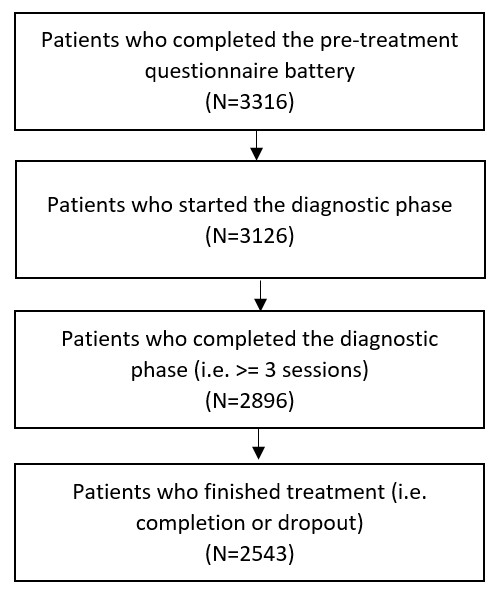
**Supplemental materials 1**

Flowchart of selected patients for the training sample

of this study.



**Supplemental materials 2**

Patient characteristics of the training and holdout samples divided into dropout and regular completion.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristics** | **Training sample**  **(*N* = 2043)** | | ***t*-value / chi²-value** | ***p*-value** | **Holdout sample**  **(*N* = 500)** | | ***t*-value / chi²-value** | ***p*-value** |
| **Regular**  **(*N* = 1418)** | **Dropout (*N* = 625)** | **Regular**  **(*N* = 346)** | **Dropout (*N* = 154)** |
| Age (*M* years (*SD*)) | 36.7 (12.7) | 34.4 (12.7) | 3.75 | <.001 | 36.4 (14.3) | 34.2 (13.0) | 1.70 | 0.09 |
| Gender female (*n* (%)) | 1279 (62.6) | 400 (64.0) | 0.67 | 0.41 | 213 (61.6) | 84 (55.5) | 1.89 | 0.17 |
| German nationality (*n* (%)) | 1352 (95.3) | 590 (94.4) | 0.64 | 0.43 | 321 (92.8) | 137 (89.0) | 1.55 | 0.21 |
| Marital status (*n* married (%)) | 422 (29.7) | 146 (23.4) | 8.75 | <.01 | 89 (25.7) | 30 (19.5) | 1.96 | 0.16 |
| Education (*n* > 12 years (%)) | 662 (46.7) | 210 (33.6) | 29.83 | <.001 | 179 (51.7) | 53 (34.4) | 12.17 | <.001 |
| Inability to work (*n* (%)) | 269 (19.0) | 149 (24.3) | 6.03 | <.05 | 58 (16.8) | 27 (17.5) | 0.01 | 0.93 |
| Intake of medication (*n* (%)) | 1011 (71.3) | 429 (68.6) | 1.35 | 0.25 | 232 (67.1) | 103 (66.9) | 0.00 | 1 |
| Primary diagnosis (*n* (%))  Affective disorder  Anxiety disorder  Adjustment disorder / PTSD  Other | 667 (51.1)  217 (16.6)  191 (14.6)  343 (24.2) | 287 (49.2)  64 (11.0)  108 (18.5)  166 (26.6) | 0.50  9.72  4.29  1.18 | 0.48  <.01  <.05  0.28 | 170 (49.4)  40 (11.6)  69 (20.1)  67 (19.4) | 59 (38.3)  20 (13.0)  38 (24.7)  37 (24.0) | 4.85  0.08  1.08  1.14 | <.05  0.78  0.30  0.29 |
| Comorbidity (*n* (%))  Two diagnoses  Three or more diagnoses | 424 (30.0)  525 (37.0) | 181 (29.0)  289 (47.7) | 0.14  20.04 | 0.71  <.001 | 126 (36.4)  70 (20.2) | 48 (31.2)  51 (33.1) | 1.07  8.96 | 0.30  <.01 |

Note: Other diagnoses included e.g., obsessive-compulsive disorders, eating disorders, personality disorders, psychosis, and substance use disorders. For continuous variables, a t - test was used, for categorical variables, a chi² - test was used.

