Grant Agreement no. 777167

BOUNCE

Predicting Effective Adaptation to Breast Cancer to Help Women to BOUNCE Back

Research and Innovation Action

SC1-PM-17-2017: *Personalised computer models and in-silico systems for well-being*

Deliverable: D4.3b - Final Design and Implementation of the Preliminary In Silico Resilience Trajectory Predictor

Due date of deliverable: (31-03-2022)

Actual submission date: (28-04-2022)

Start date of Project: 01 November 2017 Duration: 54 months (including a 6 month extension)

Responsible WP4

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777167

Dissemination level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0. Document Info

0.1. Authors

Author	Company	E-mail
Part 1		
Panagiotis Simos	FORTH	akis.simos@gmail.com
Georgios Manikis	FORTH	gmanikis@gmail.com
Vangelis Karademas	FORTH	karademas@uoc.gr
Konstantina Kourou	FORTH	konstadina.kourou@gmail.com
Eugenia Mylona	FORTH	mylona.eugenia@gmail.com
Part 2		
Eleni Kolokotroni	ICCS	ekolok@mail.ntua.gr
Georgios Stamatakos	ICCS	gestam@central.ntua.gr

0.2. Document's history

Document version #	Date	Change
V0.1	07-01-2022	Starting version, template
V0.2	25-03-2022	First complete draft of Part 1
V0.3	01-04-2022	Integrated version of Part 1 (send to WP members)
V0.4	12-04-2022	Received Part 2 from ICCS
V0.5	20-04-2022	Updated version (send to project internal reviewers)
Sign off	28-04-2022	Signed off version (for approval to PMT members)
V1.0	28-04-2022	Approved Version to be submitted to EU

0.3. Document data

Keywords	
Editor Address data	Name: Panagiotis Simos Partner: FORTH Address: N. Plastira 100. GR-70013 Vassilika Vouton Phone: Fax: E-mail: akis.simos@uoc.gr
29/	29/04/2022

Contents

NOTE: The character F before a number refers to work done by FORTH whereas the character I before a number refers to work done by ICCS.

Part 1 (FORTH)

1.1. Introduction

This deliverable presents the results of the supervised and unsupervised learning analysis pipelines implemented for modelling resilience as ***outcome*** and resilience as ***process*** over the first 18 months post-diagnosis.

With respect to ***Resilience as Outcome***, we developed and tested generalizable models for optimal prediction of 12- and 18-month patient outcomes (in terms of symptoms of mental health and overall quality of life (QoL)) by aggregating all available patient information from the early phase of illness (i.e., M0 and M3 measurement waves). The flexible and comprehensive BOUNCE Machine Learning (ML) framework was applied to predict patients' resilience based on the available set of variables. Potential predictors included: (i) patient-reported outcomes (i.e., mental health, distress level, health- and global Quality of Life (QoL), and functionality), (ii) sociodemographic variables (i.e., education level and employment status) and perceived social or health-related support, (iii) potentially stressful events taking place during the follow up period (including perceived side-effects), (iv) psychological characteristics and coping reactions (i.e. perceptions of illness, optimism, emotional self-regulation strategies. etc.), (v) lifestyle factors (i.e., diet and exercise), (vi) clinical variables (cancer stage, molecular tumor type, type and timing of medical treatments), and (vii) biological indicators of systemic processes (e.g., anemia, creatinine and bilirubin, blood cell counts etc.). Self-reported symptoms of anxiety and depression as well as subjective ratings of QoL registered near the time of BC diagnosis were included as predictors in order to optimize model performance.

Supervised modelling work described in the first version of D4.3 (D4.3.a) focussed on resilience as outcome as indicated by endpoint indices of overall mental health and global QoL. Specifically, we had implemented models predicting M12 HADS total scores and EORTC global QoL rating regardless of the corresponding patient scores at the time of diagnosis. Here, we expand these analyses in three ways (this work is presented in **Section 1.2** of the present document).

Firstly, by addressing an additional, clinically challenging question, namely identifying patients who display stable-good mental health (or QoL) between M0 and M12/M18 and patients who display good mental health (or QoL) at M0 and clinically significant deterioration in mental health (or QoL, respectively) at M12/M18. In this manner, patient progress on two well-established, yet complementary, well-being indicators is considered as representing two characteristic resilience trajectories.

Secondly, by implementing models predicting overall mental health and global QoL status at M18 given that this endpoint information was not available when preparing D4.3a.

Thirdly, by implementing additional models utilizing identical endpoints (i.e., M12 or M18, overall mental health or global QoL status (or decline as compared to the time of diagnosis, respectively), which take into account lifestyle and psychological characteristics as experienced by the patient while cancer treatments are underway (i.e., at M6 of the longitudinal study).

Within the framework of assessing resilience as outcome we pursued a second aim, namely to identify key modifiable factors that determine patients' well-being outcomes. Along these lines we implemented additional supervised ML models which considered all available patient information listed in the previous paragraph except mental health and QoL ratings obtained during the first three months post-diagnosis. As a result, we could improve model sensitivity by selecting clinically important patient traits like optimism, coping strategies, self-regulation strategies, and relevant clinical characteristics. To further this aim, we conducted Local Interpretation analysis for randomly selected patients included in each outcome label (i.e. stable good vs deteriorated mental health at M12) to explore the patient characteristics that may be related to model successes (correctly classified patients to each class) to refine clinical recommendations toward improving patients' psychological status. These results are presented in Section 1.3 of the present document.

In order to address ***Resilience as Process***, we developed and tested unsupervised clustering schemes to identify subgroups of patients who display distinct profiles of change in mental health symptoms (or global QoL) over the first 18 months post diagnosis. Subsequently we employed feature selection in the context of supervised classification models in order to identify the most important variables, collected during the first 3 months post-diagnosis, that uniquely contribute to each distinct trajectory profile. Section 1.4 presents these results.

