
```

```r
final_radiomics_test <- testing(splitdata)
head(final_radiomics_test[1:5])
```

```
##    Failure.binary    Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET
## 1               0  1.1985789       0.5529054683      -0.57063689   -0.4541408
## 3               0  2.7926271       0.4599082452      -0.06024275   -1.1504338
## 4               1 -0.4442487       1.1431829822       2.67468822   -0.4446190
## 8               0  0.4930166      -0.0029909626      -0.51556521   -0.5619898
## 24              1 -0.8519069       1.1229005476       0.94672092   -0.5762384
## 28              0  0.4477213      -0.0007790328       0.34523282   -0.4119103
```

```r
prep_train <- recipe(Failure.binary~., data=final_radiomics_train) %>%
  step_integer(all_nominal()) %>%
  step_nzv(all_nominal()) %>%
  step_dummy(all_nominal(), -all_outcomes(), one_hot = TRUE) %>%
  step_center(all_numeric(), -all_outcomes()) %>%
  step_scale(all_numeric(), -all_outcomes()) %>%
  prep(training = final_radiomics_train, retain = TRUE) %>%
  juice()

prep_test <- recipe(Failure.binary~., data=final_radiomics_test) %>%
  step_integer(all_nominal()) %>%
  step_nzv(all_nominal()) %>%
  step_dummy(all_nominal(), -all_outcomes(), one_hot = TRUE) %>%
  step_center(all_numeric(), -all_outcomes()) %>%
  step_scale(all_numeric(), -all_outcomes()) %>%
  prep(testing = final_radiomics_test, retain = TRUE) %>%
  juice()

X_train<- as.matrix(prep_train[setdiff(names(prep_train),
                                          "Failure.binary")])
Y_train<- prep_train$Failure.binary

X_test<- as.matrix(prep_test[setdiff(names(prep_test),
                                        "Failure.binary")])
Y_test<- prep_test$Failure.binary
```

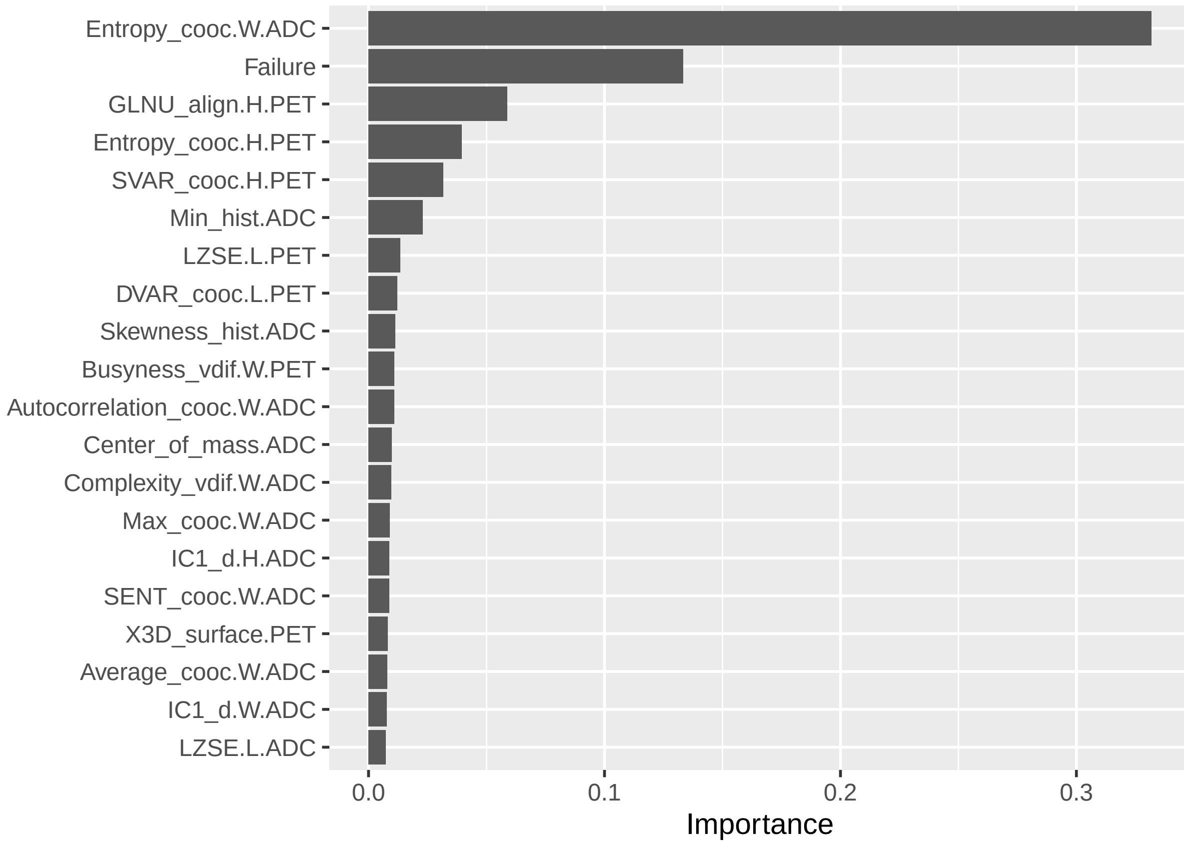
Model 1

Model 1.1: Modelling the data using XGBOOST

```r
# optimal parameter list
params <- list(
  eta = 0.01,
  max_depth = 3,
  min_child_weight = 3,
  subsample = 0.5,
  colsample_bytree = 0.5
)


# Model of Train Data
set.seed(123)
xgb.fit.final <- xgboost(
  params = params,
  data = X_train,
  label = Y_train-1,
  nrounds = 4000,
  objective = "binary:logistic",
  verbose = 0
)
summary(xgb.fit.final)
```

```
##                 Length  Class              Mode
## handle                1 xgb.Booster.handle externalptr
## raw             2906834 -none-             raw
## niter                 1 -none-             numeric
## evaluation_log        2 data.table         list
## call                 14 -none-             call
## params                7 -none-             list
## callbacks             1 -none-             list
## feature_names       147 -none-             character
## nfeatures             1 -none-             numeric
```

```r
# Top 20 important features during Training
vip::vip(xgb.fit.final, num_features = 20)
```




```r
# Prediction performance of the model using training data set
pred_xgboost_train<- predict(xgb.fit.final, X_train, type = "prob")
pred_xgboost_train
```

```
##   [1] 0.099857010 0.058816653 0.014479009 0.046121638 0.014174885 0.043349206
##   [7] 0.052952662 0.267831147 0.007973359 0.012994351 0.017465856 0.047533493
##  [13] 0.038613003 0.030375447 0.026322233 0.053214654 0.013139801 0.112728432
##  [19] 0.108347528 0.041213144 0.031961415 0.037135199 0.036562968 0.022781346
##  [25] 0.015294581 0.021129945 0.010686113 0.032881964 0.023751076 0.017315181
##  [31] 0.020384636 0.019715860 0.055517863 0.027203469 0.057680108 0.014648456
##  [37] 0.013708800 0.013956445 0.079623953 0.052200902 0.018291030 0.028915629
##  [43] 0.054857679 0.026446063 0.012269739 0.092204437 0.018490253 0.033492487
##  [49] 0.149206325 0.064320475 0.058247797 0.065827601 0.060390830 0.265067965
##  [55] 0.196747810 0.070988320 0.049936168 0.042830337 0.070069231 0.025953034
##  [61] 0.440651000 0.215244457 0.259388864 0.280341029 0.024854125 0.019391682
##  [67] 0.016616840 0.010625809 0.060528897 0.027998473 0.021445725 0.061455321
##  [73] 0.172972918 0.033041745 0.454813421 0.071569338 0.029419405 0.009356284
##  [79] 0.009943988 0.086578146 0.010247789 0.026394285 0.011763698 0.027550003
##  [85] 0.017781980 0.016190708 0.215698332 0.018005263 0.026681492 0.243469357
##  [91] 0.032860588 0.016484503 0.375563532 0.015363351 0.041187692 0.021189865
##  [97] 0.014912988 0.013414036 0.025056668 0.027815564 0.034197401 0.146598831
## [103] 0.022172604 0.038975149 0.683638930 0.954720080 0.886610568 0.630057275
## [109] 0.952911735 0.973130882 0.921227276 0.945268869 0.969312429 0.979291737
## [115] 0.868677258 0.968729317 0.983300567 0.485143244 0.940416038 0.977697313
## [121] 0.978055716 0.958996594 0.940067410 0.933341682 0.944213688 0.984498084
## [127] 0.987913430 0.618517697 0.980422199 0.987508714 0.939536452 0.896966159
```

```
## [133] 0.639649391 0.735558033 0.960798860 0.765768349 0.864851296 0.965727210
## [139] 0.926994443 0.696273327 0.651460171 0.854598820 0.961534739 0.985732913
## [145] 0.940430045 0.392735660 0.844619572 0.942693293 0.967622995 0.965173960
## [151] 0.951722741 0.942448497 0.858735561 0.914785445 0.594496548 0.880704045
## [157] 0.735231042
```

```
perf1 <- prediction(pred_xgboost_train,final_radiomics_train$Failure.binary) %>%
  performance(measure = "tpr", x.measure = "fpr")
perf1
```

```
## A performance instance
##   'False positive rate' vs. 'True positive rate' (alpha: 'Cutoff')
##   with 158 data points
```

```
# Prediction performance of the model using testing data set
pred_xgboost_test<- predict(xgb.fit.final, X_test, type = "prob")
pred_xgboost_test
```

```
##  [1] 0.404377669 0.358926207 0.950822949 0.060761694 0.953233600 0.078627832
##  [7] 0.975705087 0.035459876 0.605009973 0.006428809 0.031629287 0.010345870
## [13] 0.121987216 0.444863558 0.102037884 0.281896681 0.131317899 0.204408988
## [19] 0.037064303 0.078438655 0.828603208 0.517040551 0.199922308 0.483729303
## [25] 0.701296926 0.005774284 0.785132825 0.969533443 0.883065999 0.346980691
## [31] 0.030128101 0.940216124 0.017736306 0.963804543 0.178022757 0.290226817
## [37] 0.109095208 0.028470060 0.027936855 0.505909383
```

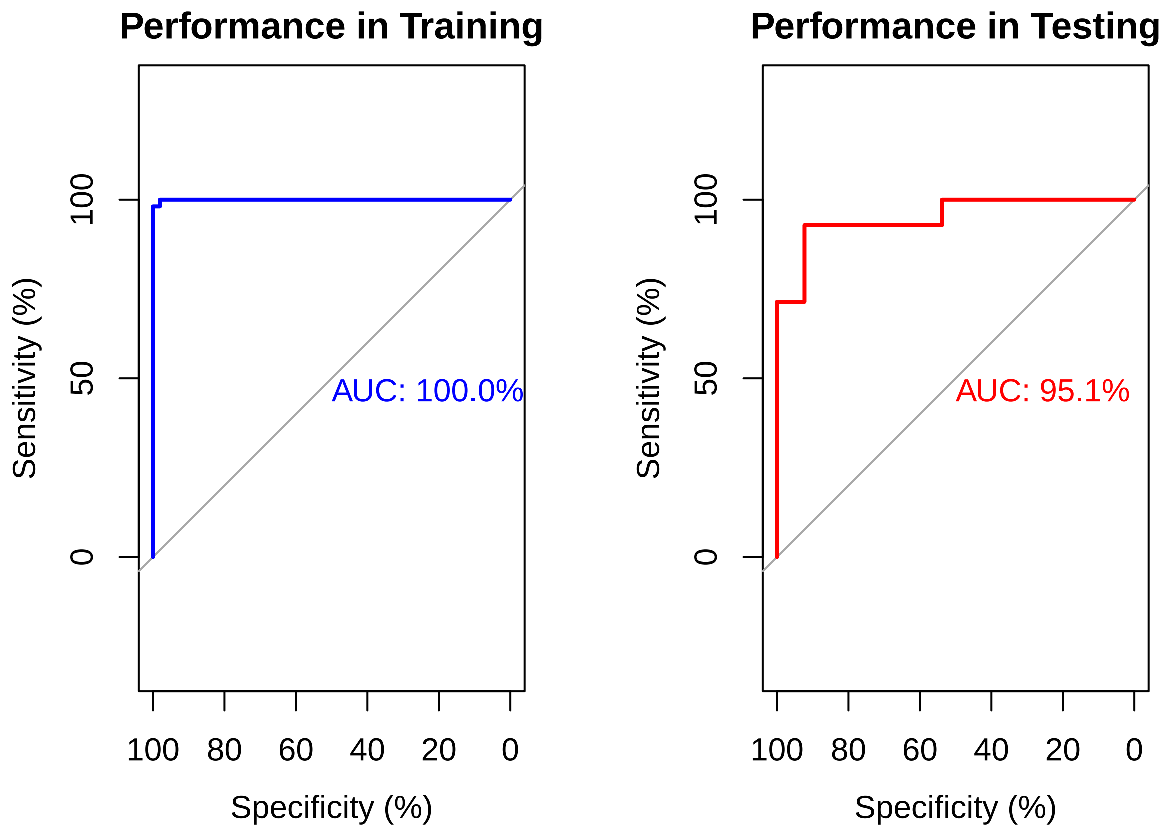
```
perf2 <- prediction(pred_xgboost_test, final_radiomics_test$Failure.binary) %>%
  performance(measure = "tpr", x.measure = "fpr")
perf2
```

```
## A performance instance
##   'False positive rate' vs. 'True positive rate' (alpha: 'Cutoff')
##   with 41 data points
```

```
# Training and Testing data performance plot
par(mfrow = c(1,2))

# Training prediction performane
roc(final_radiomics_train$Failure.binary ~ pred_xgboost_train,
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="blue", lwd=2, print.auc=TRUE,
    main = "Performance in Training")
```

```
##
## Call:
## roc.formula(formula = final_radiomics_train$Failure.binary ~     pred_xgboost_train, plot = TRUE, leg
##
## Data: pred_xgboost_train in 104 controls (final_radiomics_train$Failure.binary 0) < 53 cases (final_r
## Area under the curve: 99.96%
```

