Simulated examples for the survival ensemble methods

Select data type

ELSA_Diabetes_Study2

Time point for event prediction:

10

Random seed for calibration and validation

42

K_Outer loop CV (for validation)

3

K_ inner CV folds (model tuning)

3

Simulated data: random seed (generation):

4242

Sample size:

150

Observation time

5

Expected event prevalence by study end

0.5

Expected drop out rate

0.3

Custom data: path to data file

~/Desktop/Study_KCL/PhD P

Predictors to use in the model

"baseline_age_", "genderdur

Time variable name

time

Event indicator variable name

event

Sample statistics | CoxPH | SRF | Ens1: CoxPH->SRF | Ens2: CoxPH in clusters | Ens3: extended CoxPH | Summary | Conclusions

Internally cross-validated results:

Show 25 entries Search:

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test	0.7744	0.0775	0.0942	0.7442	1.2128	0.0401	10
train	0.9315	0.0631	0.262	0.8955	3.0708	0.0368	10

Showing 1 to 2 of 2 entries Previous 1 Next

Internally cross-validated Test results for each CV fold:

Show 25 entries Search:

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test.1	0.7663	0.0868	0.0948	0.7405	1.2223	0.0652	10
test.2	0.7611	0.0752	0.0824	0.7224	1.1853	0.033	10
test.3	0.7959	0.0707	0.1054	0.7699	1.2307	0.0221	10