1.2. Resilience as Outcome: Supervised Models predicting deterioration of Mental Health and QoL at M12 and M18

In this section we apply supervised ML models to address a specific and particularly challenging clinical problem, namely the prediction of psychological resilience among patients who did not report significant mental health-related symptoms at the time of diagnosis and are thus less likely to be systematically monitored for signs of mental health deterioration during the course of cancer treatment and thereafter. To address this goal all available variables collected at M0 and M3—including global QoL, anxiety and depression symptoms at the time of diagnosis and shortly after (M3)—were included in the models as potentials predictors. Supplementary models with identical predicted endpoints, utilizing available data from M6 alone, or M0 and M6, combined, were also implemented. These models addressed two possible clinical scenarios, namely (i) that a patient's psychological and lifestyle characteristics are not recorded at the time of diagnosis but while cancer treatments are well underway, and (ii) that the full set of baseline and 6-month follow up data are available but the patient did not provide psychological measurements on month 3.

1.2.1 Dataset description

1.2.1.1. Participants included in the analyses

Of the total cohort of 706 women enrolled at M0 539 (76.3%) and 495 (70.1%) were followed up to twelve (M12) or 18 months (M18).

M12 prediction models-Mental Health. A total of 376 women met the criteria for inclusion into one of the two groups of interest (Stable-Good and Deteriorated Mental Health between M0 and M12): 326 maintained low HADS scores throughout the first year after diagnosis (Stable-Good Mental Health group), while the remaining 50 patients had clinically significant symptomatology at M12 (Deteriorated Mental Health group). Cases who missed two or more measurement waves and patients who displayed substantial fluctuation across the five available measurement waves (e.g., low-high-low-low-high scores or low-low-high-low-high) were not included in any group (n=46; unclassified cases). Among the remaining patients, 60 reported reduced symptoms between M0 and M12, 55 women displayed significant symptoms at both M0 and M12.

M12 prediction models-QoL. A total of 270 women met the criteria for inclusion into one of the two groups of interest (Stable-Good and Deteriorated QoL between M0 and M12): 216 maintained high EORTC scores throughout the first year after diagnosis (Stable-Good QoL group), while the remaining 54 patients reported poor QoL at M12 (Deteriorated QoL group). Among the remaining patients, 84 reported improved QoL between M0 and M18, 44 patients reported stable poor QoL, and 51 were unclassified.

M18 prediction models-Mental Health. A total of 311 women met the criteria for inclusion into one of the two groups of interest (Stable-Good and Deteriorated Mental Health between M0 and M18): 268 maintained low HADS scores throughout the first 18 months after diagnosis

(Stable-Good Mental Health group), while the remaining 43 patients had clinically significant symptomatology at M18 (Deteriorated Mental Health group). Among the remaining patients, 62 reported reduced symptoms between M0 and M18, 56 women displayed significant symptoms at both M0 and M18, and 14 were unclassified.

M18 prediction models-QoL. A total of 241 women met the criteria for inclusion into one of the two groups of interest (Stable-Good and Deteriorated QoL between M0 and M18): 182 maintained high EORTC scores throughout the first 18 months diagnosis (Stable-Good QoL group), while the remaining 59 patients rated their QoL as significantly poorer at M18 (Deteriorated QoL group). Among the remaining patients, 84 reported improved QoL between M0 and M18, 29 women displayed stable poor QoL at both M0 and M18, and 24 were unclassified.

1.2.1.2. Grouping variables

Mental Health prediction models. Self-reported mental health status at either 12 or 18 months post-diagnosis, indexed by the total score on the 14-item Hospital Anxiety and Depression Scale (HADS), served as the outcome variable in these models. Higher scores indicate more frequent psychological symptoms. The clinically validated cut-off score of 16/42 points in a wide range of languages was used to identify patients who reported potentially clinically significant symptoms at M0 and at M12[1,2]. For each model, patients were assigned to two classes: (a) those who reported non-clinically significant symptoms of anxiety and depression at M0 (i.e., immediately following BC diagnosis) and clinically significant symptomatology at M12 or M18 according to validated cut-offs on HADS total score (Deteriorated Mental Health group), and (b) those who reported mild symptomatology throughout the first 12 or 18 months post diagnosis (Stable-Good Mental Health group).

QoL prediction models. Self-rated, overall quality of life at either 12 or 18 months post-diagnosis, was assessed using the two questions from The Global Health Status scale from the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30[3] questionnaire. Higher scores indicate better overall QoL. In the absence of a clinically validated cut-off score we used the 25th percentile of the total sample distribution of scores at M0 to identify patients who rated their QoL as relatively poor (corresponding to a score of 75 points). For each model, patients were assigned to two classes: (a) those who reported EORTC>75 points at M0 and poor QoL at M12 or M18 (as indicated by scores ≤75 points) (Deteriorated QoL group), and (b) those who reported relatively good QoL (>75 points) throughout the first 12 or 18 months post diagnosis (Stable-Good QoL group).

[1] Wu Y, Levis B, Sun Y, He C, Krishnan A, Neupane D, Bhandari PM, Negeri Z, Benedetti A, Thombs BD; DEPRESsion Screening Data (DEPRESSD) HADS Group. Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: systematic review and individual participant data meta-analysis. BMJ. 2021 May 10;373:n972. doi: 10.1136/bmj.n972.

[2] Vodermaier A, Millman RD. Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. Support Care Cancer. 2011 Dec;19(12):1899-908. doi:10.1007/s00520-011-1251-4.

[3] Aaronson, N.K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N.J., Filiberti, A., … & Takeda, F. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, *85,* 365–376. https://doi.org/10.1093/jnci/85.5.365

1.2.1.3. Predictor variables considered in the models

Sociodemographic*.* The following variables were registered at baseline: Age (in years), education level (categorized as low [0-9 years], and high [>9 years]), relational status (alone, or with partner), children (yes, no), employment status (currently employed or not), type of employment (full-time, retired, or self-employed vs. unemployed, housewife, or part-time employment), monthly income (very low vs average/high; adjusted for the Gross Domestic Product of home country of each participant). Two additional variables were aggregated over the first 3 months post diagnosis: sick leave taken (in days), and significant life stressors (other than BC) during the first three months post diagnosis (categorized as none/single event vs two or more events).