```
# Testing set prediction performance
roc(final_radiomics_test$Failure.binary ~ pred_xgboost_test,
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="red", lwd=2, print.auc=TRUE,
    main = "Performance in Testing")
```

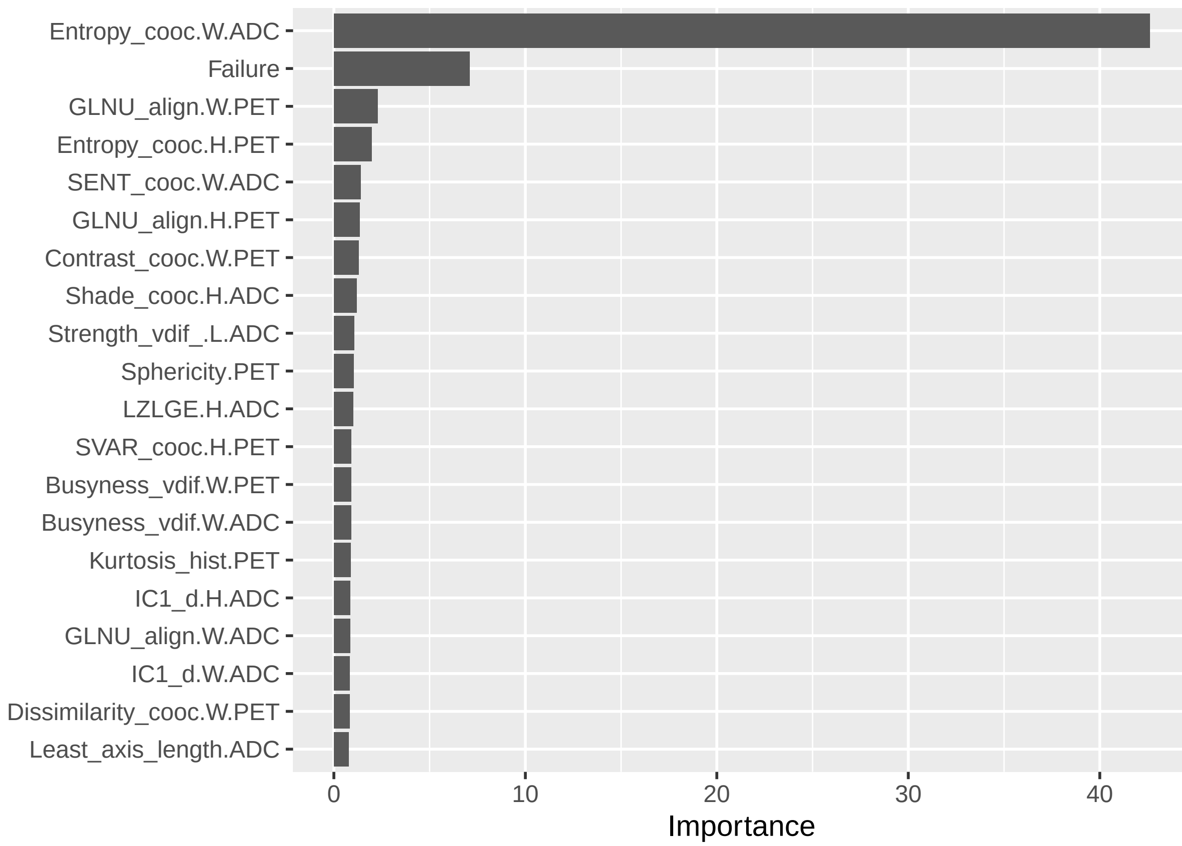



```
##
## Call:
## roc.formula(formula = final_radiomics_test$Failure.binary ~ pred_xgboost_test,     plot = TRUE, legac
##
## Data: pred_xgboost_test in 26 controls (final_radiomics_test$Failure.binary 0) < 14 cases (final_radi
## Area under the curve: 95.05%
```

Model 1.2: Modelling the data using GBM

```
set.seed(123)
gbm_model <- gbm(
  formula = Failure.binary ~ .,
  data = final_radiomics_train,
  distribution = "gaussian",
  n.trees = 500,
  shrinkage = 0.1,
  interaction.depth = 3,
```

```
  n.minobsinnode = 10,
  cv.folds = 10
)

# Top 20 important features during Training
vip::vip(gbm_model, num_features = 20)
```




```
# Prediction performance of the model using training data set
pred_gbm_train<- predict(gbm_model, newdata=as.data.frame(X_train),
                         type = "response")
pred_gbm_train
```

```
##   [1] 1.0566497 1.0816403 0.9693307 1.0196355 0.9893454 0.9351025 1.1278266
##   [8] 1.5476636 0.9690415 0.9661116 0.9309952 1.1560535 1.0312913 1.0315849
##  [15] 0.9331039 0.9962456 0.9822937 1.1300473 1.0564736 1.0369094 0.9998221
##  [22] 1.0291438 0.9834592 0.9952928 1.0271875 0.9719281 0.9270533 1.0128893
##  [29] 1.0175781 1.0798480 1.0071605 1.0540205 1.0221665 1.0514597 0.9909503
##  [36] 0.9428613 1.0314571 0.9815587 1.1003302 1.0907252 1.0201658 1.0930131
##  [43] 1.1563157 0.9780830 0.9410110 1.0926590 0.9315293 1.0114950 1.1804196
##  [50] 0.9619457 1.1035300 1.1682644 1.0637943 1.3719950 1.3079941 1.0591457
##  [57] 1.0155701 1.0716342 1.0397281 0.9782105 1.4500397 1.2002860 1.0711105
##  [64] 1.5737326 0.9809746 0.9265685 0.9696695 0.8998626 1.0638641 0.9922283
##  [71] 1.0797131 1.0399246 1.0614436 1.0069654 1.6553242 1.0805268 0.9395916
##  [78] 0.9504520 0.9482118 1.0973026 1.0795828 1.0586595 1.0273664 1.1073017
##  [85] 0.9560699 0.9744405 1.3678170 0.9698946 1.0534984 1.3353315 1.0892608
```

```
##  [92] 1.0753950 1.6086557 1.0658027 1.0838032 1.0352556 0.9868398 1.0399511
##  [99] 0.9433450 1.0072161 1.0313690 1.3517551 0.9816960 1.0128522 1.5463945
## [106] 1.9895515 1.7938371 1.5309483 2.0779838 1.9648821 1.9901807 1.6761529
## [113] 2.0119019 2.0448209 1.8690355 1.9530526 2.0386907 1.4368486 1.9412657
## [120] 2.0432251 2.0192088 1.9649723 1.8581213 1.9645322 1.8953674 1.9905785
## [127] 2.0268372 1.4078311 2.1438709 2.0223600 1.9449216 1.9333967 1.6668446
## [134] 1.8504634 1.9198445 1.6846205 1.6352294 1.9855378 1.8559478 1.8718856
## [141] 1.6298320 1.9120780 1.8157256 2.0871928 2.0552012 1.3035513 1.7154484
## [148] 1.9402573 2.0501549 2.0015447 2.0324168 2.0026052 1.9163733 1.9128407
## [155] 1.7689935 1.6731549 1.9170913
```

```
perf3 <- prediction(pred_gbm_train,final_radiomics_train$Failure.binary) %>%
  performance(measure = "tpr", x.measure = "fpr")
perf3
```

```
## A performance instance
##   'False positive rate' vs. 'True positive rate' (alpha: 'Cutoff')
##   with 158 data points
```

```
# Prediction performance of the model using testing data set
pred_gbm_test<- predict(gbm_model, newdata=as.data.frame(X_test),
                        type = "response")
pred_gbm_test
```

```
##  [1] 1.4319948 1.4955901 1.9581071 1.0877395 2.0022408 1.0565499 1.9566118
##  [8] 1.0202829 1.2409809 0.9101864 1.0086530 0.9270703 1.1024023 1.8247972
## [15] 1.1715897 1.1311502 1.0687053 1.1484260 1.0650793 1.2015374 1.8009021
## [22] 1.4901733 1.1761204 1.6342046 1.6761921 0.9114302 1.7754062 1.7617687
## [29] 1.6474194 1.1923029 0.9474163 1.9376117 1.0074654 1.9834138 1.4608164
## [36] 1.6102214 1.1172166 0.9199796 0.9770945 1.9667019
```

```
perf4 <- prediction(pred_gbm_test, final_radiomics_test$Failure.binary) %>%
  performance(measure = "tpr", x.measure = "fpr")
perf4
```

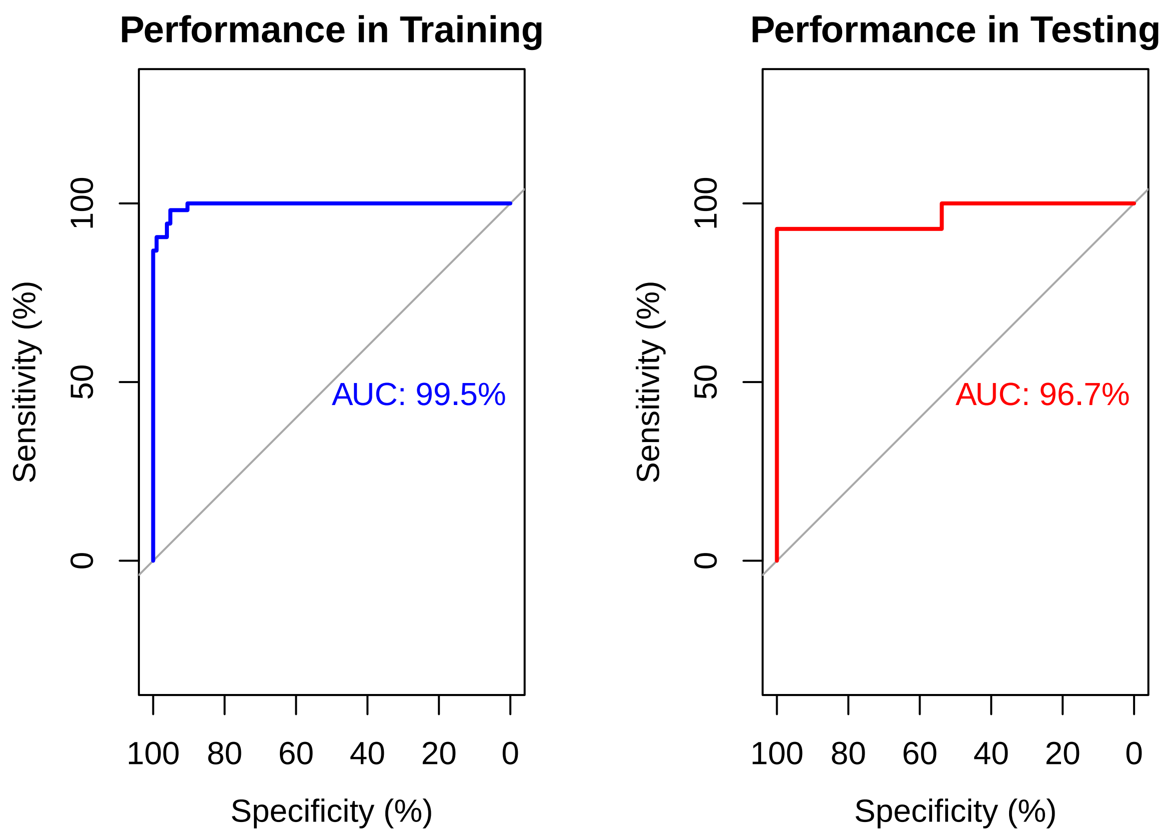
```
## A performance instance
##   'False positive rate' vs. 'True positive rate' (alpha: 'Cutoff')
##   with 41 data points
```

```
# Training and Testing data performance plot
par(mfrow = c(1,2))

# Training prediction performane
roc(final_radiomics_train$Failure.binary ~ pred_gbm_train,
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="blue", lwd=2, print.auc=TRUE,
    main = "Performance in Training")
```

```
##
## Call:
## roc.formula(formula = final_radiomics_train$Failure.binary ~     pred_gbm_train, plot = TRUE, legacy
```

```
##
## Data: pred_gbm_train in 104 controls (final_radiomics_train$Failure.binary 0) < 53 cases (final_radi
## Area under the curve: 99.46%
```

```
# Testing set prediction performance
roc(final_radiomics_test$Failure.binary ~ pred_gbm_test,
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="red", lwd=2, print.auc=TRUE,
    main = "Performance in Testing")
```

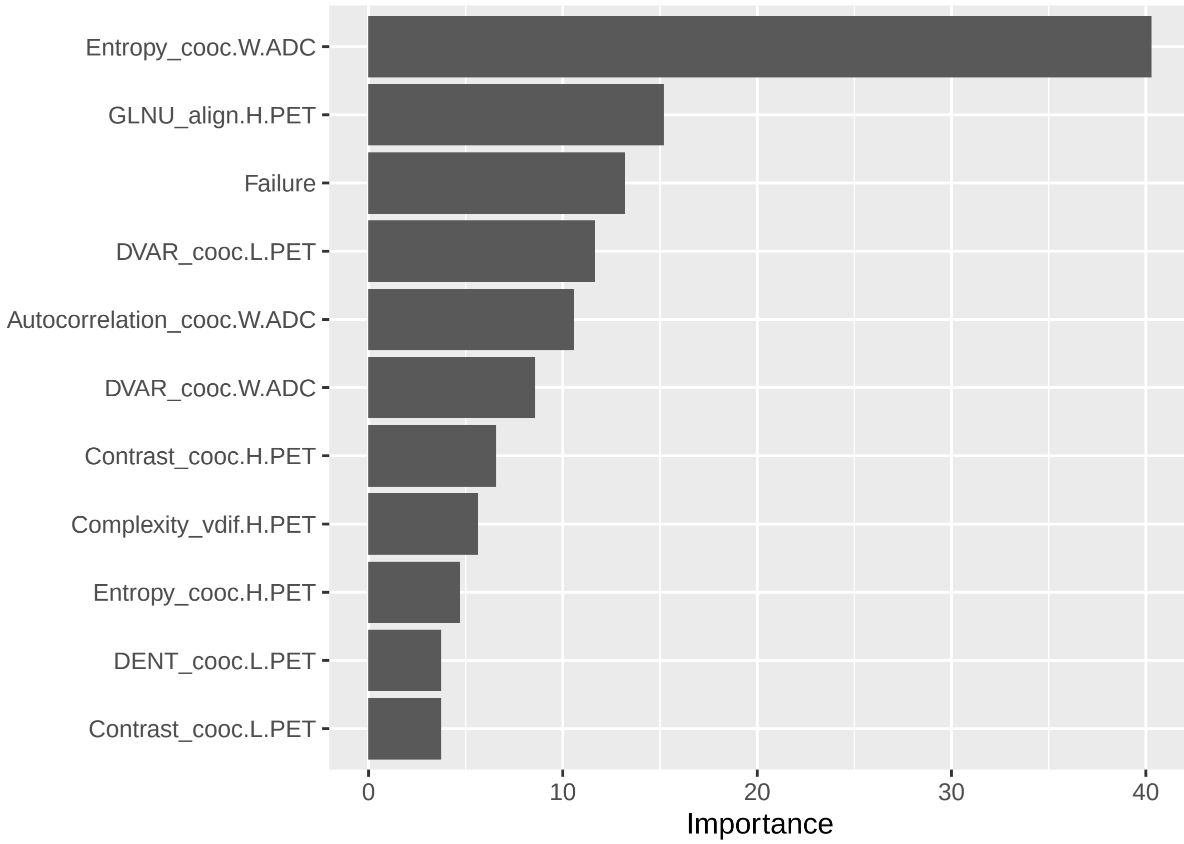