Showing 1 to 3 of 3 entries Previous 1 Next

```
$test
   T    AUCROC         BS  BS_scaled   C_score
1 10 0.7662861 0.08675366 0.09483187 0.7404850
2 10 0.7610997 0.07517238 0.08243972 0.7223657
3 10 0.7959463 0.07070586 0.10541233 0.7698869
  Calib_slope Calib_alpha test cv_n
1    1.222318  0.06523804    1    1
2    1.185295  0.03300928    1    2
3    1.230688  0.02211466    1    3

$train
   T    AUCROC        BS BS_scaled   C_score Calib_slope
1 10 0.9164321 0.0627577 0.2237111 0.8780920    2.651160
2 10 0.9114885 0.0689563 0.2131885 0.8776219    2.782932
3 10 0.9666902 0.0575690 0.3491984 0.9309040    3.778457
  Calib_alpha test cv_n
1  0.03423430    0    1
2  0.04119281    0    2
3  0.03488901    0    3

$testaverage
          T      AUCROC          BS   BS_scaled
10.00000000  0.77444405  0.07754397  0.09422797
    C_score Calib_slope Calib_alpha        test
 0.74424587  1.21276679  0.04012066  1.00000000

$trainaverage
          T      AUCROC          BS   BS_scaled
10.00000000  0.93153691  0.06309433  0.26203266
    C_score Calib_slope Calib_alpha        test
 0.89553930  3.07084965  0.03677204  0.00000000

$model_list
$model_list[[1]]
$model_list[[1]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V1    3       45        50  8.9 0.7574443 0.06423064
    BS_scaled   C_score Calib_alpha Calib_slope
V1 0.07790647 0.7316096  0.03298471    1.167896

$model_list[[1]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1     5       15        50  8.9 0.7524432 0.06409274
V2     5       20        50  8.9 0.7491188 0.06423795
V3     5       25        50  8.9 0.7500564 0.06436698
V4     5       30        50  8.9 0.7473205 0.06441021
V5     5       35        50  8.9 0.7520207 0.06436109
V6     5       40        50  8.9 0.7506937 0.06429571
V7     5       45        50  8.9 0.7555107 0.06418097
V8     5       50        50  8.9 0.7526464 0.06418817
V11    3       45        50  8.9 0.7574443 0.06423064
V21    5       45        50  8.9 0.7555107 0.06418097
V31    7       45        50  8.9 0.7485881 0.06444108
V41   10       45        50  8.9 0.7489870 0.06479384
V51   15       45        50  8.9 0.7411255 0.06515950
```

```
     BS_scaled   C_score Calib_alpha Calib_slope
V1  0.07988610 0.7266474  0.02987083   0.8803919
V2  0.07780145 0.7218242  0.03128357   0.8786895
V3  0.07594906 0.7251048  0.03141545   0.8944504
V4  0.07532852 0.7219599  0.03232438   0.9219095
V5  0.07603363 0.7265248  0.03220964   0.9408165
V6  0.07697233 0.7254902  0.03293138   0.9659345
V7  0.07861951 0.7299378  0.03293988   0.9939605
V8  0.07851616 0.7271501  0.03299204   0.9924619
V11 0.07790647 0.7316096  0.03298471   1.1678961
V21 0.07861951 0.7299378  0.03293988   0.9939605
V31 0.07488540 0.7229955  0.03249506   0.8995296
V41 0.06982110 0.7236501  0.03264760   0.8408969
V51 0.06457173 0.7149470  0.03252309   0.7497635

$model_list[[1]]$model
                         Sample size: 3978
                    Number of deaths: 305
                     Number of trees: 500
           Forest terminal node size: 45
       Average no. of terminal nodes: 58.024
No. of variables tried at each split: 3
              Total no. of variables: 23
       Resampling used to grow trees: swor
    Resample size used to grow trees: 2514
                            Analysis: RSF
                              Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                         (OOB) CRPS: 0.04245173
   (OOB) Requested performance error: 0.26577737


$model_list[[1]]$vimp10
   cox_predict         bmi_0_         trig_0
   0.055099492    0.031270070    0.013059578
         hyp_0          t2dm_           pc2_
   0.006160379    0.005029568    0.003850401
    wealth_low   baseline_hdl   EduLevel_low
   0.002880856    0.002607893    0.002207414
          pc4_ B_smokstatus_0           age_
   0.001883513    0.001560937    0.001526236
  exercise_vig        B_dep_0       stroke_0
   0.001193894    0.001184870    0.001144235

$model_list[[1]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                    coef exp(coef)  se(coef)      z
sz20_          -0.028345  0.972053  0.055940 -0.507
pc1_            0.019860  1.020059  0.057520  0.345
pc2_           -0.091728  0.912353  0.057566 -1.593
pc3_           -0.030055  0.970392  0.060139 -0.500
```

```
pc4_           -0.039656  0.961120  0.058256 -0.681
age_            0.146215  1.157445  0.071871  2.034
sex            -0.369663  0.690967  0.131234 -2.817
bmi_0_          0.469355  1.598963  0.055908  8.395
hyp_0           0.561660  1.753581  0.120607  4.657
cvd_0          -0.028051  0.972339  0.168995 -0.166
B_dep_0         0.379309  1.461275  0.150855  2.514
trig_0          0.102940  1.108425  0.035682  2.885
baseline_hdl   -0.299819  0.740952  0.196535 -1.526
stroke_0        0.330850  1.392151  0.303594  1.090
B_smokstatus_0  0.378792  1.460519  0.151321  2.503
exercise_light  0.007339  1.007366  0.245089  0.030
exercise_vig   -0.343196  0.709499  0.145865 -2.353
EduLevel_low    0.366422  1.442564  0.220926  1.659
EduLevel_med    0.178474  1.195392  0.210247  0.849