Life-style*.* The following variables were registered at baseline: Current smoker, alcohol consumption (no drinking or occasional consumption, defined as: ≤2 servings of beer and/or ≤1 servings of spirits per week, moderate, defined as: 3-6 servings of beer and/or ≤4 servings of spirits per week, heavy, defined as: >6 servings of beer and/or >4 servings of spirits per week), self-defined diet (Mediterranean, special diet [e.g., vegan, lactose-free), undefined), physical exercise (defined as: no/occasional [<60 min/week], moderate [60-180 min/week], heavy [>180 min/week]).

Medical*.* Health-related variables collected at baseline: Eastern Cooperative Oncology Group (ECOG) performance status, obesity, family history of BC, pre-existing chronic physical illness (other than metabolic), psychotropic medications (including sleep medications), pre-existing metabolic disease, pre-existing anxiety or dysthymia, anemia, menopausal status (premenopausal, perimenopausal, postmenopausal), serum levels of alanine aminotransferase, creatinine, and bilirubin, blood cell count (thrombocyte count, neutrophil/leukocyte ratio [NLR]).

Breast cancer-related*:* cancer stage (I vs II or III), tumor molecular profile (Luminal A, Luminal B, Triple Negative, HER2 Enriched) progesterone receptor positivity, estrogen receptor positivity, HER2 positivity, Ki67 levels (≥25); treatment-related: surgery at M0, surgery at M3, onset of chemotherapy at M0, onset of chemotherapy at M3, onset of radiotherapy at M0, onset of radiotherapy at M3, type of breast surgery (lumpectomy vs mastectomy), type of chemotherapy (adjuvant or neoadjuvant), type of endocrine therapy (letrozole, exemestane, anastrozole, ovarian suppression, tamoxifen), anti-HER2 therapy, systematic mental health support through M3.

Finally, patient *psychosocial* characteristics were assessed using standardized questionnaires that had been appropriately adapted and translated into the different languages of the four clinical sites of the BOUNCE prospective study. The following domains were assessed: (i) several personality characteristics, (ii) coping and the ability to cope, (iii) perceived social support, (iv) resilience as trait, (v) illness perception and related behaviors, (vi) global QoL, anxiety and depression symptoms and, (vii) patient affect at the time of measurement. Measures included the following:

Positive and Negative affect. The Positive and Negative Affectivity Schedule (PANAS) [1] was used to evaluate positive (10 adjectives; Cronbach's $\alpha = 0.84$) and negative affect (10 adjectives; Cronbach's $\alpha = 0.75$). A 5-point Likert type scale was adopted to assess affect over the past week. Higher scores represent higher levels of positive and negative affect, respectively.

Fear of Cancer Recurrence Inventory. The 9-item Fear of Cancer Recurrence Inventory (FCRI) questionnaire was used to measure the fear of a recurrence event [2]. Each item of the questionnaire is rated on a Likert type scale ranging from 0 ("not at all" or "never") to 4 ("a great deal" or "all the time"). The total score can be obtained by summing the responses to all items. Higher scores indicate higher levels of FCR.

Quality of Life. To evaluate patients' global health status, the BR-23 module of the European Organization for Research and Treatment of Cancer (EORTC) QLQ questionnaire was used [3]. This module comprises of 23 questions related to the (i) disease symptoms, (ii) side effects of treatment (surgery, chemotherapy, radiotherapy and hormonal treatment), (iii) body image, (iv) sexual functioning and (v) future perspective. It should be noted that a linear transformation was applied to the raw scores to reach a range from 0 to 100.

Illness perception and coping responses. The brief version of the Cancer Behavior Inventory (CBI-B) measure [4] was used to assess a general sense of perceived self-efficacy to cope with the illness-related difficulties and needs. A single score measure of coping self-efficacy was yielded (Cronbach's $\alpha = 0.89$) with higher scores indicating higher confidence in coping with illness. The Mental Adjustment to Cancer scale (MAC) [5] was used to estimate patients' coping responses to cancer. The scale includes five reliable dimensions: (i) fighting spirit, (ii) helplessness, (iii) anxious preoccupation, and (iv) avoidance. A 4-point Likert type scale indicate the coping responses of BC patients. Also, the Perceived Ability to Cope with Trauma (PACT) questionnaire was used to estimate the flexibility in coping across different potentially traumatic events [6]. Two scales are measured related to: (i) the focus on processing the trauma (trauma focus), and (ii) the focus on moving beyond the trauma (forward focus). An overall PACT flexibility score was created to evaluate both types of coping. Finally, to assess any potential positive responses to the entire stressful experience, we used the total score on the 14-Post-Traumatic Growth Inventory (PTGI short form) questionnaire (with higher scores indicating better posttraumatic growth) [7].

Social support and family resilience. The modified Medical Outcomes Study Social Support Survey (mMOS-SS) was used to assess social support, which has been shown to provide many benefits related to overall health and well-being [8]. It consists of 8 items and the total score was calculated by summing all response values (Cronbach's $\alpha = 0.92$). Higher total and subscale mMOS-SS scores reflect stronger social support. For the assessment of family resilience the Walsh Family Resilience Questionnaire [9] was used. For the purposes of the BOUNCE study, two subscales were used: (i) communication and cohesion and (ii) perceived family coping. A higher total score indicates higher levels of family resilience.

Resilience as a personality characteristic (trait). The Connor-Davidson Resilience Scale was used to assess resilience as a trait (CD-RISC) [10]. The scale includes 10 items for quantifying the level of self-perceived resilience (e.g. ability to adapt to change; achieving my goals). Each item is rated on a 5-point Likert type scale from 0 (“not true at all”) to 4 (“true nearly all the time”) with higher total scores reflecting higher resilience levels (Cronbach’s $\alpha = 0.89$).

Emotion regulation and relevant strategies. The Cognitive Emotion Regulation Questionnaire (CERQ – short) was used to identify the cognitive emotion regulation strategies (or cognitive coping strategies) that BC patients followed when experiencing negative events or situations [11]. A 5-item Likert type scale was used for each item ranging from 1 (“(almost) never”) to 5 (“(almost) always”). In addition, the Mindful Attention Awareness Scale (MAAS) [12] was used to assess the patients’ characteristic of mindfulness. A total score is considered by summing all patients’ responses with higher scores reflecting higher levels of dispositional mindfulness.