```
##
## Call:
## roc.formula(formula = final_radiomics_test$Failure.binary ~ pred_gbm_test,     plot = TRUE, legacy.a
##
## Data: pred_gbm_test in 26 controls (final_radiomics_test$Failure.binary 0) < 14 cases (final_radiomi
## Area under the curve: 96.7%
```

Model 1.3: Modelling the data using Rpart

```
set.seed(123)
# Modelling
rpart_train_model <- rpart(Failure.binary~.,data=final_radiomics_train,
                           method="class")
```

```r
rpart_test_model <- rpart(Failure.binary~.,data=final_radiomics_test,
                          method="class")


# Top 20 important features during Training
vip::vip(rpart_train_model, num_features = 20)
```




```r
# Prediction performance of the model using training data set
pred_rpart_train<- predict(rpart_train_model, newdata=as.data.frame(X_train),
                           type = "prob", na.action = na.pass)
pred_rpart_train
```

```
##            0          1
## 1   0.9479167 0.05208333
## 2   0.9479167 0.05208333
## 3   0.9479167 0.05208333
## 4   0.9479167 0.05208333
## 5   0.9479167 0.05208333
## 6   0.9479167 0.05208333
## 7   0.9479167 0.05208333
## 8   0.1296296 0.87037037
## 9   0.9479167 0.05208333
## 10  0.9479167 0.05208333
## 11  0.9479167 0.05208333
```

```
## 12  0.9479167 0.05208333
## 13  0.9479167 0.05208333
## 14  0.9479167 0.05208333
## 15  0.9479167 0.05208333
## 16  0.9479167 0.05208333
## 17  0.9479167 0.05208333
## 18  0.1296296 0.87037037
## 19  0.9479167 0.05208333
## 20  0.9479167 0.05208333
## 21  0.9479167 0.05208333
## 22  0.9479167 0.05208333
## 23  0.9479167 0.05208333
## 24  0.9479167 0.05208333
## 25  0.9479167 0.05208333
## 26  0.9479167 0.05208333
## 27  0.9479167 0.05208333
## 28  0.9479167 0.05208333
## 29  0.9479167 0.05208333
## 30  0.9479167 0.05208333
## 31  0.9479167 0.05208333
## 32  0.9479167 0.05208333
## 33  0.9479167 0.05208333
## 34  0.9479167 0.05208333
## 35  0.9479167 0.05208333
## 36  0.9479167 0.05208333
## 37  0.9479167 0.05208333
## 38  0.9479167 0.05208333
## 39  0.9479167 0.05208333
## 40  0.9479167 0.05208333
## 41  0.9479167 0.05208333
## 42  0.9479167 0.05208333
## 43  0.9479167 0.05208333
## 44  0.9479167 0.05208333
## 45  0.9479167 0.05208333
## 46  0.9479167 0.05208333
## 47  0.9479167 0.05208333
## 48  0.9479167 0.05208333
## 49  0.9479167 0.05208333
## 50  0.9479167 0.05208333
## 51  0.9479167 0.05208333
## 52  0.9479167 0.05208333
## 53  0.9479167 0.05208333
## 54  0.1296296 0.87037037
## 55  0.9479167 0.05208333
## 56  0.9479167 0.05208333
## 57  0.9479167 0.05208333
## 58  0.9479167 0.05208333
## 59  0.9479167 0.05208333
## 60  0.9479167 0.05208333
## 61  0.1296296 0.87037037
## 62  0.1296296 0.87037037
## 63  0.9479167 0.05208333
## 64  0.1296296 0.87037037
## 65  0.9479167 0.05208333
```

```
## 66  0.9479167 0.05208333
## 67  0.9479167 0.05208333
## 68  0.9479167 0.05208333
## 69  0.9479167 0.05208333
## 70  0.9479167 0.05208333
## 71  0.9479167 0.05208333
## 72  0.9479167 0.05208333
## 73  0.9479167 0.05208333
## 74  0.9479167 0.05208333
## 75  0.1296296 0.87037037
## 76  0.8571429 0.14285714
## 77  0.9479167 0.05208333
## 78  0.9479167 0.05208333
## 79  0.9479167 0.05208333
## 80  0.8571429 0.14285714
## 81  0.9479167 0.05208333
## 82  0.9479167 0.05208333
## 83  0.9479167 0.05208333
## 84  0.9479167 0.05208333
## 85  0.9479167 0.05208333
## 86  0.9479167 0.05208333
## 87  0.8571429 0.14285714
## 88  0.9479167 0.05208333
## 89  0.9479167 0.05208333
## 90  0.8571429 0.14285714
## 91  0.9479167 0.05208333
## 92  0.9479167 0.05208333
## 93  0.8571429 0.14285714
## 94  0.9479167 0.05208333
## 95  0.9479167 0.05208333
## 96  0.9479167 0.05208333
## 97  0.9479167 0.05208333
## 98  0.9479167 0.05208333
## 99  0.9479167 0.05208333
## 100 0.9479167 0.05208333
## 101 0.9479167 0.05208333
## 102 0.8571429 0.14285714
## 103 0.9479167 0.05208333
## 104 0.9479167 0.05208333
## 105 0.9479167 0.05208333
## 106 0.1296296 0.87037037
## 107 0.1296296 0.87037037
## 108 0.9479167 0.05208333
## 109 0.1296296 0.87037037
## 110 0.1296296 0.87037037
## 111 0.1296296 0.87037037
## 112 0.1296296 0.87037037
## 113 0.1296296 0.87037037
## 114 0.1296296 0.87037037
## 115 0.1296296 0.87037037
## 116 0.1296296 0.87037037
## 117 0.1296296 0.87037037
## 118 0.9479167 0.05208333
## 119 0.1296296 0.87037037
```

```
## 120 0.1296296 0.87037037
## 121 0.1296296 0.87037037
## 122 0.1296296 0.87037037
## 123 0.1296296 0.87037037
## 124 0.1296296 0.87037037
## 125 0.1296296 0.87037037
## 126 0.1296296 0.87037037
## 127 0.1296296 0.87037037
## 128 0.9479167 0.05208333
## 129 0.1296296 0.87037037
## 130 0.1296296 0.87037037
## 131 0.1296296 0.87037037
## 132 0.1296296 0.87037037
## 133 0.1296296 0.87037037
## 134 0.1296296 0.87037037
## 135 0.1296296 0.87037037
## 136 0.1296296 0.87037037
## 137 0.9479167 0.05208333
## 138 0.1296296 0.87037037
## 139 0.1296296 0.87037037
## 140 0.1296296 0.87037037
## 141 0.1296296 0.87037037
## 142 0.1296296 0.87037037
## 143 0.1296296 0.87037037
## 144 0.1296296 0.87037037
## 145 0.1296296 0.87037037
## 146 0.9479167 0.05208333
## 147 0.1296296 0.87037037
## 148 0.8571429 0.14285714
## 149 0.1296296 0.87037037
## 150 0.1296296 0.87037037
## 151 0.1296296 0.87037037
## 152 0.1296296 0.87037037
## 153 0.1296296 0.87037037
## 154 0.1296296 0.87037037
## 155 0.1296296 0.87037037
## 156 0.1296296 0.87037037
## 157 0.1296296 0.87037037
```

```r
# Prediction performance of the model using testing data set
pred_rpart_test<- predict(rpart_test_model, newdata=as.data.frame(X_test),
                        type = "prob", na.action = na.pass)
pred_rpart_test
```

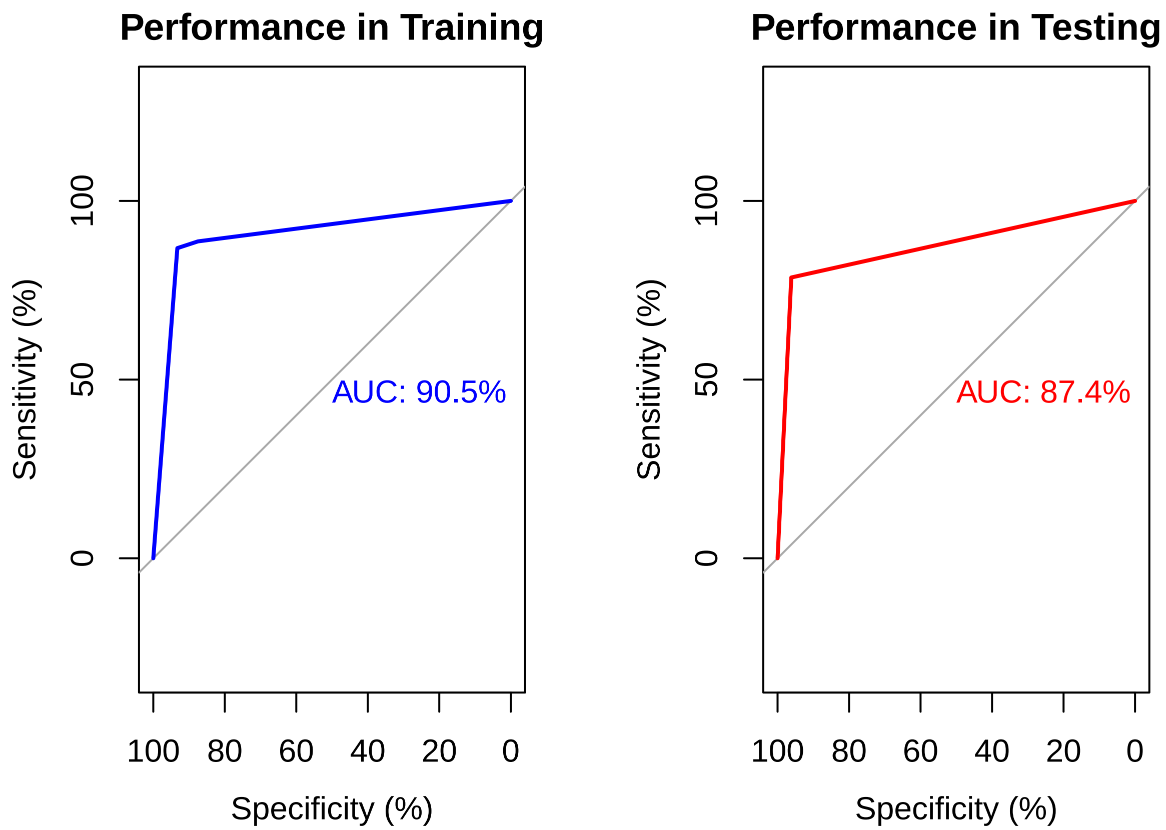
```
##             0          1
## 1  0.92592593 0.07407407
## 2  0.92592593 0.07407407
## 3  0.07692308 0.92307692
## 4  0.92592593 0.07407407
## 5  0.07692308 0.92307692
## 6  0.92592593 0.07407407
## 7  0.07692308 0.92307692
## 8  0.92592593 0.07407407
## 9  0.92592593 0.07407407
```

```
## 10 0.92592593 0.07407407
## 11 0.92592593 0.07407407
## 12 0.92592593 0.07407407
## 13 0.92592593 0.07407407
## 14 0.07692308 0.92307692
## 15 0.92592593 0.07407407
## 16 0.92592593 0.07407407
## 17 0.92592593 0.07407407
## 18 0.92592593 0.07407407
## 19 0.92592593 0.07407407
## 20 0.92592593 0.07407407
## 21 0.07692308 0.92307692
## 22 0.92592593 0.07407407
## 23 0.92592593 0.07407407
## 24 0.07692308 0.92307692
## 25 0.07692308 0.92307692
## 26 0.92592593 0.07407407
## 27 0.07692308 0.92307692
## 28 0.92592593 0.07407407
## 29 0.92592593 0.07407407
## 30 0.92592593 0.07407407
## 31 0.92592593 0.07407407
## 32 0.07692308 0.92307692
## 33 0.92592593 0.07407407
## 34 0.07692308 0.92307692
## 35 0.92592593 0.07407407
## 36 0.07692308 0.92307692
## 37 0.92592593 0.07407407
## 38 0.92592593 0.07407407
## 39 0.92592593 0.07407407
## 40 0.07692308 0.92307692
```

```r
# Training and Testing data performance plot
par(mfrow = c(1,2))

# Training prediction performane
roc(final_radiomics_train$Failure.binary ~ pred_rpart_train[,2],
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="blue", lwd=2, print.auc=TRUE,
    main = "Performance in Training")
```

```
##
## Call:
## roc.formula(formula = final_radiomics_train$Failure.binary ~     pred_rpart_train[, 2], plot = TRUE,
##
## Data: pred_rpart_train[, 2] in 104 controls (final_radiomics_train$Failure.binary 0) < 53 cases (fina
## Area under the curve: 90.53%
```

```r
# Testing set prediction performance
roc(final_radiomics_test$Failure.binary ~ pred_rpart_test[,2],
    plot=TRUE, legacy.axes=FALSE,
    percent=TRUE, col="red", lwd=2, print.auc=TRUE,
    main = "Performance in Testing")
```




```
##
## Call:
## roc.formula(formula = final_radiomics_test$Failure.binary ~ pred_rpart_test[,     2], plot = TRUE, lo
##
## Data: pred_rpart_test[, 2] in 26 controls (final_radiomics_test$Failure.binary 0) < 14 cases (final_r
## Area under the curve: 87.36%
```

Model 2

```
set.seed(123)
Train_Features <- data.matrix(final_radiomics_train[,-1])
Train_Labels <- final_radiomics_train[,1]
Test_Features <- data.matrix(final_radiomics_test[,-1])
Test_Labels <- final_radiomics_test[,1]

# Reshaping the dataset
colnames(Train_Features) <- paste0("V", 1:ncol(Train_Features))
Train_Features <- Train_Features / 255

colnames(Test_Features) <- paste0("V", 1:ncol(Test_Features))
Test_Features <- Test_Features / 255

# Converting the labels into categorical
Train_Labels <- to_categorical(Train_Labels, num_classes = 2)
```

```r
Test_Labels <- to_categorical(Test_Labels, num_classes = 2)