wealth_med      0.085987  1.089792  0.159795  0.538
wealth_low      0.230550  1.259292  0.162348  1.420
t2dm_           0.319483  1.376417  0.059816  5.341
                      p
sz20_           0.61237
pc1_            0.72988
pc2_            0.11106
pc3_            0.61724
pc4_            0.49605
age_            0.04191
sex             0.00485
bmi_0_          < 2e-16
hyp_0          3.21e-06
cvd_0           0.86817
B_dep_0         0.01192
trig_0          0.00392
baseline_hdl    0.12713
stroke_0        0.27581
B_smokstatus_0  0.01231
exercise_light  0.97611
exercise_vig    0.01863
EduLevel_low    0.09720
EduLevel_med    0.39595
wealth_med      0.59050
wealth_low      0.15558
t2dm_          9.24e-08

Likelihood ratio test=257.7  on 22 df, p=< 2.2e-16
n= 3978, number of events= 305


$model_list[[2]]
$model_list[[2]]$beststats
   mtry nodesize nodedepth time    AUCROC        BS
V1    3       50        50    9 0.7767595 0.0720635
    BS_scaled   C_score Calib_alpha Calib_slope
V1 0.09193106 0.7486539   0.0366775    1.309322

$model_list[[2]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1     5       15        50    9 0.7666048 0.07197338
```

```
V2     5       20        50    9 0.7658599 0.07206223
V3     5       25        50    9 0.7683058 0.07192512
V4     5       30        50    9 0.7684578 0.07202294
V5     5       35        50    9 0.7709751 0.07206851
V6     5       40        50    9 0.7704632 0.07188872
V7     5       45        50    9 0.7699974 0.07217563
V8     5       50        50    9 0.7725238 0.07190498
V11    3       50        50    9 0.7767595 0.07206350
V21    5       50        50    9 0.7725238 0.07190498
V31    7       50        50    9 0.7663602 0.07230879
V41   10       50        50    9 0.7655590 0.07240470
V51   15       50        50    9 0.7641787 0.07271219
     BS_scaled   C_score Calib_alpha Calib_slope
V1  0.09306665 0.7393086  0.03336042   0.8960075
V2  0.09194706 0.7408122  0.03378378   0.9455051
V3  0.09367479 0.7411962  0.03407004   0.9853218
V4  0.09244218 0.7411933  0.03564269   1.0008196
V5  0.09186795 0.7436081  0.03528510   1.0327777
V6  0.09413353 0.7427847  0.03601608   1.0560081
V7  0.09051817 0.7450198  0.03575406   1.0713556
V8  0.09392858 0.7464099  0.03617636   1.0998736
V11 0.09193106 0.7486539  0.03667750   1.3093224
V21 0.09392858 0.7464099  0.03617636   1.0998736
V31 0.08884024 0.7395762  0.03551413   0.9842565
V41 0.08763164 0.7375523  0.03555250   0.9118257
V51 0.08375706 0.7355275  0.03591630   0.8406177

$model_list[[2]]$model
                         Sample size: 3979
                     Number of deaths: 336
                      Number of trees: 500
           Forest terminal node size: 50
       Average no. of terminal nodes: 52.166
No. of variables tried at each split: 3
              Total no. of variables: 23
       Resampling used to grow trees: swor
    Resample size used to grow trees: 2515
                             Analysis: RSF
                               Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                           (OOB) CRPS: 0.04534768
   (OOB) Requested performance error: 0.25188944


$model_list[[2]]$vimp10
   cox_predict         bmi_0_         trig_0
  0.0636945012   0.0263490307   0.0171049348
         hyp_0           age_          t2dm_
  0.0088539436   0.0063876320   0.0056133904
  baseline_hdl       stroke_0     wealth_low
  0.0046737470   0.0018113059   0.0017268723
          pc1_   EduLevel_low           pc4_
  0.0016132373   0.0015386472   0.0012770149
  exercise_vig           pc2_ exercise_light
  0.0011574051   0.0008165957   0.0006028412
```

```
$model_list[[2]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                   coef exp(coef) se(coef)      z
sz20_           0.05845   1.06019  0.05648  1.035
pc1_            0.02100   1.02122  0.05473  0.384
pc2_           -0.01324   0.98684  0.05520 -0.240
pc3_            0.03302   1.03358  0.05921  0.558
pc4_            0.03321   1.03376  0.05220  0.636
age_            0.24278   1.27478  0.06807  3.566
sex            -0.27238   0.76156  0.12402 -2.196
bmi_0_          0.47992   1.61595  0.04985  9.628
hyp_0           0.56260   1.75523  0.11439  4.918
cvd_0          -0.04503   0.95597  0.15915 -0.283
B_dep_0         0.32119   1.37877  0.14796  2.171
trig_0          0.10805   1.11410  0.03174  3.404
baseline_hdl   -0.60433   0.54644  0.19215 -3.145
stroke_0        0.56474   1.75900  0.25133  2.247
B_smokstatus_0  0.36836   1.44537  0.14856  2.480
exercise_light  0.02430   1.02459  0.23387  0.104
exercise_vig   -0.21397   0.80738  0.13363 -1.601
EduLevel_low    0.42431   1.52853  0.22276  1.905
EduLevel_med    0.34439   1.41113  0.21110  1.631
wealth_med     -0.00133   0.99867  0.14948 -0.009
wealth_low      0.17185   1.18750  0.15294  1.124
t2dm_           0.31351   1.36823  0.05610  5.589
                      p
sz20_          0.300705
pc1_           0.701234
pc2_           0.810360
pc3_           0.577049
pc4_           0.524662
age_           0.000362
sex            0.028072
bmi_0_          < 2e-16
hyp_0          8.73e-07
cvd_0          0.777231