Other personality characteristics. Sense of coherence was assessed based on the Sense of Coherence (SOC)-13 questionnaire (Cronbach’s $\alpha = 0.81$, for the total score). Comprehensibility (5 items), manageability (4 items), and meaningfulness (4 items) were measured on a 7-point (Likert-type) response scale (from 1 (lower) to 7 (higher)) with higher total scores indicating higher level of sense of coherence. Generalized optimism was assessed with the Life Orientation Test (LOT)–Revised (Cronbach’s $\alpha = 0.71$) [13].

1.2.2. Model Design

1.2.2.1. Supervised learning analysis pipeline

Figure F1 illustrates the pipeline adopted for the supervised learning analysis towards the design and development of robust and generalizable predictive models to minimize training errors while considering the bias-variance tradeoff. These steps are described in more detail below.

Data pre-processing and handling of missing data

Initially, raw data were rescaled to zero mean and unit variance and ordinal variables were recoded into dummy binary variables. Cases and variables with more than 10% of missingness were excluded from the final dataset. Remaining missing values were replaced by the global median value.

Feature Selection

Feature selection was conducted using a meta-transformer built on a Random Forest (RF) algorithm[4] which assigns weights to the features and ranks them according to their relative importance. The maximum number of features to be selected by the estimator was set to the default value (i.e. the square root of the total number of features) in order to identify all important variables that contribute to the risk prediction of mental health and QoL

[4] Zhou, Qifeng, Hao Zhou, and Tao Li. "Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features." Knowledge-based systems 95 (2016): 1-11.

deterioration. The feature selection scheme was incorporated into the ML-based pipeline alongside the classification algorithm to select only the relevant features for training and testing the final model.

For comparison reasons, we also set the parameter of maximum features to be selected by the estimator equal to 15 and 20, respectively, to keep the most significant and highly ranked variables for building the classification models. The obtained results were comparable with the training and testing results based on the initial default settings, suggesting a robust and generalizable ML-based prediction models for the identification of clinically important risk factors.

Model training and validation

Cross-validation: evaluating model performance. To address the rather common problem of model overfitting in machine learning applications in clinical research we adopted a cross-validation scheme with holdout data for the final model evaluation. Model overfitting occurs because a model that has less training error (i.e. misclassifications on training data) can have poor generalization (expected classification errors on new unseen data) than a model with higher training error. As a result, we took extra steps to avoid partially overlapping subsets of cases by splitting our dataset into training and testing subsets with a validation set. Hence, model testing was always performed on unseen cases which were not considered during the training phase and, consequently, did not influence the feature selection process. This procedure helps to minimize misclassifications on the training phase while also ensuring lessening of generalization errors.

Classification with Balanced Random Forest algorithm.

Univariate data imputation, feature selection and class imbalance handling were considered as a chain of transforms and estimators to build the composite estimator for models' training (i.e., the Balanced Random Forest algorithm). Cross-validated grid-search procedure with test set evaluation enhanced the generalizability of our models by finding the best parameters in the defined hyper-parameter space to achieve the best cross validation score (in terms of several evaluation metrics). The initial pool of data was randomly split into a training set (80% of the whole dataset) and a holdout set (20%) for models' training and testing. A 5-fold cross validation was selected for the model selection during the grid search procedure. The number of trials for the inner loop during the model training was set to 50 for consistency purposes. Python programming language was used along with the scikit-learn ML library[5], for the design and development of the BOUNCE flexible and comprehensive ML-based pipeline.

Class imbalance handling was addressed by random under-sampling to balance the subsets combined inside an ensemble. The Balanced Random Forest classifier[6,7]combines the down

[5] Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.

[6] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," The Journal of Machine Learning Research, vol. 18, pp. 559-563, 2017.

[7] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, pp. 539-550, 2008.

sampling majority class technique and the ensemble learning approach, artificially adjusting the class distribution so that classes are represented equally in each tree in the forest. In this manner, each bootstrap sample contains balanced down-sampled data. Applying random-under sampling to balance the different bootstraps in an RF classifier could have classification performance superior to most of the existing conventional ML-based estimators while alleviating the problem of learning from imbalanced datasets.

The following metrics were calculated to assess the performance of the classification models: specificity (true negative rate), sensitivity (true positive rate), accuracy, precision, and F-measure. The Receiver Operating Characteristic (ROC) curve was also computed to represent the trade-off between the false negative and false positive rates for every possible cut off. The Area Under the ROC curve (AUC) was used as a scoring metric during the search over the specified parameters to assess subsequently the performance of the cross-validated model on the test set.

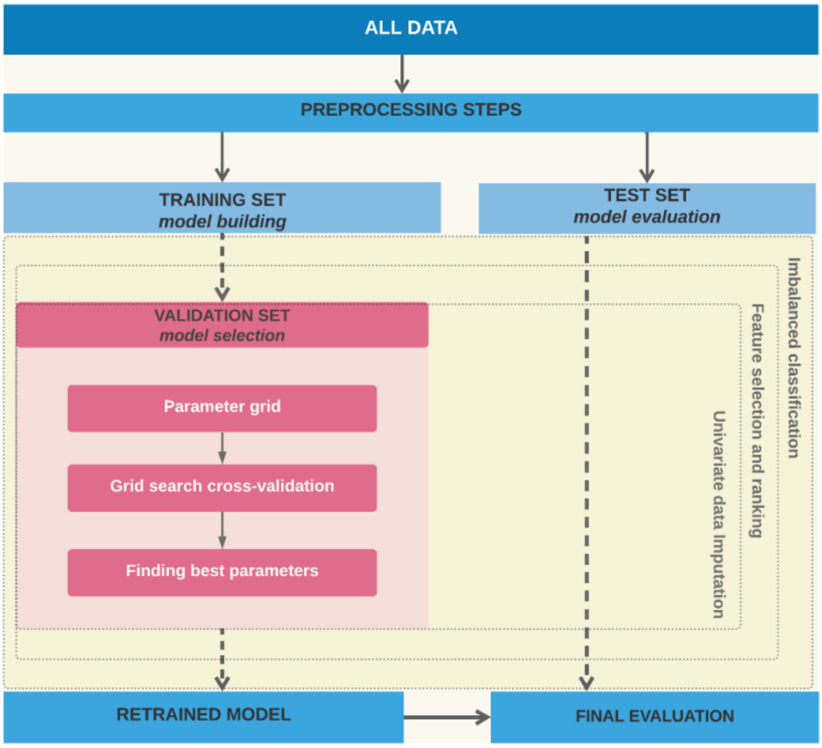


Figure F1. The data division scheme of the proposed supervised learning pipeline for training, testing and validation sets. Cross validation grid search with test set were utilized for more accurate and generalizable