# Model training
set.seed(123)
model <- keras_model_sequential() %>%
  layer_dense(units = 256, activation = "sigmoid", input_shape = ncol(Train_Features)) %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 128, activation = "sigmoid") %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 128, activation = "sigmoid") %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 64, activation = "sigmoid") %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 64, activation = "sigmoid") %>%
  layer_dropout(rate = 0.3) %>%
  layer_dense(units = 2, activation = "softmax")
summary(model)
```

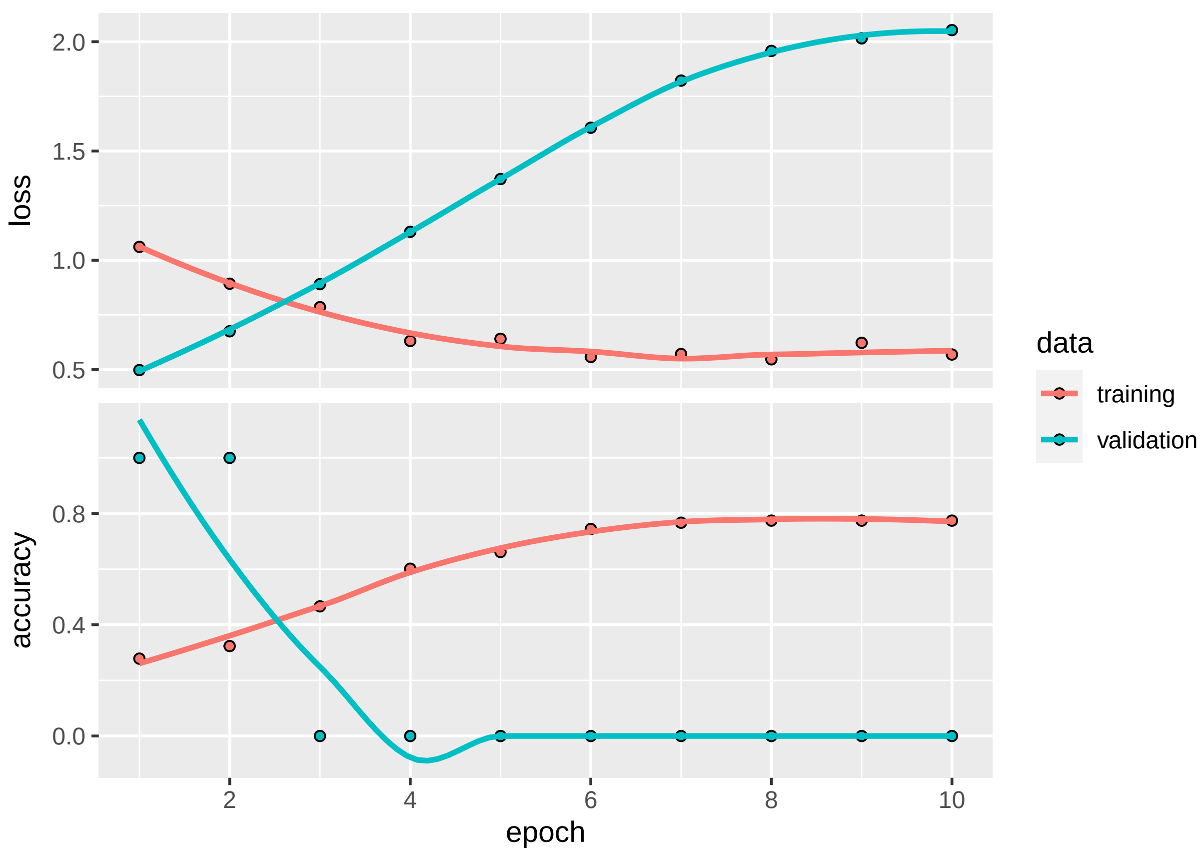
```
## Model: "sequential"
## ________________________________________________________________________________
##  Layer (type)                       Output Shape                    Param #
## ================================================================================
##  dense_5 (Dense)                    (None, 256)                     37888
##  dropout_4 (Dropout)                (None, 256)                     0
##  dense_4 (Dense)                    (None, 128)                     32896
##  dropout_3 (Dropout)                (None, 128)                     0
##  dense_3 (Dense)                    (None, 128)                     16512
##  dropout_2 (Dropout)                (None, 128)                     0
##  dense_2 (Dense)                    (None, 64)                      8256
##  dropout_1 (Dropout)                (None, 64)                      0
##  dense_1 (Dense)                    (None, 64)                      4160
##  dropout (Dropout)                  (None, 64)                      0
##  dense (Dense)                      (None, 2)                       130
## ================================================================================
## Total params: 99,842
## Trainable params: 99,842
## Non-trainable params: 0
## ________________________________________________________________________________
```

```r
# Backpropagation
model %>% compile(
  loss = "categorical_crossentropy",
  optimizer = optimizer_rmsprop(),
  metrics = c("accuracy")
)

# Compiling the model
model %>% compile(
  loss = "categorical_crossentropy",
  optimizer = optimizer_adam(),
  metrics = c("accuracy")
)
```

```r
# Train the model
set.seed(123)
model_fit <- model %>%
  fit(Train_Features, Train_Labels, epochs = 10, batch_size = 128,
      validation_split = 0.15)

# Display Output
model_fit
```

```
##
## Final epoch (plot to see history):
##         loss: 0.5685
##     accuracy: 0.7744
##     val_loss: 2.053
## val_accuracy: 0
```

```r
plot(model_fit)
```




```r
# Model evaluation
model %>%
  evaluate(Test_Features, Test_Labels)
```

```
##      loss  accuracy
## 0.8078468 0.6500000
```

```r
# Model prediction
model %>% predict(Test_Features)
```

```
##            [,1]      [,2]
##  [1,] 0.8716558 0.1283442
##  [2,] 0.8716552 0.1283448
##  [3,] 0.8716555 0.1283446
##  [4,] 0.8716555 0.1283445
##  [5,] 0.8716553 0.1283446
##  [6,] 0.8716556 0.1283443
##  [7,] 0.8716550 0.1283450
##  [8,] 0.8716559 0.1283441
##  [9,] 0.8716551 0.1283449
## [10,] 0.8716553 0.1283448
## [11,] 0.8716562 0.1283438
## [12,] 0.8716562 0.1283438
## [13,] 0.8716557 0.1283443
## [14,] 0.8716551 0.1283450
## [15,] 0.8716555 0.1283445
## [16,] 0.8716555 0.1283446
## [17,] 0.8716555 0.1283445
## [18,] 0.8716555 0.1283445
## [19,] 0.8716553 0.1283448
## [20,] 0.8716558 0.1283442
## [21,] 0.8716547 0.1283453
## [22,] 0.8716551 0.1283450
## [23,] 0.8716558 0.1283442
## [24,] 0.8716555 0.1283445
## [25,] 0.8716555 0.1283444
## [26,] 0.8716555 0.1283444
## [27,] 0.8716559 0.1283441
## [28,] 0.8716551 0.1283449
## [29,] 0.8716553 0.1283448
## [30,] 0.8716553 0.1283446
## [31,] 0.8716550 0.1283450
## [32,] 0.8716549 0.1283450
## [33,] 0.8716552 0.1283448
## [34,] 0.8716543 0.1283457
## [35,] 0.8716560 0.1283441
## [36,] 0.8716548 0.1283452
## [37,] 0.8716555 0.1283446
## [38,] 0.8716553 0.1283447
## [39,] 0.8716555 0.1283444
## [40,] 0.8716548 0.1283452
```

Model 3

```r
# The data
data <- radiomics_df
head(data[1:5])
```

```
##      Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET Max_hist.PET
```

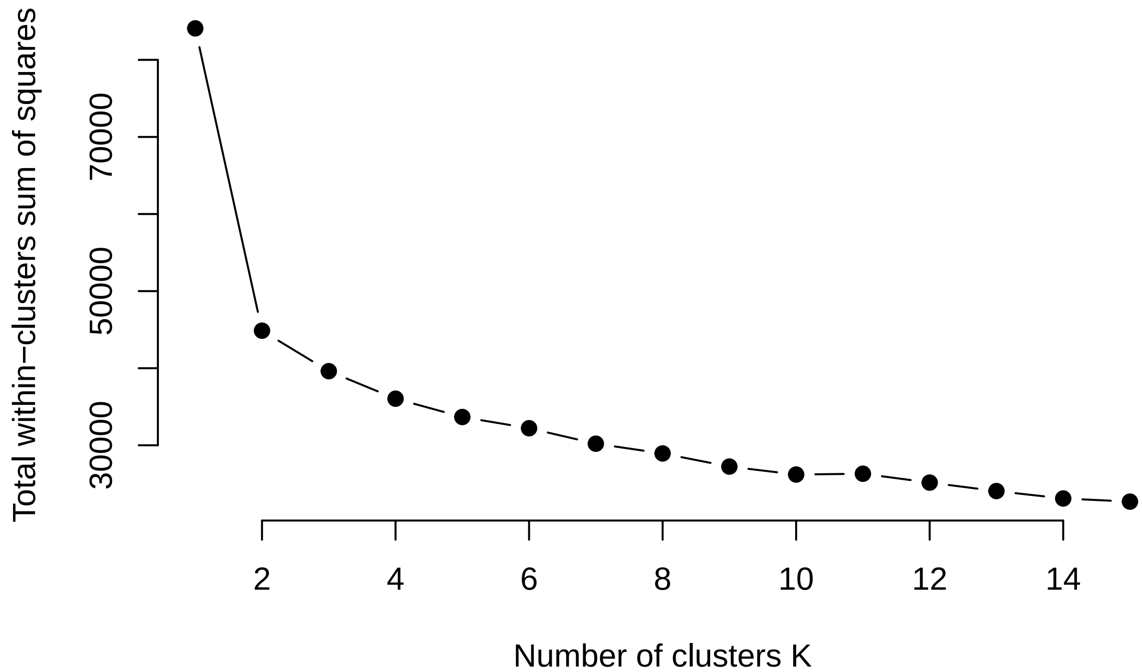
```
## 1  1.1985789         0.55290547      -0.57063689   -0.4541408   -0.4361311
## 2 -0.7212472        -0.06486729      -0.78903636    0.4998369    0.1486951
## 3  2.7926271         0.45990825      -0.06024275   -1.1504338   -1.1768823
## 4 -0.4442487         1.14318298       2.67468822   -0.4446190   -0.1516658
## 5  0.6898772         0.34499368      -0.06740573   -0.9887407   -1.1061760
## 6 -1.1289054         0.84917904       0.07354603   -1.1864923   -1.2223057
```

```
summary(data[1:5])
```

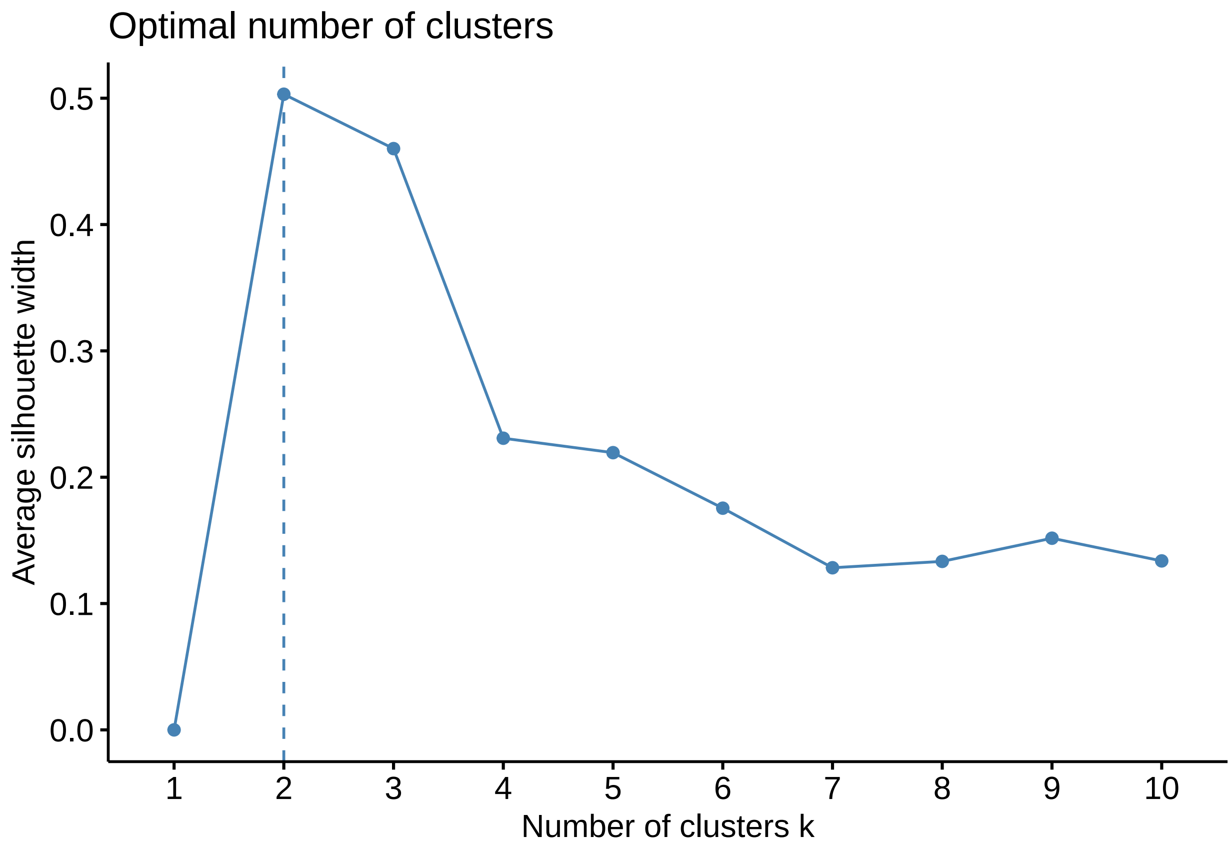
```
##     Failure         Entropy_cooc.W.ADC   GLNU_align.H.PET   Min_hist.PET
##  Min.   :-1.1289    Min.   :-2.6407173   Min.   :-0.9982   Min.   :-1.4098
##  1st Qu.:-0.7892    1st Qu.:-0.6921994   1st Qu.:-0.6721   1st Qu.:-0.6742
##  Median :-0.3066    Median : 0.0001827   Median :-0.1783   Median :-0.2256
##  Mean   : 0.0000    Mean   : 0.0000000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.6028    3rd Qu.: 0.6719746   3rd Qu.: 0.1947   3rd Qu.: 0.4998
##  Max.   : 3.7247    Max.   : 2.1464095   Max.   : 5.3894   Max.   : 3.9898
##   Max_hist.PET
##  Min.   :-1.3604
##  1st Qu.:-0.7578
##  Median :-0.2204
##  Mean   : 0.0000
##  3rd Qu.: 0.6421
##  Max.   : 3.7697
```

Model 3.1: K-Means