Patient characteristics of the training and holdout sample

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristics** | **Training sample**  **(*N* = 2043)** | **Holdout sample**  **(*N* = 500)** | ***t*-value / chi²-value** | ***p*-value** |
| Age (*M* years (*SD*)) | 36.0 (12.8) | 35.7 (14.0) | 0.46 | 0.65 |
| Gender female (*n* (%)) | 1279 (63.8) | 297 (59.4) | 1.62 | 0.20 |
| German nationality (*n* (%)) | 1942 (95.1) | 458 (91.6) | 8.40 | <.01 |
| Marital status (*n* married (%)) | 568 (27.8) | 119 (23.8) | 3.06 | 0.08 |
| Education (*n* > 12 years (%)) | 872 (42.7) | 232 (46.4) | 2.11 | 0.15 |
| Inability to work (*n* (%)) | 418 (20.5) | 85 (17.0) | 2.82 | 0.09 |
| Intake of medication (*n* (%)) | 1440 (70.0) | 335 (66.7) | 2.15 | 0.14 |
| Number of dropouts (*n* (%)) | 625 (30.6) | 154 (30.8) | 0.00 | 0.97 |
| Primary diagnosis (*n* (%))  Affective disorder  Anxiety disorder  Adjustment disorder / PTSD\*  Other | 954 (50.5)  281 (14.9)  299 (15.8)  509 (24.9) | 229 (46.0)  60 (12.0)  107 (21.5)  104 (20.8) | 3.08  2.36  8.51  3.50 | 0.08  0.12  <.01  0.06 |
| Comorbidity (*n* (%))  Two diagnoses\*  Three or more diagnoses\* | 605 (29.6)  823 (40.3) | 174 (34.8)  121 (24.2) | 4.84  43.83 | <.05  <.001 |

Note: Other diagnoses included e.g., obsessive-compulsive disorders, eating disorders, personality disorders, psychosis, and substance use

disorders. For continuous variables, a t - test was used, for categorical variables, a chi² - test was used.

**Supplemental materials 3**

Mean scores of the models generated by each algorithm with all significant predictors from Zimmermann et al. (2017).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overall rank | Algorithm | Brier score | AUC | Training AUC |
| 1 | ADA | .2259 (4) | .6393 (1) | .6456 (1) |
| 2 | GLMAIC | .2344 (5) | .6378 (2) | .6392 (2) |
| 3 | GLMBOOST | .2344 (6) | .6368 (3) | .6381 (7) |
| 3 | GBM | .2204 (2) | .6363 (7) | .6388 (5) |
| 5 | GLM | .2346 (9) | .6366 (4) | .6379 (8) |
| 5 | LDA | .2345 (7) | .6364 (6) | .6378 (10) |
| 7 | XGB | .2173 (1) | .6222 (13) | .6359 (12) |
| 8 | BAYESGLM | .2346 (10) | .6366 (5) | .6379 (9) |
| 9 | EN | .2345 (8) | .6325 (9) | .6389 (4) |
| 10 | SVM | .2350 (11) | .6330 (8) | .6345 (13) |
| 10 | RF | .2247 (3) | .5931 (16) | .6027 (16) |
| 12 | NNET | .2360 (13) | .6314 (11) | .6382 (6) |
| 13 | MONMLP | .2382 (16) | .6320 (10) | .6374 (11) |
| 13 | NB | .2361 (14) | .6239 (12) | .6282 (14) |
| 15 | AVNN | .2373 (15) | .6205 (14) | .6391 (3) |
| 16 | CART | .2352 (12) | .5783 (18) | .5881 (18) |
| 17 | MARS | .2635 (19) | .6033 (15) | .3995 (21) |
| 17 | CTREE | .2452 (17) | .5816 (17) | .6041 (15) |
| 19 | C4.5 | .2649 (20) | .5723 (19) | .5902 (17) |
| 19 | LOGIT | .2527 (18) | .5572 (21) | .5625 (20) |
| 21 | kNN | .2743 (21) | .5663 (20) | .5794 (19) |