1.2.3. Results

1.2.3.1. Cross-validation classification performance predicting mental health and QoL deterioration at M12 based on M0 and M3 data

1.2.3.1.1. Prediction of Mental health deterioration

As shown in Table FI, Model 2 correctly predicted **one-year** mental health deterioration for 84% of patients. Moreover, the model identified the patients who had stable good mental health status at M12 with approximately 85% certainty. The shape of the Receiver Operating Characteristic Curve shown in Figure F2 (AUC=0.876) illustrates a fair balance between sensitivity and specificity.

Most important predictors included variables measured shortly after disease diagnosis, as well as variables reported at the 3-month follow-up (that is, during treatment; see Figure F3). They comprised life-style characteristics (at least moderate, regular exercise), trait resilience and other psychological characteristics presumed to be associated with illness adaptation, emotional status of the patient (particularly on month 3), and specific, illness-related physical symptoms. In addition, three biological variables ranked among the important predictors: thrombocyte count, NLR, and serum creatinine levels (although the latter did not vary significantly between groups; see Table FII).

Descriptive statistics of the selected continuous variables are shown in Table FIII, whereas group data on exercise at M0, which also emerged as an important predictor, is shown in Table FIV. As expected, the Stable Mental health group reported significantly lower symptomatology and better global QoL at both M0 and M3 ($p<0.001$). Basic sociodemographic characteristics of the two groups are listed in Table FIV.

Results were compared with the output of a reference model using logistic regression. Variables were force-entered into this model if their association with the dependent (categorical) variable approached significance ($p<0.1$). In total, 38 variables met this criterion and the model represented a good fit to the data, $X^2(38) = 133.75$, $p<0.001$, $R^2 = 0.599$. The logistic function achieved very high specificity (97%) and considerably lower sensitivity (56%).

Table FI. Model performance predicting M12 mental health and QoL deterioration according to the proposed ML-based pipeline and the RF estimator (values are means ± SD).

	Model 1	**Model 2**	**Model 3**	**Model 4**
Outcome	M12 Mental Health	M12 Mental Health	M12 Global QoL	M12 Global QoL
Predictors	M0 & M3 variables (excluding HADS and QoL indices)	All M0 & M3 variables	M0 & M3 variables (excluding HADS and QoL indices)	All M0 & M3 variables
Accuracy	0.746 ± 0.03	0.840 ± 0.11	0.671 ± 0.02	0.759 ± 0.19
Balanced accuracy	0.710 ± 0.14	0.786 ± 0.06	0.622 ± 0.04	0.791 ± 0.02
Sensitivity	0.746 ± 0.03	0.840 ± 0.11	0.671 ± 0.02	0.759 ± 0.02
Specificity	0.759 ± 0.03	0.858 ± 0.13	0.701 ± 0.12	0.740 ± 0.02
AUC	0.789 ± 0.04	0.876 ± 0.05	0.775 ± 0.06	0.832 ± 0.03
F1	0.400 ± 0.13	0.533 ± 0.10	0.385 ± 0.04	0.571 ± 0.03

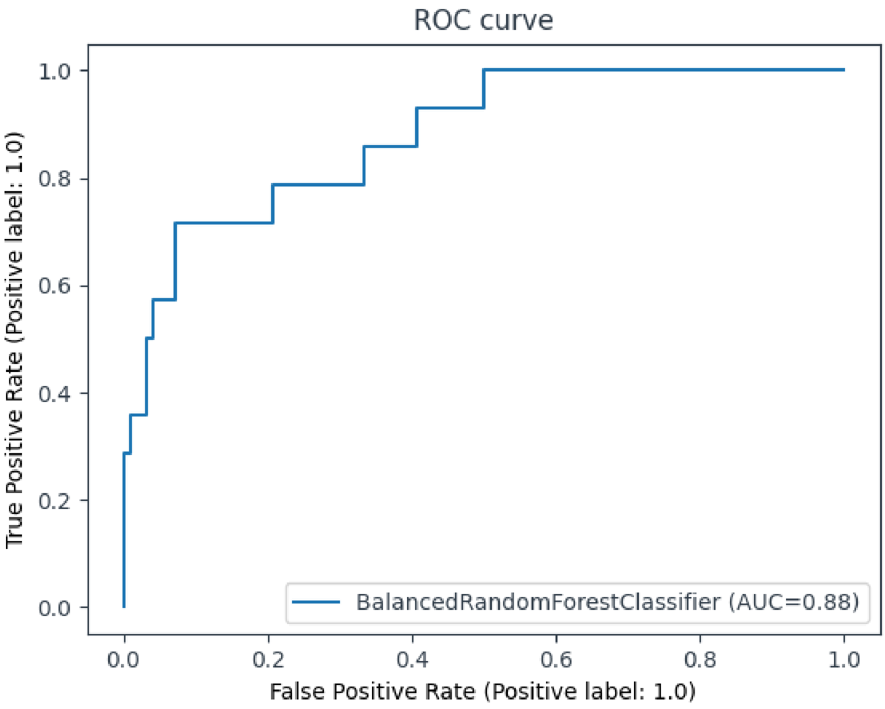


Figure F2. Receiver Operating Characteristic Curve for discriminating between Deteriorating and Stable-Good Mental health groups through M12 using all available variables registered within 3 months post-diagnosis (Model 2 in Table FI).