```
# Determining Optimal Number of Clusters
set.seed(123)

# Function to compute total within-cluster sum of square
wss <- function(k) {
  kmeans(data, k, nstart = 10)$tot.withinss
}

# Compute and plot wss for k = 1 to k = 15
k.values <- 1:15

# extract wss for 1-15 clusters
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")
```



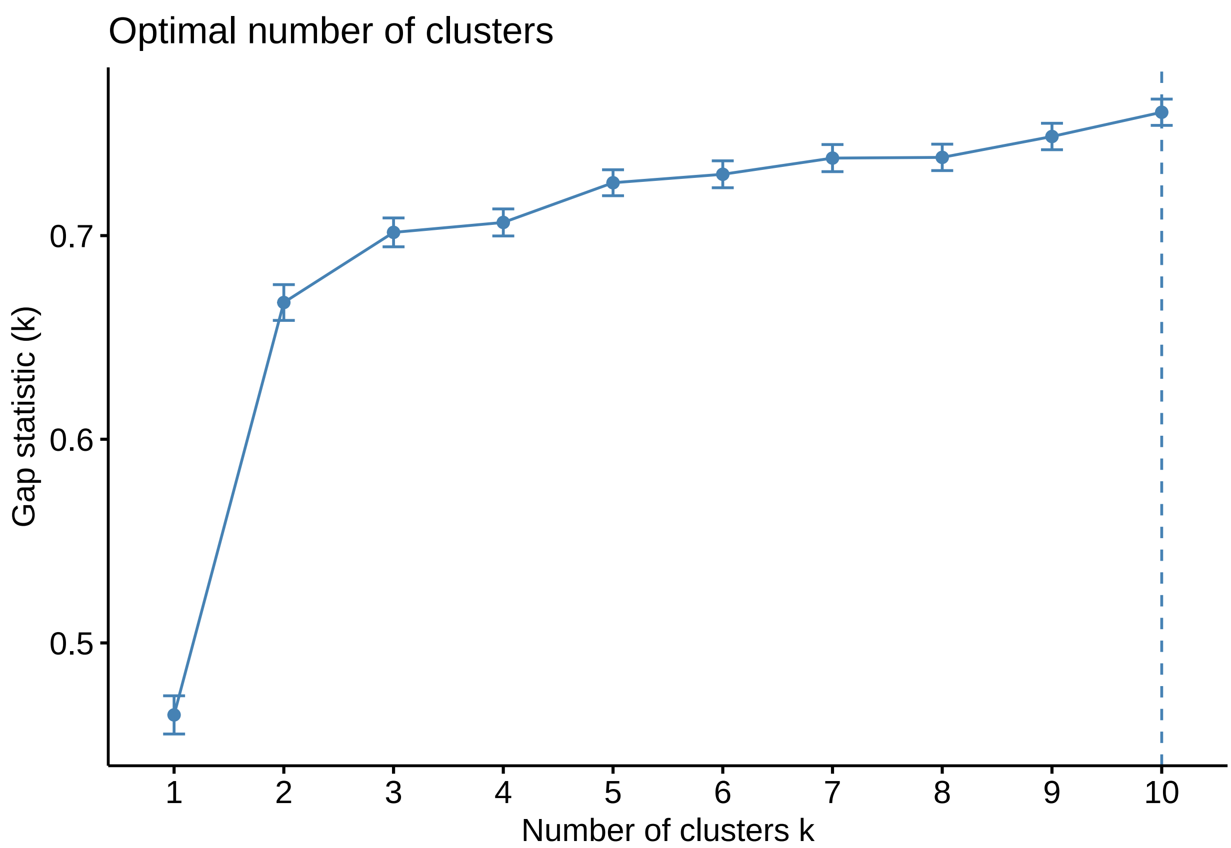

```r
# or use this
fviz_nbclust(data, kmeans, method = "silhouette")
```




```
# compute gap statistic
set.seed(123)
gap_stat <- clusGap(data, FUN = kmeans, nstart = 25,
                    K.max = 10, B = 50)
# Print the result
print(gap_stat, method = "firstmax")
```

```
## Clustering Gap statistic ["clusGap"] from call:
## clusGap(x = data, FUNcluster = kmeans, K.max = 10, B = 50, nstart = 25)
## B=50 simulated reference sets, k = 1..10; spaceH0="scaledPCA"
##  --> Number of clusters (method 'firstmax'): 10
##           logW   E.logW       gap      SE.sim
##  [1,] 7.171204 7.635853 0.4646496 0.009379996
##  [2,] 6.879524 7.546674 0.6671493 0.008786338
##  [3,] 6.798848 7.500436 0.7015873 0.007082545
##  [4,] 6.760004 7.466467 0.7064633 0.006632270
##  [5,] 6.715614 7.441579 0.7259645 0.006374244
##  [6,] 6.689522 7.419633 0.7301115 0.006603869
##  [7,] 6.661683 7.399745 0.7380616 0.006654018
##  [8,] 6.643211 7.381624 0.7384134 0.006480643
##  [9,] 6.616471 7.365139 0.7486677 0.006484664
## [10,] 6.588968 7.349544 0.7605765 0.006453097
```

```
fviz_gap_stat(gap_stat)
```




```r
# Compute k-means clustering with k = 2
set.seed(123)
k_means <- kmeans(data, 2, nstart = 25)
print(k_means)
```

```
## K-means clustering with 2 clusters of sizes 50, 147
##
## Cluster means:
##         Failure Entropy_cooc.W.ADC GLNU_align.H.PET Min_hist.PET Max_hist.PET
## 1 -0.0014733768         0.04845450      -0.07901100    0.9204612    0.9468341
## 2  0.0005011486        -0.01648112       0.02687449   -0.3130820   -0.3220524
##   Mean_hist.PET Variance_hist.PET Standard_Deviation_hist.PET Skewness_hist.PET
## 1     0.9216792         0.4594337                   0.9319222         0.9115602
## 2    -0.3134963        -0.1562700                  -0.3169804        -0.3100545
##   Kurtosis_hist.PET Energy_hist.PET Entropy_hist.PET AUC_hist.PET  H_suv.PET
## 1        0.25274217       0.6864958        1.5003007    1.6957546  0.9652219
## 2       -0.08596673      -0.2335020       -0.5103064   -0.5767873 -0.3283068
##   Volume.PET X3D_surface.PET ratio_3ds_vol.PET ratio_3ds_vol_norm.PET
## 1  0.5900077       0.3802612         0.9436984              0.9622506
## 2 -0.2006829      -0.1293406        -0.3209858             -0.3272961
##   irregularity.PET tumor_length.PET Compactness_v1.PET Compactness_v2.PET
## 1        1.6522842        1.0256292          0.8807232          0.4324058
## 2       -0.5620014       -0.3488535         -0.2995657         -0.1470768
##   Spherical_disproportion.PET Sphericity.PET Asphericity.PET Center_of_mass.PET
## 1                   0.9622506      0.4460709       0.9240341          0.6358358
## 2                  -0.3272961     -0.1517248      -0.3142973         -0.2162707
```

```
##   Max_3D_diam.PET Major_axis_length.PET Minor_axis_length.PET
## 1       0.8259982             0.8904297             1.1433164
## 2      -0.2809518            -0.3028672            -0.3888831
##   Least_axis_length.PET Elongation.PET Flatness.PET Max_cooc.L.PET
## 1             0.9772289      1.4563692    1.3553445      0.7290795
## 2            -0.3323908     -0.4953637   -0.4610015     -0.2479862
##   Average_cooc.L.PET Variance_cooc.L.PET Entropy_cooc.L.PET DAVE_cooc.L.PET
## 1           1.389215           1.1041050          1.6813985       1.2936781
## 2          -0.472522          -0.3755459         -0.5719043      -0.4400266
##   DVAR_cooc.L.PET DENT_cooc.L.PET SAVE_cooc.L.PET SVAR_cooc.L.PET
## 1       1.1366603       1.6603800       1.3889879       1.1209781
## 2      -0.3866192      -0.5647551      -0.4724449      -0.3812851
##   SENT_cooc.L.PET ASM_cooc.L.PET Contrast_cooc.L.PET Dissimilarity_cooc.L.PET
## 1       1.6614758      0.6775498           0.9285775                1.2936781
## 2      -0.5651278     -0.2304591          -0.3158427               -0.4400266
##   Inv_diff_cooc.L.PET Inv_diff_norm_cooc.L.PET IDM_cooc.L.PET
## 1            1.443028                1.6979660      1.2814891
## 2           -0.490826               -0.5775395     -0.4358807
##   IDM_norm_cooc.L.PET Inv_var_cooc.L.PET Correlation_cooc.L.PET
## 1           1.7046571          1.2896785               1.123648
## 2          -0.5798153         -0.4386661              -0.382193
##   Autocorrelation_cooc.L.PET Tendency_cooc.L.PET Shade_cooc.L.PET
## 1                  1.0338012           1.1209781        0.5578271
## 2                 -0.3516331          -0.3812851       -0.1897371
##   Prominence_cooc.L.PET IC1_.L.PET IC2_.L.PET Coarseness_vdif_.L.PET
## 1             0.7889007 -0.6341334  1.5273752              0.7537450
## 2            -0.2683336  0.2156916 -0.5195154             -0.2563758
##   Contrast_vdif_.L.PET Busyness_vdif_.L.PET Complexity_vdif_.L.PET
## 1            0.3878173            0.5565230              1.2153015
## 2           -0.1319107           -0.1892936             -0.4133678
##   Strength_vdif_.L.PET SRE_align.L.PET LRE_align.L.PET GLNU_align.L.PET
## 1            0.4934069        1.706523       1.6948229        0.4587983
## 2           -0.1678255       -0.580450      -0.5764704       -0.1560539
##   RLNU_align.L.PET RP_align.L.PET LGRE_align.L.PET HGRE_align.L.PET
## 1        0.4189336      1.7061400        1.0408063        1.0700373
## 2       -0.1424944     -0.5803197       -0.3540158       -0.3639583
##   LGSRE_align.L.PET HGSRE_align.L.PET LGHRE_align.L.PET HGLRE_align.L.PET
## 1          1.048281         1.0672364         1.0052958          1.078233
## 2         -0.356558        -0.3630056        -0.3419373         -0.366746
##   GLNU_norm_align.L.PET RLNU_norm_align.L.PET GLVAR_align.L.PET
## 1             1.1041018             1.7034139         1.1510468
## 2            -0.3755448            -0.5793925        -0.3915125
##   RLVAR_align.L.PET Entropy_align.L.PET SZSE.L.PET LZSE.L.PET LGLZE.L.PET
## 1         1.0474522           1.6880661  1.6676802  1.1852630   1.0601400
## 2        -0.3562762          -0.5741722 -0.5672382 -0.4031507  -0.3605919
##   HGLZE.L.PET SZLGE.L.PET SZHGE.L.PET LZLGE.L.PET LZHGE.L.PET GLNU_area.L.PET
## 1   1.0866745   1.0735299   1.0776043   0.8457163   0.8914749       0.4621309
## 2  -0.3696172  -0.3651462  -0.3665321  -0.2876586  -0.3032228      -0.1571874
##   ZSNU.L.PET ZSP.L.PET GLNU_norm.L.PET ZSNU_norm.L.PET GLVAR_area.L.PET
## 1  0.4218710  1.679008       1.1042309        1.681848        1.1694826
## 2 -0.1434935 -0.571091      -0.3755887       -0.572057       -0.3977832
##   ZSVAR.L.PET Entropy_area.L.PET Max_cooc.H.PET Average_cooc.H.PET
## 1   0.7548095          1.6893793      0.5052232          1.6652563
## 2  -0.2567379         -0.5746188     -0.1718446         -0.5664137
```

```
##   Variance_cooc.H.PET Entropy_cooc.H.PET DAVE_cooc.H.PET DVAR_cooc.H.PET
## 1           1.4721984          1.4404122       1.5079528       1.4645709
## 2          -0.5007478         -0.4899361      -0.5129091      -0.4981534
##   DENT_cooc.H.PET SAVE_cooc.H.PET SVAR_cooc.H.PET SENT_cooc.H.PET
## 1       1.3368883       1.6782221       1.4484331       1.1582831
## 2      -0.4547239      -0.5708239      -0.4926643      -0.3939739
##   ASM_cooc.H.PET Contrast_cooc.H.PET Dissimilarity_cooc.H.PET
## 1      0.4701159            1.344935                1.5079528
## 2     -0.1599034           -0.457461               -0.5129091
##   Inv_diff_cooc.H.PET Inv_diff_norm_cooc.H.PET IDM_cooc.H.PET
## 1           1.1377441                1.6996628      0.9576980
## 2          -0.3869878               -0.5781166     -0.3257476
##   IDM_norm_cooc.H.PET Inv_var_cooc_.H.PET Correlation_cooc.H.PET
## 1           1.7052806           0.9554037              1.1365587
## 2          -0.5800274          -0.3249672             -0.3865846
##   Autocorrelation_cooc.H.PET Tendency_cooc.H.PET Shade_cooc.H.PET
## 1                  1.5649714           1.4092944       -0.7124616
## 2                 -0.5323032          -0.4793518        0.2423339
##   Prominence_cooc.H.PET IC1_d.H.PET IC2_d.H.PET Coarseness_vdif.H.PET
## 1             1.0427158 -0.23095606   1.3345708             0.6663547
## 2            -0.3546653  0.07855648  -0.4539356            -0.2266512
##   Contrast_vdif.H.PET Busyness_vdif.H.PET Complexity_vdif.H.PET
## 1           0.4860224          0.25301766             1.0958360
## 2          -0.1653138         -0.08606043            -0.3727333
##   Strength_vdif.H.PET SRE_align.H.PET LRE_align.H.PET RLNU_align.H.PET
## 1          0.03112072       1.6638495       1.0890098        0.4166644
## 2         -0.01058528      -0.5659352      -0.3704115       -0.1417226
##   RP_align.H.PET LGRE_align.H.PET HGRE_align.H.PET LGSRE_align.H.PET
## 1      1.6436641        0.7082866        1.5743684         0.7040204
## 2     -0.5590694       -0.2409138       -0.5354994        -0.2394627
##   HGSRE_align.H.PET LGHRE_align.H.PET HGLRE_align.H.PET GLNU_norm_align.H.PET
## 1         1.6533952         0.7311054         0.7453460             0.8572435
## 2        -0.5623793        -0.2486753        -0.2535191            -0.2915794
##   RLNU_norm_align.H.PET GLVAR_align.H.PET RLVAR_align.H.PET Entropy_align.H.PET
## 1             1.5584253         1.4161797         0.4776867            1.550297
## 2            -0.5300766        -0.4816938        -0.1624785           -0.527312
##   SZSE.H.PET  LZSE.H.PET LGLZE.H.PET HGLZE.H.PET SZLGE.H.PET SZHGE.H.PET
## 1  1.4671263 -0.09759617   0.7096710   1.4890573   0.6984264   1.4294579
## 2 -0.4990226  0.03319598  -0.2413847  -0.5064821  -0.2375600  -0.4862102
##    LZLGE.H.PET LZHGE.H.PET GLNU_area.H.PET ZSNU.H.PET  ZSP.H.PET
## 1  0.001044652 -0.08592571       0.4835029  0.3648643  1.1565208
## 2 -0.000355324  0.02922643      -0.1644568 -0.1241035 -0.3933744
##   GLNU_norm.H.PET ZSNU_norm.H.PET GLVAR_area.H.PET ZSVAR_H.PET
## 1       0.8791603       1.2441418        1.3802703 -0.09449223
## 2      -0.2990341      -0.4231775       -0.4694797  0.03214021
##   Entropy_area.H.PET Max_cooc.W.PET Average_cooc.W.PET Variance_cooc.W.PET
## 1          1.6279234      0.5502762          0.9151412           0.4579807
## 2         -0.5537154     -0.1871688         -0.3112725          -0.1557757
##   Entropy_cooc.W.PET DAVE_cooc.W.PET DVAR_cooc.W.PET DENT_cooc.W.PET
## 1          1.4784780       0.9564701       0.5165571        1.450023
## 2         -0.5028837      -0.3253300      -0.1756997       -0.493205
##   SAVE_cooc.W.PET SVAR_cooc.W.PET SENT_cooc.W.PET ASM_cooc.W.PET
## 1       0.9140050       0.4135667       1.5336398      0.5955603
## 2      -0.3108861      -0.1406689      -0.5216462     -0.2025715
```