Note: The digits in the brackets refer to the rank of the algorithm for the particular parameter. The overall rank is the sum of the single rankings concerning the two parameters without the training AUC. When sums were equal, the AUC was given priority. For the full names of the ML algorithms, see Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overall rank | Algorithm | Brier score | AUC | Training AUC |
| 1 | RF | .2037 (1) | .6584 (1) | .6610 (3) |
| 2 | GBM | .2090 (2) | .6567 (2) | .6637 (1) |
| 3 | ADA | .2093 (4) | .6515 (3) | .6607 (4) |
| 4 | XGB | .2090 (3) | .6442 (5) | .6624 (2) |
| 5 | GLMBOOST | .2306 (6) | .6469 (4) | .6531 (10) |
| 6 | EN | .2324 (8) | .6435 (6) | .6599 (5) |
| 7 | LDA | .2333 (9) | .6425 (7) | .6569 (7) |
| 8 | BAYESGLM | .2338 (10) | .6424 (8) | .6569 (8) |
| 8 | SVM | .2323 (7) | .6401 (11) | .6575 (6) |
| 8 | CART | .2113 (5) | .6256 (13) | .6278 (15) |
| 11 | GLM | .2339 (11) | .6422 (9) | .6568 (9) |
| 12 | GLMAIC | .2340 (12) | .6406 (10) | .6520 (11) |
| 13 | AVNN | .2374 (14) | .6289 (12) | .6489 (12) |
| 14 | NNET | .2459 (17) | .6247 (14) | .6436 (13) |
| 14 | MONMLP | .2381 (15) | .6079 (16) | .6217 (16) |
| 16 | CTREE | .2438 (16) | .5764 (17) | .5856 (17) |
| 16 | LOGIT | .2353 (13) | .5505 (20) | .5608 (19) |
| 18 | NB | .2856 (19) | .6214 (15) | .6325 (14) |
| 19 | kNN | .3052 (21) | .5608 (18) | .5564 (20) |
| 19 | C4.5 | .2986 (20) | .5510 (19) | .5612 (18) |
| 19 | MARS | .2850 (18) | .5275 (21) | .4284 (21) |

Mean scores of the models generated by each algorithm with all significant predictors identified with an elastic net analysis in the training sample.

Note: The digits in the brackets refer to the rank of the algorithm for the particular parameter. The overall rank is the sum of the single rankings concerning the two parameters without the training AUC. When sums were equal, the AUC was given priority. For the full names of the ML algorithms, see Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overall rank | Algorithm | Brier score | AUC | Training AUC |
| 1 | GBM | .2045 (1) | .6483 (2) | .6522 (1) |
| 2 | ADA | .2068 (3) | .6527 (1) | .6503 (2) |
| 3 | RF | .2053 (2) | .6423 (4) | .6443 (6) |
| 4 | GLMBOOST | .2311 (7) | .6426 (3) | .6492 (4) |
| 4 | XGB | .2096 (4) | .6319 (6) | .6489 (5) |
| 6 | EN | .2308 (6) | .6413 (5) | .6494 (3) |
| 7 | SVM | .2335 (8) | .6301 (7) | .6296 (8) |
| 8 | GLMAIC | .2375 (9) | .6291 (9) | .6287 (10) |
| 9 | CART | .2167 (5) | .6043 (14) | .6078 (14) |
| 10 | BAYESGLM | .2382 (12) | .6293 (8) | .6303 (7) |
| 11 | AVNN | .2376 (11) | .6241 (11) | .6296 (9) |
| 12 | LDA | .2389 (13) | .6243 (10) | .6265 (11) |
| 13 | MONMLP | .2376 (10) | .6008 (15) | .5965 (15) |
| 14 | GLM | .2396 (14) | .6234 (12) | .6259 (12) |
| 15 | NNET | .2498 (17) | .6226 (13) | .6131 (13) |
| 16 | CTREE | .2468 (16) | .5796 (17) | .5844 (17) |
| 17 | LOGIT | .2419 (15) | .5373 (19) | .5560 (18) |
| 18 | NB | .3680 (21) | .5957 (16) | .5962 (16) |
| 19 | C4.5 | .3431 (20) | .5520 (18) | .5549 (19) |
| 20 | kNN | .3181 (19) | .5351 (20) | .5293 (20) |
| 20 | MARS | .2843 (18) | .5198 (21) | .4298 (21) |