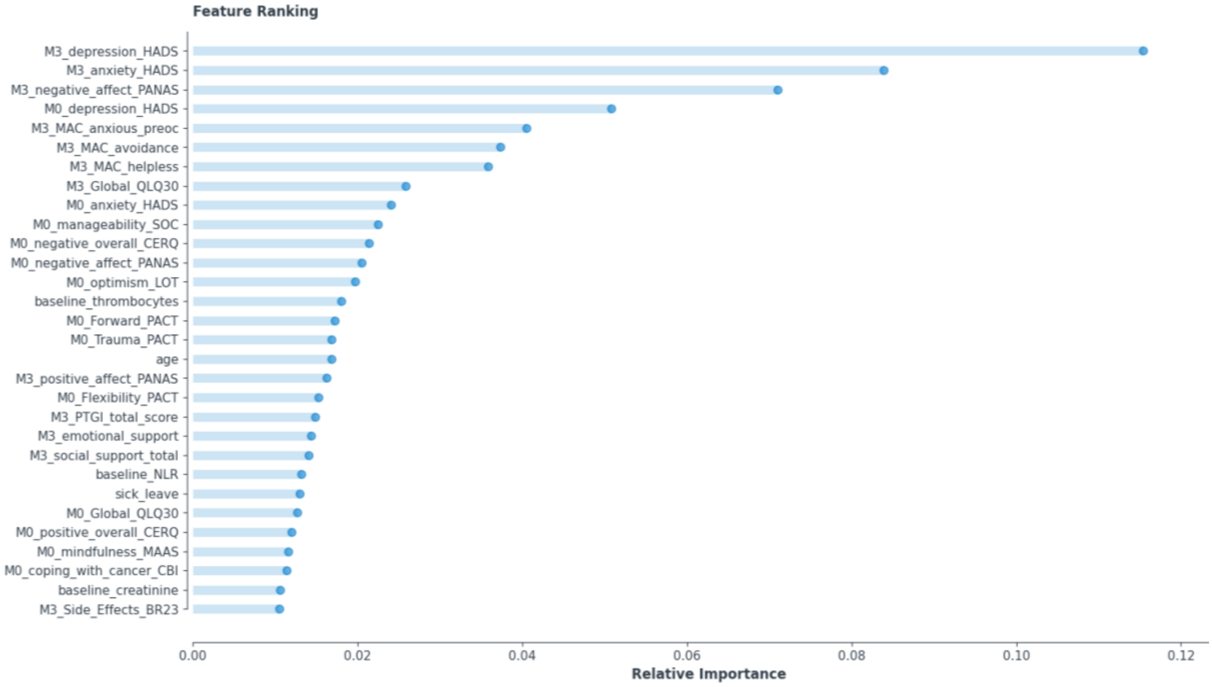


Figure F3. The selected features for Model 2 in Table FI ranked according to their relative importance for prediction of mental health change between M0 and M12. M0 indicates variables assessed at baseline and M3 variables assessed at M3.

Table FII. Cancer-related characteristics at M0 and M3 measurement waves by mental health change group.

Mental Health Status (diagnosis to month 12)	Stable-Good (n=326)	Deteriorated (n=50)	P value
Cancer stage			0.4
I	53.0	47.1	
II	39.0	43.1	
III	8.0	9.8	
Molecular tumor characteristics			
Luminal A	77.6	68.0	0.3
Luminal B	10.0	12.0	
Triple negative	5.0	12.0	
HER2 enriched	3.8	6.0	
Thrombocytes (x10^3/mL)[1]	267.7 (67.4)	244.9 (51.1)	0.027
Creatinine (mg/dl)[1]	66.7 (10.1)	68.5 (12.4)	0.2
NLR[1]	0.60 (0.20)	0.73 (0.21)	<0.001
Surgery			
…Mastectomy	22.7	30.0	0.3
…Lumpectomy	77.3	70.0	0.3
Chemotherapy	50.9	42.0	0.3
Adjuvant	36.9	29.3	0.3
Neoadjuvant	13.1	22.2	
Endocrine therapy	87.2	78.0	0.1
Anti HER2 treatment	16.7	18.0	0.8
Radiotherapy	83.2	68.2	0.02
Hospitalization by M3	9.2	10.6	0.4

Values are percentages with the exception of variables marked by[1]. NLR: Neutrophil to leukocyte ratio.

Table FIII. Psychosocial characteristics that optimally differentiate patients who displayed stable-good from those who showed deteriorating mental health according to Machine Learning Models 2 and/or 1.

Mental Health Status (diagnosis to month 12)	Stable-Good (n=326)	Deteriorated (n=50)	P value
Measured at baseline			
Manageability (SOC)^{1,2}	21.2 (3.5)	18.1 (4.0)	<0.001
Negative Affectivity (PANAS)^{1,2}	1.6 (0.5)	1.9 (0.6)	<0.001
Coping with Cancer (CBI)^{1,2}	7.5 (0.9)	6.9 (1.2)	<0.001
Trait Resilience^{1,2}	3.0 (0.5)	2.7 (0.7)	<0.001
Forward (PACT)^{1,2}	5.4 (0.9)	5.1 (1.0)	0.017
Future Perspective^{1,2}	60.4 (25.0)	42.7 (30.9)	<0.001
Optimism (LOT)^{1,2}	2.9 (0.6)	2.5 (0.6)	<0.001
Trauma (PACT)^{1,2}	5.4 (0.8)	5.1 (0.8)	0.006
Meaningfulness (SOC)^{1}	23.6 (3.3)	21.6 (4.1)	<0.001
Mindfulness (MAAS)^{1,2}	4.5 (0.7)	4.2 (0.8)	0.008
Comprehensibility (SOC)^{1}	21.7 (3.6)	19.9 (4.3)	0.002
Arm symptoms (BR-23)^{1}	11.1 (15.9)	21.3 (22.0)	<0.001
Flexibility (PACT)^{1,2}	10.2 (1.7)	9.8 (2.0)	0.01
Positive Emotion Regulation (CERQ)^{2}	3.4 (0.7)	3.3 (0.7)	0.1
HADS Anxiety^{2}	5.2 (2.8)	7.3 (2.5)	<0.001
HADS Depression	2.5 (2.1)	4.4 (2.2)	<0.001
Global QoL	78.5 (15.6)	69.2 (19.7)	<0.001
Measured at month 3			
Negative Affectivity (PANAS)^{1,2}	1.4 (0.5)	2.1 (0.6)	<0.001
Anxious Preoccupation (MAC)^{1,2}	1.9 (0.5)	2.4 (0.6)	<0.001
Helplessness (MAC)^{1,2}	1.3 (0.3)	1.6 (0.5)	<0.001
Social Support^{1,2}	4.2 (0.8)	3.7 (0.8)	<0.001
Treatment Side Effects (BR-23)^{1,2}	23.9 (16.4)	30.1 (16.2)	0.01
Avoidance (MAC)^{1,2}	2.3 (0.7)	2.7 (0.6)	<0.001
Body Image (BR-23)^{2}	80.3 (20.7)	70.7 (22.6)	<0.001
Community Cohesion (FARE)^{1}	6.2 (0.9)	6.0 (0.9)	0.2
Emotional Support^{1,2}	4.2 (0.8)	3.8 (0.8)	<0.001
Future Perspective^{1,2}	65.6 (20.6)	49.9 (27.1)	<0.001
Positive Affectivity (PANAS)^{1,2}	3.45 (0.6)	3.3 (0.6)	0.04
PTGI^{2}	2.4 (1.2)	2.8 (1.1)	0.027
HADS Anxiety^{2}	4.1 (2.6)	7.7 (2.7)	<0.001
HADS Depression^{2}	2.9 (2.6)	7.3 (3.2)	<0.001
Global QoL^{2}	72.8 (18.9)	62.1 (17.3)	<0.001