```
##   Contrast_cooc.W.PET Dissimilarity_cooc.W.PET Inv_diff_cooc.W.PET
## 1           0.5325478                0.9564701           1.2750883
## 2          -0.1811387               -0.3253300          -0.4337035
##   Inv_diff_norm_cooc.W.PET IDM_cooc.W.PET IDM_norm_cooc.W.PET
## 1                1.6983343       1.044167           1.7048157
## 2               -0.5776647      -0.355159          -0.5798693
##   Inv_var_cooc.W.PET Correlation_cooc.W.PET Autocorrelation_cooc.W.PET
## 1          1.1637708              1.1228422                  0.4576739
## 2         -0.3958404             -0.3819191                 -0.1556714
##   Tendency_cooc.W.PET Shade_cooc.W.PET Prominence_cooc.W.PET IC1_d.W.PET
## 1           0.4135667       0.07642004           0.022900737 -0.26887955
## 2          -0.1406689      -0.02599321          -0.007789366  0.09145563
##   IC2_d.W.PET Coarseness_vdif.W.PET Contrast_vdif.W.PET Busyness_vdif.W.PET
## 1   1.4455561             0.7071892           0.8252351           0.4153574
## 2  -0.4916858            -0.2405405          -0.2806922          -0.1412780
##   Complexity_vdif.W.PET Strength_vdif.W.PET SRE_align.W.PET LRE_align.W.PET
## 1             0.2991726           0.4249851        1.697315       1.4801473
## 2            -0.1017594          -0.1445527       -0.577318      -0.5034515
##   GLNU_align.W.PET RLNU_align.W.PET RP_align.W.PET LGRE_align.W.PET
## 1        0.4738278        0.4182280      1.6901986        0.8300003
## 2       -0.1611659       -0.1422544     -0.5748975       -0.2823130
##   HGRE_align.W.PET LGSRE_align.W.PET HGSRE_align.W.PET LGHRE_align.W.PET
## 1        0.4630749         0.8904857         0.4557129         0.5563026
## 2       -0.1575085        -0.3028863        -0.1550044        -0.1892186
##   HGLRE_align.W.PET GLNU_norm_align.W.PET RLNU_norm_align.W.PET
## 1         0.4921754             0.8494549              1.658483
## 2        -0.1674066            -0.2889302             -0.564110
##   GLVAR_align.W.PET RLVAR_align.W.PET Entropy_align.W.PET SZSE.W.PET
## 1         0.4593218         0.5957178           1.5543465  1.6121174
## 2        -0.1562319        -0.2026251          -0.5286893 -0.5483392
##    LZSE.W.PET LGLZE.W.PET HGLZE.W.PET SZLGE.W.PET SZHGE.W.PET  LZLGE.W.PET
## 1  0.21517025   0.8709408   0.4690713   0.9938480   0.4481637 -0.004326372
## 2 -0.07318716  -0.2962384  -0.1595481  -0.3380435  -0.1524366  0.001471555
##   LZHGE.W.PET GLNU_area.W.PET ZSNU.W.PET  ZSP.W.PET GLNU_norm.W.PET
## 1   0.5263985       0.4910918  0.3971868  1.4948131       0.8826796
## 2  -0.1790471      -0.1670380 -0.1350976 -0.5084398      -0.3002311
##   ZSNU_norm.W.PET GLVAR_area.W.PET ZSVAR.W.PET Entropy_area.W.PET Min_hist.ADC
## 1       1.4869647        0.4655759  0.06408427          1.6167770    0.5724098
## 2      -0.5057703       -0.1583592 -0.02179737         -0.5499242   -0.1946972
##   Max_hist.ADC Mean_hist.ADC Variance_hist.ADC Standard_Deviation_hist.ADC
## 1    1.5075750     1.4864908         0.7599395                   1.2359485
## 2   -0.5127806    -0.5056091        -0.2584828                  -0.4203906
##   Skewness_hist.ADC Kurtosis_hist.ADC Energy_hist.ADC Entropy_hist.ADC
## 1         0.3899909         0.4662845       0.7015053        1.6284344
## 2        -0.1326500        -0.1586002      -0.2386073       -0.5538893
##   AUC_hist.ADC Volume.ADC X3D_surface.ADC ratio_3ds_vol.ADC
## 1    1.6655300  0.5687484       0.7349831         1.1042095
## 2   -0.5665068 -0.1934518      -0.2499942        -0.3755815
##   ratio_3ds_vol_norm.ADC irregularity.ADC Compactness_v1.ADC Compactness_v2.ADC
## 1              1.6106322        1.6397737          1.1221987          1.3007130
## 2             -0.5478341       -0.5577462         -0.3817002         -0.4424194
##   Spherical_disproportion.ADC Sphericity.ADC Asphericity.ADC Center_of_mass.ADC
## 1                   1.6106322      1.6242350       1.1989866          0.5373920
## 2                  -0.5478341     -0.5524609      -0.4078186         -0.1827864
```

```
##   Max_3D_diam.ADC Major_axis_length.ADC Minor_axis_length.ADC
## 1       1.0866100             1.2316275             1.1312333
## 2      -0.3695952            -0.4189209            -0.3847732
##   Least_axis_length.ADC Elongation.ADC Flatness.ADC Max_cooc.L.ADC
## 1             1.0417403      1.4824827    1.4052040      0.8250964
## 2            -0.3543334     -0.5042458   -0.4779606     -0.2806450
##   Average_cooc.L.ADC Variance_cooc.L.ADC Entropy_cooc.L.ADC DAVE_cooc.L.ADC
## 1           1.456079           0.9533869          1.6827114       1.2819538
## 2          -0.495265          -0.3242813         -0.5723508      -0.4360387
##   DVAR_cooc.L.ADC DENT_cooc.L.ADC SAVE_cooc.L.ADC SVAR_cooc.L.ADC
## 1       0.9295089       1.6521421       1.4558899       0.9317704
## 2      -0.3161595      -0.5619531      -0.4952006      -0.3169287
##   SENT_cooc.L.ADC ASM_cooc.L.ADC Contrast_cooc.L.ADC Dissimilarity_cooc.L.ADC
## 1       1.2584756      0.7127202           0.8811662                1.2819538
## 2      -0.4280529     -0.2424218          -0.2997164               -0.4360387
##   Inv_diff_cooc.L.ADC Inv_diff_norm_cooc.L.ADC IDM_cooc.L.ADC
## 1           1.5058302                1.7039344      1.3642322
## 2          -0.5121871               -0.5795695     -0.4640245
##   IDM_norm_cooc.L.ADC Inv_var_cooc.L.ADC Correlation_cooc.L.ADC
## 1           1.7073272           1.379898              1.2216811
## 2          -0.5807235          -0.469353             -0.4155378
##   Autocorrelation_.L.ADC Tendency_cooc.L.ADC Shade_.L.ADC Prominence_cooc.L.ADC
## 1              1.1050198           0.9317704   0.29259000             0.5515288
## 2             -0.3758571          -0.3169287  -0.09952041            -0.1875948
##   IC1_.L.ADC IC2_.L.ADC Coarseness_vdif_.L.ADC Contrast_vdif_.L.ADC
## 1 -0.6732168  1.5121032              0.6939723            0.6587722
## 2  0.2289853 -0.5143208             -0.2360450           -0.2240722
##   Busyness_vdif_.L.ADC Complexity_vdif_.L.ADC Strength_vdif_.L.ADC
## 1            0.6475886              1.2753146            0.4214397
## 2           -0.2202682             -0.4337805           -0.1433468
##   SRE_align.L.ADC LRE_align.L.ADC GLNU_align.L.ADC RLNU_align.L.ADC
## 1       1.7052408       1.6811893        0.5682374        0.5910147
## 2      -0.5800139      -0.5718331       -0.1932780       -0.2010254
##   RP_align.L.ADC LGRE_align.L.ADC HGRE_align.L.ADC LGSRE_align.L.ADC
## 1      1.7034645        0.7243458        1.2086645         0.7235521
## 2     -0.5794097       -0.2463761       -0.4111104        -0.2461061
##   HGSRE_align.L.ADC LGHRE_align.L.ADC HGLRE_align.L.ADC GLNU_norm_align.L.ADC
## 1         1.2124123         0.7234431         1.1801466             1.2291014
## 2        -0.4123852        -0.2460691        -0.4014104            -0.4180617
##   RLNU_norm_align.L.ADC GLVAR_align.L.ADC RLVAR_align.L.ADC Entropy_align.L.ADC
## 1             1.6955541         0.9930121         1.1385331           1.6982212
## 2            -0.5767191        -0.3377592        -0.3872562          -0.5776262
##   SZSE.L.ADC LZSE.L.ADC LGLZE.L.ADC HGLZE.L.ADC SZLGE.L.ADC SZHGE.L.ADC
## 1  1.6968578  1.3430968   0.7262967   1.2295659   0.7219542   1.2399482
## 2 -0.5771625 -0.4568356  -0.2470397  -0.4182197  -0.2455627  -0.4217511
##   LZLGE.L.ADC LZHGE.L.ADC GLNU_area.L.ADC ZSNU.L.ADC  ZSP.L.ADC GLNU_norm.L.ADC
## 1   0.6651854    1.077189       0.5782984  0.5919629  1.6748354       1.2251432
## 2  -0.2262535   -0.366391      -0.1967001 -0.2013479 -0.5696719      -0.4167154
##   ZSNU_norm.L.ADC GLVAR_area.L.ADC ZSVAR.L.ADC Entropy_area.L.ADC
## 1       1.6570978         1.012871   0.6758567          1.7010816
## 2      -0.5636387        -0.344514  -0.2298832         -0.5785992
##   Max_cooc.H.ADC Average_cooc.H.ADC Variance_cooc.H.ADC Entropy_cooc.H.ADC
## 1      0.7039103          1.6967547           1.7053247          1.7011475
## 2     -0.2394253         -0.5771274          -0.5800424         -0.5786216
```

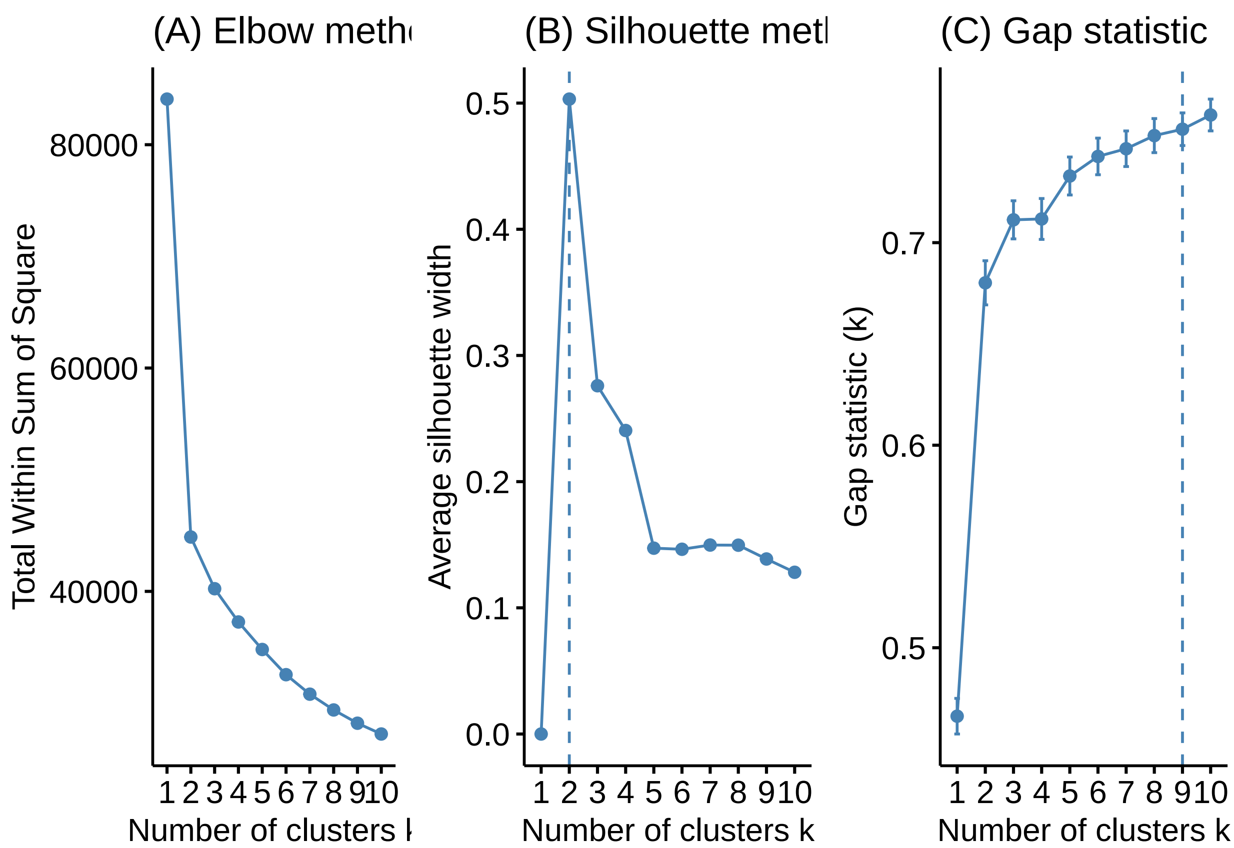
```
##   DAVE_cooc.H.ADC DVAR_cooc.H.ADC DENT_cooc.H.ADC SAVE_cooc.H.ADC
## 1       1.5698813       1.4861394       1.7017575       1.6967573
## 2      -0.5339732      -0.5054896      -0.5788291      -0.5771283
##   SVAR_cooc.H.ADC SENT_cooc.H.ADC ASM_cooc.H.ADC Contrast_cooc.H.ADC
## 1       1.6206816       1.6803084      0.6607170           1.3858879
## 2      -0.5512522      -0.5715335     -0.2247337          -0.4713904
##   Dissimilarity_cooc.H.ADC Inv_diff_cooc.H.ADC Inv_diff_norm_cooc.H.ADC
## 1                1.5698813           1.5546888                1.7028145
## 2               -0.5339732          -0.5288057               -0.5791886
##   IDM_cooc.H.ADC IDM_norm_cooc.H.ADC Inv_var_cooc.H.ADC Correlation_cooc.H.ADC
## 1      1.4136874           1.7054539          1.4364367              1.1993586
## 2     -0.4808461          -0.5800864         -0.4885839             -0.4079451
##   Autocorrelation_cooc.H.ADC Tendency_cooc.H.ADC Shade_cooc.H.ADC
## 1                  1.6722184           1.6206816        0.3887230
## 2                 -0.5687818          -0.5512522       -0.1322187
##   Prominence_cooc.H.ADC IC1_d.H.ADC IC2_d.H.ADC Coarseness_vdif.H.ADC
## 1             1.5404751  -0.5455177   1.5085932             0.6780216
## 2            -0.5239711   0.1855502  -0.5131269            -0.2306196
##   Contrast_vdif.H.ADC Busyness_vdif.H.ADC Complexity_vdif.H.ADC
## 1           1.5316725           0.6153610              1.503704
## 2          -0.5209771          -0.2093065             -0.511464
##   Strength_vdif.H.ADC SRE_align.H.ADC LRE_align.H.ADC GLNU_align.H.ADC
## 1           0.3677298       1.7071497       1.7038845        0.5901231
## 2          -0.1250782      -0.5806632      -0.5795526       -0.2007222
##   RLNU_align.H.ADC RP_align.H.ADC LGRE_align.H.ADC HGRE_align.H.ADC
## 1        0.5924412       1.706814        1.0946139        1.7100780
## 2       -0.2015106      -0.580549       -0.3723177       -0.5816592
##   LGSRE_align.H.ADC HGSRE_align.H.ADC LGHRE_align.H.ADC HGLRE_align.H.ADC
## 1         1.0760014         1.7093907         1.1710039         1.7053139
## 2        -0.3659869        -0.5814254        -0.3983006        -0.5800387
##   GLNU_norm_align.H.ADC RLNU_norm_align.H.ADC GLVAR_align.H.ADC
## 1             0.9735389             1.7053279         1.7100152
## 2            -0.3311357            -0.5800435        -0.5816378
##   RLVAR_align.H.ADC Entropy_align.H.ADC SZSE.H.ADC LZSE.H.ADC LGLZE.H.ADC
## 1         1.0687509           1.7093530  1.7049082  1.6336887   1.0589022
## 2        -0.3635207          -0.5814126 -0.5799008 -0.5556764  -0.3601708
##   HGLZE.H.ADC SZLGE.H.ADC SZHGE.H.ADC LZLGE.H.ADC LZHGE.H.ADC GLNU_area.H.ADC
## 1    1.709075   1.0114862   1.7031396   1.0813161   1.5698347       0.5919958
## 2   -0.581318  -0.3440429  -0.5792992  -0.3677946  -0.5339574      -0.2013591
##   ZSNU.H.ADC  ZSP.H.ADC GLNU_norm.H.ADC ZSNU_norm.H.ADC GLVAR_area.H.ADC
## 1  0.5972096  1.7013318       0.9745507        1.692802        1.7072803
## 2 -0.2031325 -0.5786843      -0.3314798       -0.575783       -0.5807076
##   ZSVAR.H.ADC Entropy_area.H.ADC Max_cooc.W.ADC Average_cooc.W.ADC
## 1   0.8431301          1.7066118      0.6868122           1.199285
## 2  -0.2867790         -0.5804802     -0.2336096          -0.407920
##   Variance_cooc.W.ADC DAVE_cooc.W.ADC DVAR_cooc.W.ADC DENT_cooc.W.ADC
## 1           0.7283676       1.3033631       0.7679414       1.6768624
## 2          -0.2477441      -0.4433208      -0.2612045      -0.5703613
##   SAVE_cooc.W.ADC SVAR_cooc.W.ADC SENT_cooc.W.ADC ASM_cooc.W.ADC
## 1       1.1909017       0.6843706       1.2023295      0.6601442
## 2      -0.4050686      -0.2327791      -0.4089556     -0.2245389
##   Contrast_cooc.W.ADC Dissimilarity_cooc.W.ADC Inv_diff_cooc.W.ADC
## 1           0.7994120                1.3033631           1.3827605
## 2          -0.2719088               -0.4433208          -0.4703267
```