B_dep_0        0.029943
trig_0         0.000663
baseline_hdl   0.001661
stroke_0       0.024641
B_smokstatus_0 0.013154
exercise_light 0.917258
exercise_vig   0.109347
EduLevel_low   0.056809
EduLevel_med   0.102802
wealth_med     0.992903
wealth_low     0.261150
t2dm_          2.28e-08

Likelihood ratio test=310.2  on 22 df, p=< 2.2e-16
n= 3979, number of events= 336
```

```
$model_list[[3]]
$model_list[[3]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V2    5       20        50    9 0.7410925 0.07398025
    BS_scaled  C_score Calib_alpha Calib_slope
V2 0.06943004 0.713259  0.03310082   0.8649443

$model_list[[3]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1     5       15        50    9 0.7366901 0.07429957
V2     5       20        50    9 0.7410925 0.07398025
V3     5       25        50    9 0.7351740 0.07433120
V4     5       30        50    9 0.7342701 0.07434394
V5     5       35        50    9 0.7371426 0.07407061
V6     5       40        50    9 0.7373241 0.07401269
V7     5       45        50    9 0.7355204 0.07417644
V8     5       50        50    9 0.7362332 0.07414613
V11    3       20        50    9 0.7360386 0.07394098
V21    5       20        50    9 0.7410925 0.07398025
V31    7       20        50    9 0.7280921 0.07482518
V41   10       20        50    9 0.7261060 0.07517868
V51   15       20        50    9 0.7230698 0.07613958
     BS_scaled   C_score Calib_alpha Calib_slope
V1  0.06541346 0.7104335  0.03245704   0.8315036
V2  0.06943004 0.7132590  0.03310082   0.8649443
V3  0.06501550 0.7086842  0.03378657   0.8635963
V4  0.06485528 0.7086244  0.03429826   0.8854501
V5  0.06829343 0.7103689  0.03465459   0.9081545
V6  0.06902200 0.7130094  0.03513163   0.9337602
V7  0.06696220 0.7101800  0.03543247   0.9356905
V8  0.06734341 0.7103670  0.03576064   0.9589342
V11 0.06992399 0.7096533  0.03384507   0.9621704
V21 0.06943004 0.7132590  0.03310082   0.8649443
V31 0.05880193 0.7028878  0.03256426   0.7591329
V41 0.05435536 0.7011319  0.03224581   0.7023665
V51 0.04226864 0.6961665  0.03079759   0.6366920

$model_list[[3]]$model
                         Sample size: 3979
                    Number of deaths: 345
                     Number of trees: 500
          Forest terminal node size: 20
       Average no. of terminal nodes: 107.028
No. of variables tried at each split: 5
              Total no. of variables: 23
       Resampling used to grow trees: swor
    Resample size used to grow trees: 2515
                            Analysis: RSF
                              Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.04793599
   (OOB) Requested performance error: 0.29073272
```

```
$model_list[[3]]$vimp10
   cox_predict         bmi_0_         trig_0
  0.0633668227   0.0279939763   0.0125464277
          age_          t2dm_           pc2_
  0.0078066203   0.0037898724   0.0026323100
        hyp_0           pc4_   EduLevel_low
  0.0024262964   0.0014300096   0.0010435545
       B_dep_0   exercise_vig       stroke_0
  0.0008701044   0.0008426293   0.0007508497
exercise_light          sz20_          cvd_0
  0.0004282929   0.0002451051   0.0001607238

$model_list[[3]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

                    coef exp(coef)  se(coef)      z
sz20_           0.065039  1.067201  0.055387  1.174
pc1_            0.032623  1.033161  0.054653  0.597
pc2_           -0.089742  0.914167  0.053964 -1.663
pc3_            0.016462  1.016598  0.054782  0.300
pc4_           -0.042009  0.958861  0.053633 -0.783
age_            0.142562  1.153225  0.068243  2.089
sex            -0.315088  0.729725  0.124046 -2.540
bmi_0_          0.486410  1.626467  0.048283 10.074
hyp_0           0.424529  1.528870  0.112253  3.782
cvd_0          -0.006792  0.993231  0.161008 -0.042
B_dep_0         0.234661  1.264480  0.148876  1.576
trig_0          0.117352  1.124515  0.040517  2.896
baseline_hdl   -0.433751  0.648073  0.189558 -2.288
stroke_0        0.472919  1.604671  0.282917  1.672
B_smokstatus_0  0.387311  1.473014  0.141095  2.745
exercise_light -0.129075  0.878908  0.240804 -0.536
exercise_vig   -0.186913  0.829516  0.129140 -1.447
EduLevel_low    0.391485  1.479176  0.214204  1.828
EduLevel_med    0.333466  1.395797  0.201573  1.654
wealth_med     -0.038004  0.962709  0.143598 -0.265
wealth_low      0.097583  1.102503  0.148756  0.656
t2dm_           0.294049  1.341849  0.056125  5.239
                      p
sz20_          0.240290
pc1_           0.550563
pc2_           0.096314
pc3_           0.763797
pc4_           0.433472
age_           0.036703
sex            0.011082
bmi_0_          < 2e-16
hyp_0          0.000156
cvd_0          0.966352
B_dep_0        0.114976
trig_0         0.003775
baseline_hdl   0.022124
```

```
stroke_0       0.094607
B_smokstatus_0 0.006051
exercise_light 0.591947
exercise_vig   0.147794
EduLevel_low   0.067605
EduLevel_med   0.098063
wealth_med     0.791276
wealth_low     0.511828
t2dm_          1.61e-07

Likelihood ratio test=257.4  on 22 df, p=< 2.2e-16
n= 3979, number of events= 345



$time
Time difference of 4.668681 mins
```