Notes: Values are means (SD). Superscripts indicate important features within Model 1 and 2, respectively. Abbreviations; SOC: Sense of coherence, LOT: Life orientation test, MAAS: Mindful attention awareness scale, PANAS: Positive and negative affect , CERQ: Cognitive emotion regulation questionnaire, CBI: Cancer behavior inventory, PACT: Perceived ability to cope with trauma, BR-23: Breast cancer–23, MAC: Mental adjustment to cancer, PTGI: Post-traumatic growth inventory, FARE: Family resilience, HADS: Hospital Anxiety and Depression Scale.

Table FIV. Patient sociodemographic, clinical and lifestyle characteristics at M0 and M3 measurement waves by mental health change group.

Mental Health Status (diagnosis to month 12)	Stable-Good (n=326)	Deteriorated (n=50)	P value
Age (mean (SD) in years)	56.5 (8.1)	54.5 (8.3)	0.1
Education (≤9 years)	94.2	84.0	0.016
Has children	86.6	78.0	0.1
Has partner	74.9	86.0	0.1
Currently employed	75.3	70.7	0.5
Full-time, retired, or self-employed	89.3	78.0	0.035
Unemployed, housewife, or part-time	10.7	22.2	
Low income	16.8	34.1	0.012
Life stressors (≥2)	32.4	30.0	0.8
Obesity	19.3	14.3	0.5
Family history of BC	34.2	42.0	0.3
Physical comorbidity (chronic)	37.5	34.0	0.7
Metabolic Comorbidity	25.6	20.0	0.5
History of anxiety disorder/dysthymia	9.7	16.0	0.2
Psychotropic medication	13.8	18.0	0.4
Menopausal status			0.6
Premenopausal	29.3	34.0	
Perimenopausal	5.2	4.0	
Postmenopausal	65.4	62.0	
Mental health support by month 3	14.3	21.7	0.2
Sick leave (mean (SD) in days))	58.8 (81.1)	78.7 (111.8)	0.1
Current smoker	32.1	30.6	0.8
Alcohol consumption			
No/occasional	31.2	32.0	0.2
Moderate	58.7	66.0	
Heavy	10.1	2.0	
Diet			
Mediterranean	26.94	42.0	0.03
Special diet	16.0	14.0	0.8
Exercise			
No/occasional	21.2	50.0	<0.001
Moderate	40.6	30.0	
Heavy	38.2	20.0	

Values are percentages unless otherwise indicated

1.2.3.1.2. Prediction of QoL deterioration

As shown in Table FI, Model 4 correctly predicted **one-year** global QoL deterioration for 76% of patients. Moreover, the model identified the patients who had stable good QoL through M12 with approximately 74% certainty. The shape of the Receiver Operating Characteristic Curve shown in Figure F4 (AUC=0.876) illustrates a fair balance between sensitivity and specificity.

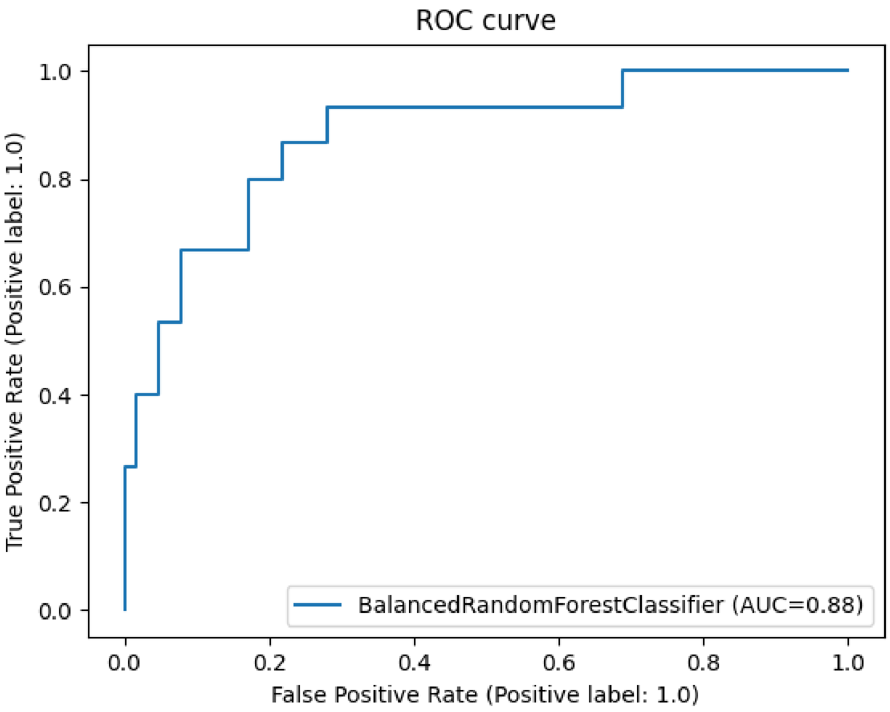


Figure F4. Receiver Operating Characteristic Curve for discriminating between Deteriorating and Stable-Good QoL groups through M12 using all available variables registered within 3 months post-diagnosis (Model 4 in Table FI).