```
##   Inv_diff_norm_cooc.W.ADC IDM_cooc.W.ADC IDM_norm_cooc.W.ADC
## 1                1.7038802      1.3112119           1.7073083
## 2               -0.5795511     -0.4459904          -0.5807171
##   Inv_var_cooc.W.ADC Correlation_cooc.W.ADC Autocorrelation_cooc.W.ADC
## 1          1.3074526              1.2225367                  0.8447953
## 2         -0.4447118             -0.4158288                 -0.2873453
##   Tendency_cooc.W.ADC Shade_cooc.W.ADC Prominence_cooc.W.ADC IC1_d.W.ADC
## 1           0.6843706        0.2567335             0.3775512  -0.6756692
## 2          -0.2327791       -0.0873243            -0.1284188   0.2298194
##   IC2_d.W.ADC Coarseness_vdif.W.ADC Contrast_vdif.W.ADC Busyness_vdif.W.ADC
## 1   1.6012140             0.7114542           0.6249552           1.0116700
## 2  -0.5446306            -0.2419912          -0.2125698          -0.3441054
##   Complexity_vdif.W.ADC Strength_vdif.W.ADC SRE_align.W.ADC LRE_align.W.ADC
## 1             0.6003182           0.5784705       1.7073214       1.7065667
## 2            -0.2041899          -0.1967587      -0.5807216      -0.5804649
##   GLNU_align.W.ADC RLNU_align.W.ADC RP_align.W.ADC LGRE_align.W.ADC
## 1        0.6326468        0.5857336      1.7071535        0.6918953
## 2       -0.2151860       -0.1992291     -0.5806645       -0.2353386
##   HGRE_align.W.ADC LGSRE_align.W.ADC HGSRE_align.W.ADC LGHRE_align.W.ADC
## 1        0.8626770         0.6918084         0.8616174         0.6894568
## 2       -0.2934276        -0.2353090        -0.2930672        -0.2345091
##   HGLRE_align.W.ADC GLNU_norm_align.W.ADC RLNU_norm_align.W.ADC
## 1          0.866512             0.9154487             1.7063312
## 2         -0.294732            -0.3113771            -0.5803848
##   GLVAR_align.W.ADC RLVAR_align.W.ADC Entropy_align.W.ADC SZSE.W.ADC LZSE.W.ADC
## 1         0.7640782         0.9834635            1.661714  1.7066974  1.6823970
## 2        -0.2598905        -0.3345114           -0.565209 -0.5805093 -0.5722439
##   LGLZE.W.ADC HGLZE.W.ADC SZLGE.W.ADC SZHGE.W.ADC LZLGE.W.ADC LZHGE.W.ADC
## 1   0.6918923   0.8639228   0.6899145   0.8602645   0.6450074   0.8755515
## 2  -0.2353375  -0.2938513  -0.2346648  -0.2926070  -0.2193903  -0.2978066
##   GLNU_area.W.ADC ZSNU.W.ADC  ZSP.W.ADC GLNU_norm.W.ADC ZSNU_norm.W.ADC
## 1       0.6327545  0.5822861  1.7050925       0.9137899        1.699026
## 2      -0.2152226 -0.1980565 -0.5799634      -0.3108129       -0.577900
##   GLVAR_area.W.ADC ZSVAR.W.ADC Entropy_area.W.ADC
## 1        0.7713592   1.0785430           1.672228
## 2       -0.2623671  -0.3668514          -0.568785
##
## Clustering vector:
##   [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
##  [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [112] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1
## [149] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [186] 1 1 1 1 1 1 1 1 1 1 1 1
##
## Within cluster sum of squares by cluster:
## [1] 21058.70 23808.27
##  (between_SS / total_SS =  46.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```r
# Plot of Final kmeans clustering
kmeans_plot <- fviz_cluster(k_means, data = data) +
  ggtitle("(3.1) K-Means Clustering")
kmeans_plot
```

Model 3.2: Hierarchical

```r
set.seed(123)

# Plot cluster results
p1 <- fviz_nbclust(data, FUN = hcut, method = "wss",
                   k.max = 10) +
  ggtitle("(A) Elbow method")
p2 <- fviz_nbclust(data, FUN = hcut, method = "silhouette",
                   k.max = 10) +
  ggtitle("(B) Silhouette method")
p3 <- fviz_nbclust(data, FUN = hcut, method = "gap_stat",
                   k.max = 10) +
  ggtitle("(C) Gap statistic")

# Display plots side by side
gridExtra::grid.arrange(p1, p2, p3, nrow = 1)
```



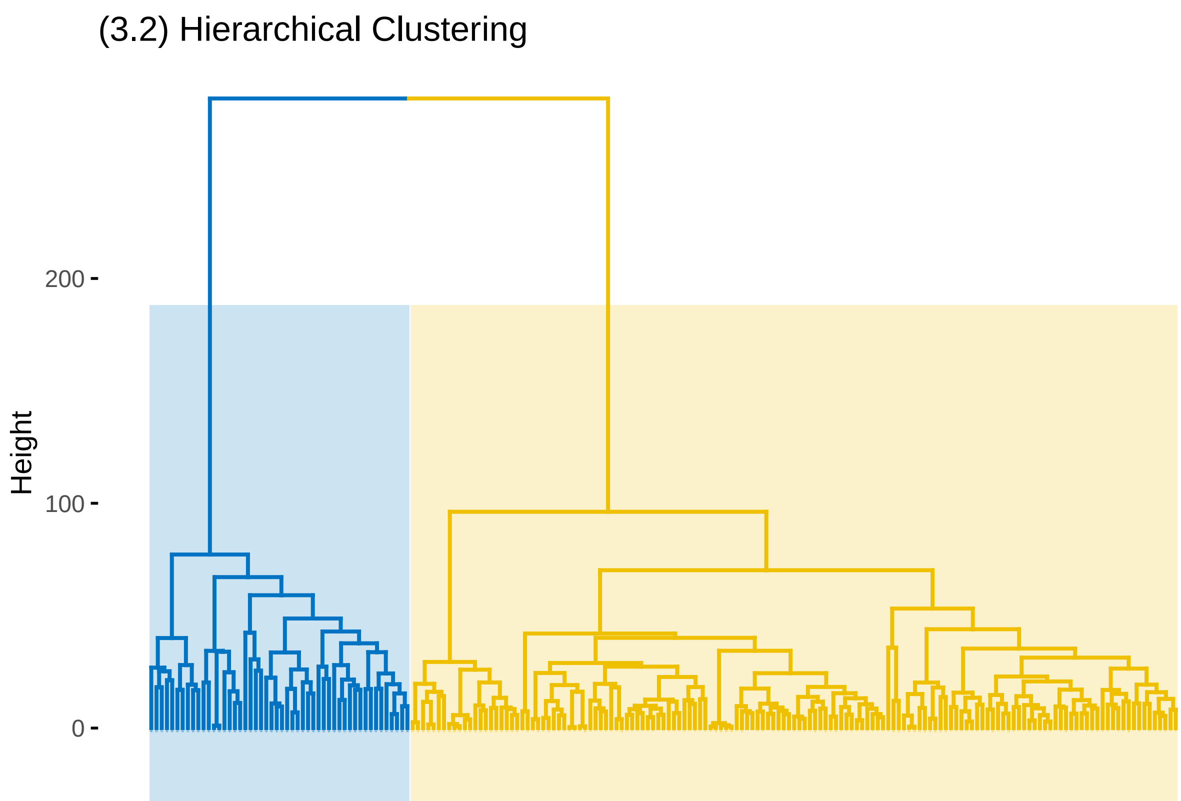

```r
# Dissimilarity matrix
d <- dist(data, method = "euclidean")

# Construct Hierarchical clustering
hierarchical <- hclust(d, method = "ward.D2" )
summary(hierarchical)
```

```
##             Length Class  Mode
## merge       392    -none- numeric
## height      196    -none- numeric
## order       197    -none- numeric
## labels        0    -none- NULL
## method        1    -none- character
## call          3    -none- call
## dist.method   1    -none- character
```

```r
# Cut tree into 2 groups
sub_grp <- cutree(hierarchical, k = 2)