Mean scores of the models generated by each algorithm with all variables.

Note: The digits in the brackets refer to the rank of the algorithm for the particular parameter. The overall rank is the sum of the single rankings concerning the two parameters without the training AUC. When sums were equal, the AUC was given priority. For the full names of the ML algorithms, see Table 2.

**Supplemental materials 4**

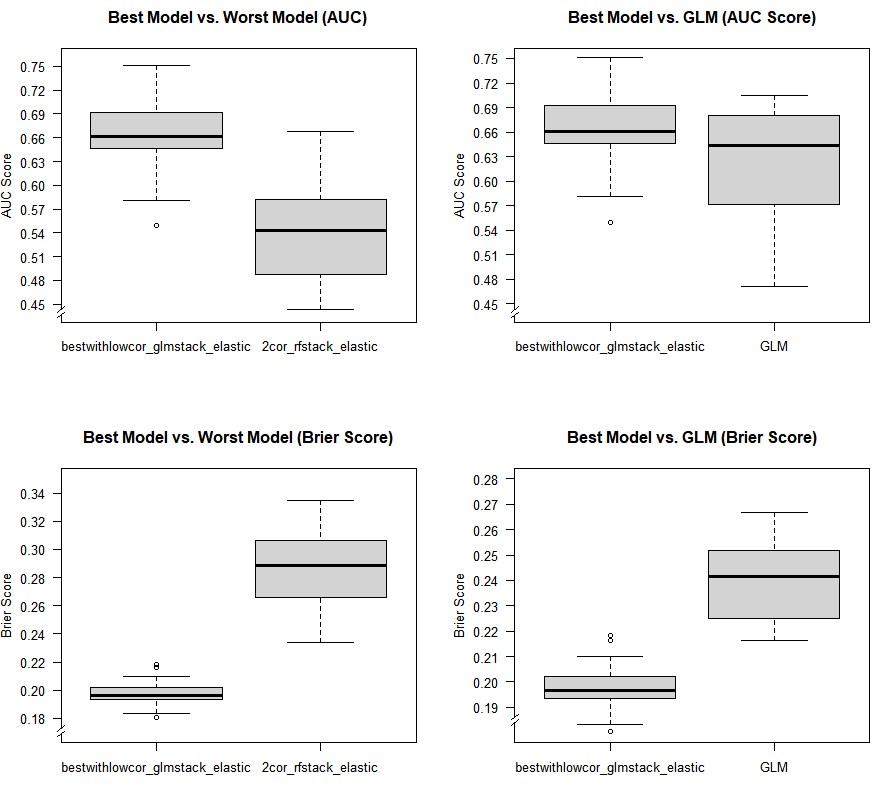
Overview of ensemble correlations by included predictors.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Two least correlating** | **Three least correlating** | **Best algorithm with least correlating** |
| **All variables** | CART + C4.5 | kNN + C4.5 + GBM | GBM + kNN |
| **Significant variables** | MARS + C4.5 | MARS + kNN + C4.5 | ADA + LOGIT |
| **Elastic net variables** | LOGIT + MARS | LOGIT + MARS + C4.5 | RF + kNN |

Note: Significant variables = using significant variables from Zimmermann et al. (2017); Elastic net variables = using variables with predictive power in a preceding elastic net analysis.

kNN = K-fold Nearest Neighbors; C4.5 = C4.5-like Trees; MARS = Bagged Multivariate Adaptive Regression Splines; LOGIT = Boosted Logistic Regression; CART = Bagged Classification and Regression Tree; ADA = Boosted Classification Tree; RF = Random Forest; GBM = Stochastic Gradient Boosting.