# Number of members in each cluster
table(sub_grp)
```

```
## sub_grp
##   1   2
## 147  50
```

```r
# Plot full dendogram
hierarchical_plot <- fviz_dend(
  hierarchical,
  k = 2,
  horiz = FALSE,
  rect = TRUE,
  rect_fill = TRUE,
  rect_border = "jco",
  k_colors = "jco",
  cex = 0.1
) +
  ggtitle("(3.2) Hierarchical Clustering")
hierarchical_plot
```



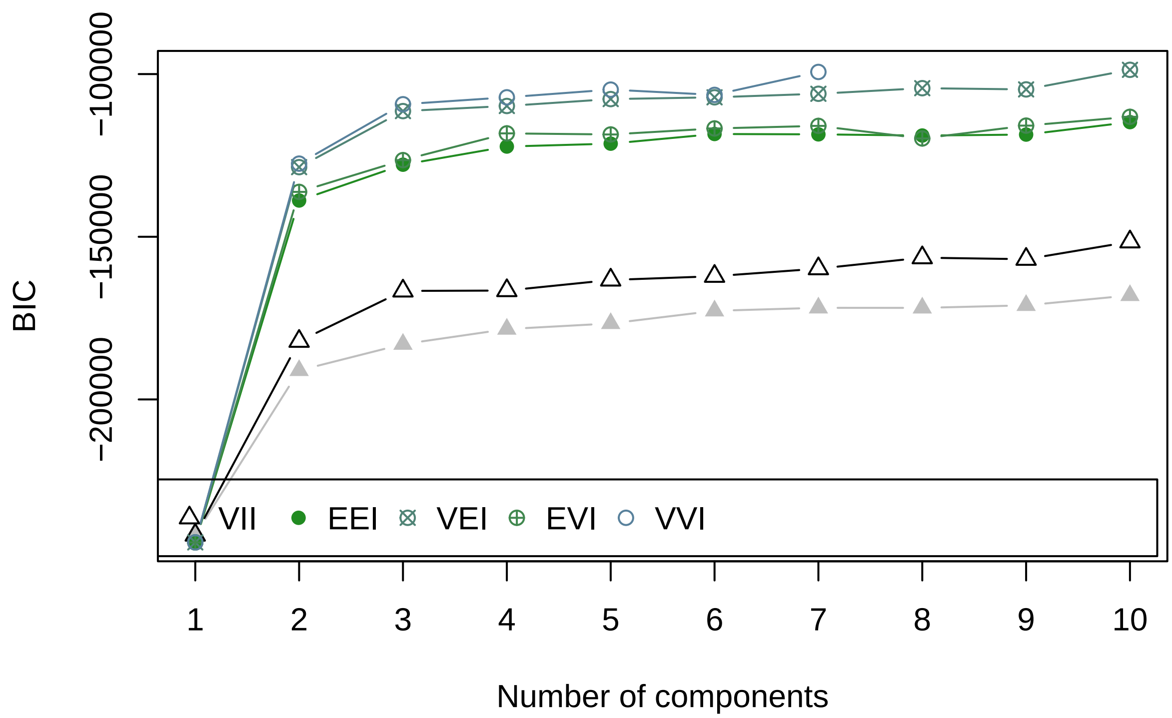
Model 3.3: Model-Based

```r
  # Apply GMM model with 10 components
set.seed(123)
radiomics_mc <- Mclust(data, 1:10)
summary(radiomics_mc)
```

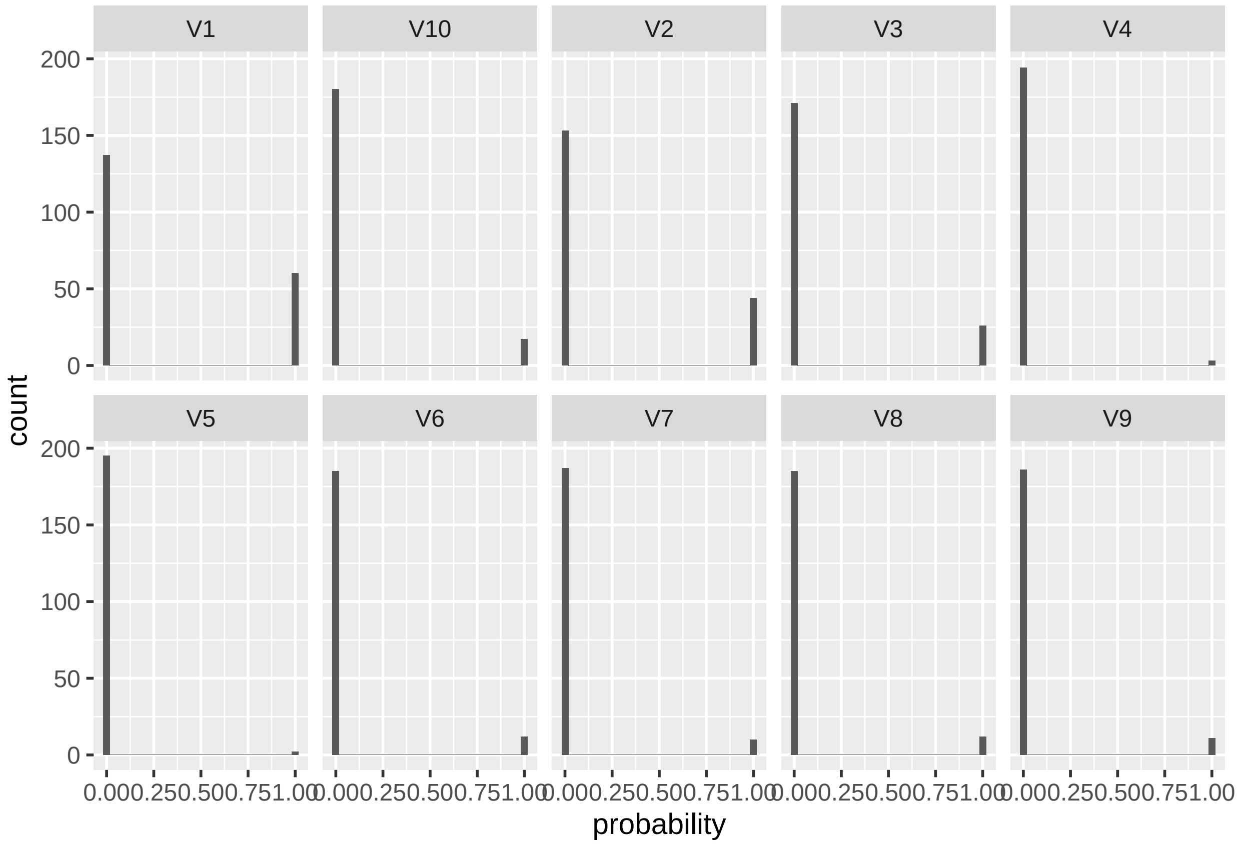
```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
```

```
## -----------------------------------------------------
##
## Mclust VEI (diagonal, equal shape) model with 10 components:
##
##  log-likelihood   n   df       BIC       ICL
##       -36831.86 197 4737 -98690.26 -98690.27
##
## Clustering table:
##  1  2  3  4  5  6  7  8  9 10
## 60 44 26  3  2 12 10 12 11 17
```

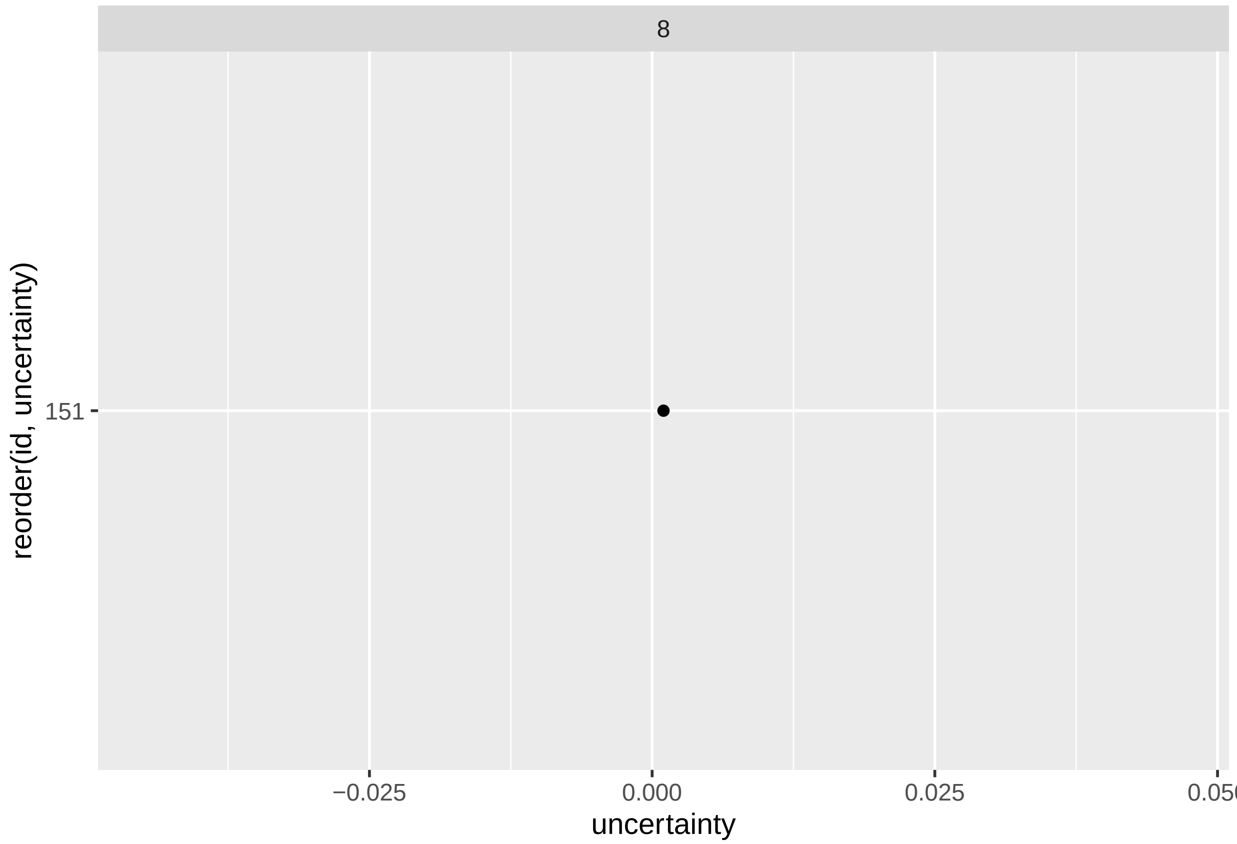
```
plot(radiomics_mc, what = 'BIC',
     legendArgs = list(x = "bottomright", ncol = 10))
```



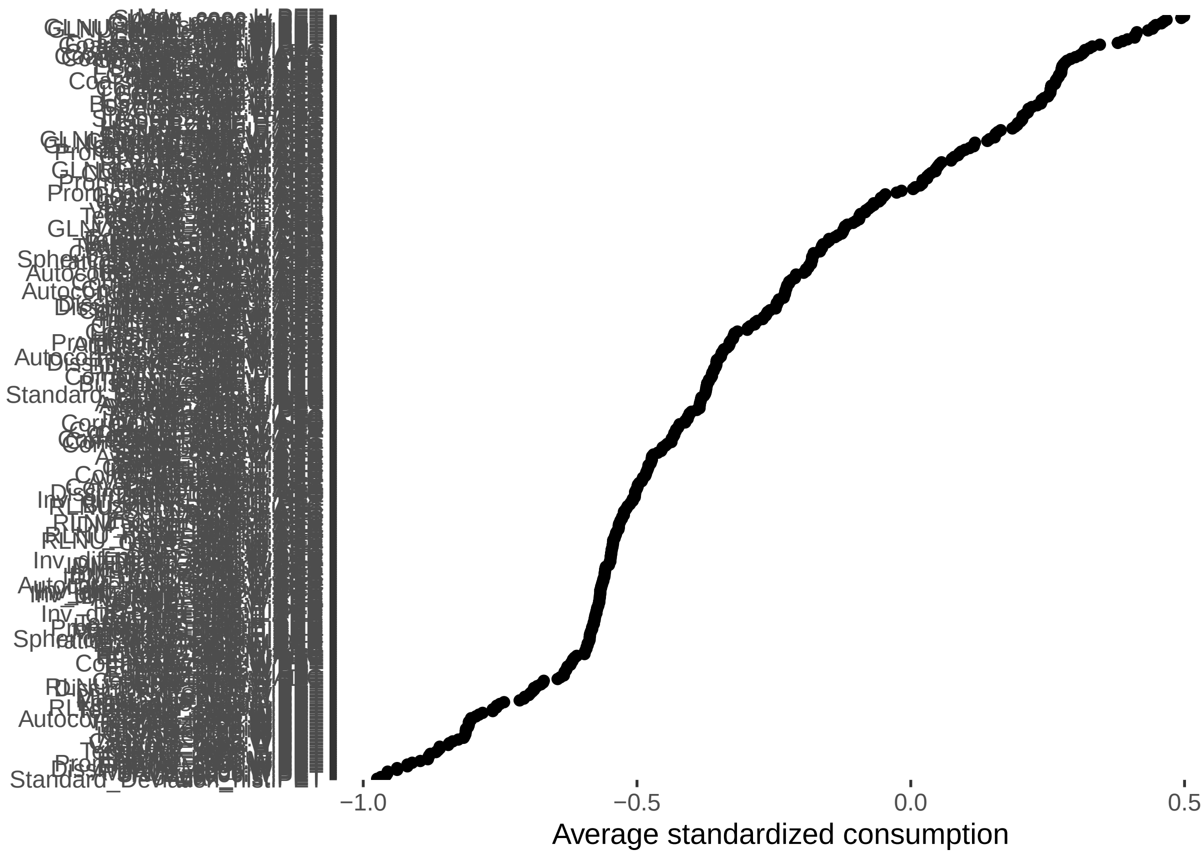

```
probabilities <- radiomics_mc$z

probabilities <- probabilities %>%
  as.data.frame() %>%
  mutate(id = row_number()) %>%
  tidyr::gather(cluster, probability, -id)

ggplot(probabilities, aes(probability)) +
  geom_histogram() +
  facet_wrap(~ cluster, nrow = 2)
```




```
uncertainty <- data.frame(
  id = 1:nrow(data),
  cluster = radiomics_mc$classification,
  uncertainty = radiomics_mc$uncertainty
)

uncertainty %>%
  group_by(cluster) %>%
  filter(uncertainty > 0.0001) %>%
  ggplot(aes(uncertainty, reorder(id, uncertainty))) +
  geom_point() +
  facet_wrap(~ cluster, scales = 'free_y', nrow = 1)
```




```
cluster2 <- data %>%
  scale() %>%
  as.data.frame() %>%
  mutate(cluster = radiomics_mc$classification) %>%
  filter(cluster == 2) %>%
  select(-cluster)

cluster2 %>%
  tidyr::gather(product, std_count) %>%
  group_by(product) %>%
  summarize(avg = mean(std_count)) %>%
  ggplot(aes(avg, reorder(product, avg))) +
  geom_point() +
  labs(x = "Average standardized consumption", y = NULL)
```


−1.0
−0.5
0.0
0.5
Average standardized consumption