**Supplemental materials 5**

Boxplots for the best models (overall), the worst models (overall) and the GLM-models.

Note: GLM = Generalized Linear Model; bestwithlowcor\_glmstack\_elastic = ensemble of the best algorithm with its least correlating algorithm (i.e., random forest, nearest neighbor) using variables with predictive power in a preceding elastic net analysis stacked via GLM; 2cor\_rfstack\_elastic = ensemble of the two least correlating algorithms (i.e., boosted logistic regression, bagged multivariate adaptive regression splines) using variables with predictive power in a preceding elastic net analysis stacked via GLM; AUC = area under the curve.

**Supplemental materials 6**

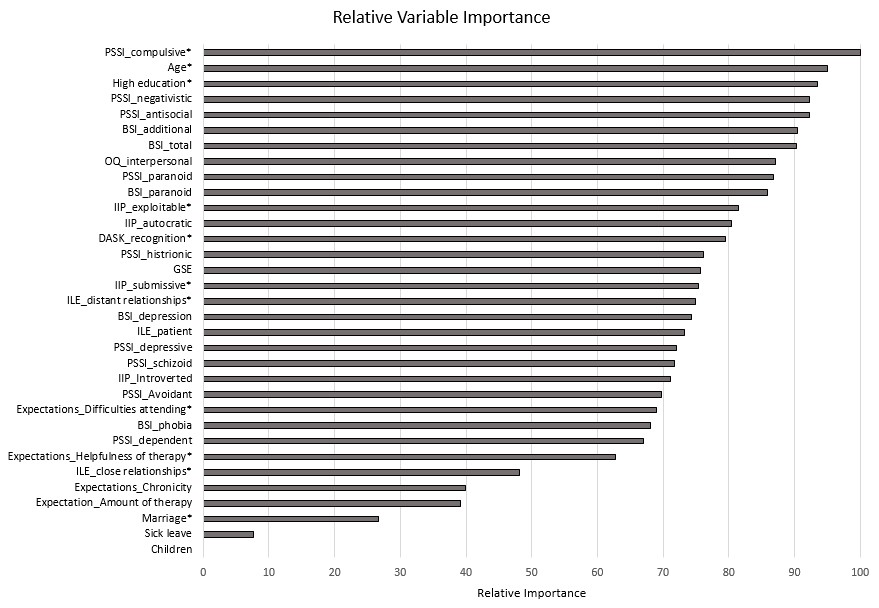
Confusion matrix of the best model generated by the best ensemble

(i.e., best algorithm with its least correlating algorithm (RF, kNN) stacked via GLM using variables with predictive power in a preceding elastic net analysis)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Observed** | |  |
| **Predicted** | Regular | Dropout | Total |
| Regular | 224 | 61 | 285 |
| Dropout | 122 | 93 | 215 |
| Total | 346 | 154 | 500 |

Note: kNN = K-fold Nearest Neighbors; RF = Random Forest; GLM = Generalized Linear Model.

**Supplemental materials 7**

Relative variable importance for the final model tested in the holdout sample.

Note: All variables not shown here were excluded from the ensemble by the preceding elastic net analysis (i.e., relative importance = 0). Since the Nearest Neighbor algorithm uses the entire data set and uses Euclidean distance to determine predictions, no variable importance is available. Therefore, only the values from the Random Forest algorithm are shown here.

\* The correlation between these variables and dropout is negative.