Both M0 and M3 variables featured among the highest-ranking ones (see Figure F5). They comprised primarily of treatment side effects and other psychological characteristics presumed to be associated with illness adaptation (such as mindfulness, emotion regulation strategies, and coping styles) as well as emotional status and subjective QoL of the patient (particularly on month 3). In addition, trait resilience and additional life stressors taking place at or immediately preceding M3 and age emerged as important predictors of one-year QoL deterioration. Biological indices that featured among the top predictor variables in this model include the NLR variable measured at the time of diagnosis.

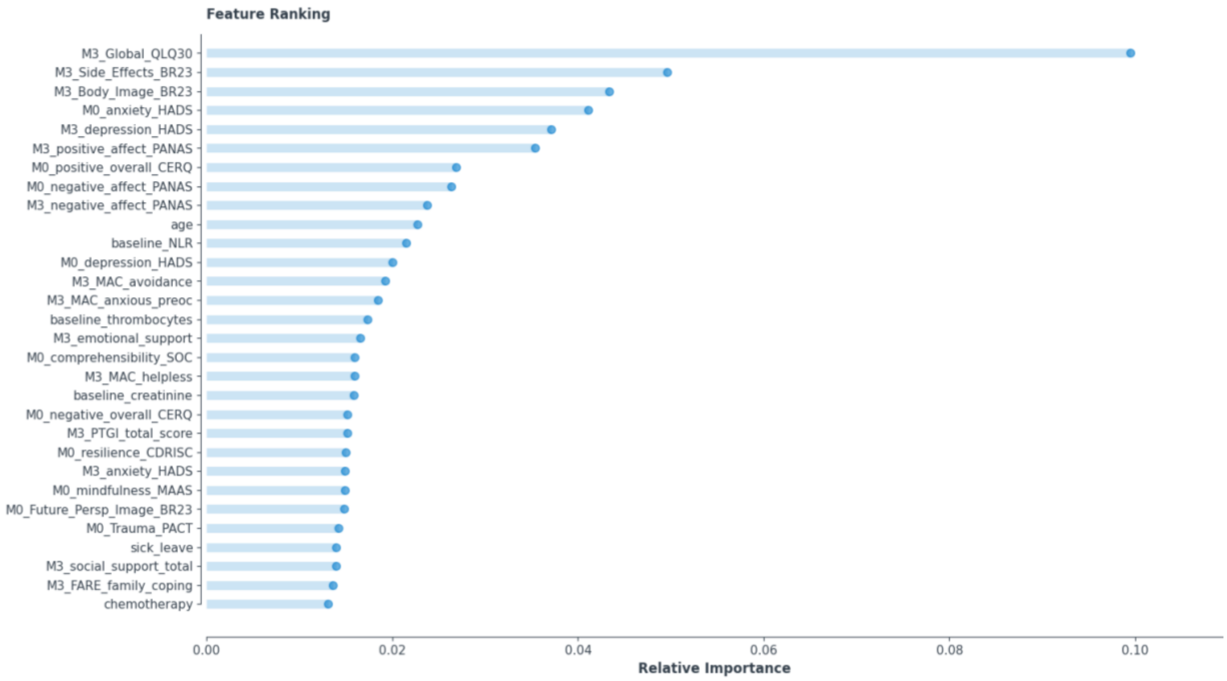


Figure F5. The selected features for Model 4 in Table FI ranked according to their relative importance for prediction of global QoL change between M0 and M12. M0 indicates variables assessed at baseline and M3 variables assessed at M3.

1.2.3.2. Cross-validation classification performance predicting mental health and QoL deterioration at M18 based on M0 and M3 data

1.2.3.2.1. Prediction of Mental health deterioration

As shown in Table FV, Model 6 correctly predicted 18-month overall mental health deterioration for 76% of patients. Moreover, the model identified the patients who maintained mild symptomatology through M18 with approximately 77% certainty (AUC=0.853; see Figure F6).

As in the case of the prediction of M12 mental health deterioration both M0 and M3 variables featured among the highest-ranking ones (see Figure F7). They comprised primarily of psychological characteristics presumed to be associated with illness adaptation (such as optimism, perceived emotional support by others, and coping styles) as well as emotional status and subjective QoL of the patient (particularly on month 3). Importantly, the two biological indices which were among the top-ranking variables in predicting M12 mental health deterioration, also feature among the most important features in Model 6 (NLR and platelet count at M0). In addition, at least moderate exercise at the time of diagnosis, absence of treatment side effects and physical symptoms, and engaging in well-being promoting activities at M3 also contributing to remaining free of symptoms of anxiety and depression one and half year post diagnosis.

Table FV. Model performance predicting M18 mental health and QoL deterioration according to the proposed ML-based pipeline and the RF estimator (values are means ± SD).

	Model 5	**Model 6**	**Model 7**	**Model 8**
Outcome	M18 Mental Health	M18 Mental Health	M18 Global QoL	M18 Global QoL
Predictors	M0 & M3 variables (excluding HADS and QoL indices)	All M0 & M3 variables	M0 & M3 variables (excluding HADS and QoL indices)	All M0 & M3 variables
Accuracy	0.656 ± 0.02	0.763 ± 0.05	0.671 ± 0.02	0.767 ± 0.07
Balanced accuracy	0.686 ± 0.02	0.757 ± 0.03	0.707 ± 0.02	0.761 ± 0.04
Sensitivity	0.656 ± 0.02	0.763 ± 0.05	0.671 ± 0.02	0.767 ± 0.07
Specificity	0.645 ± 0.04	0.765 ± 0.06	0.637 ± 0.02	0.773 ± 0.09
AUC	0.786 ± 0.09	0.853 ± 0.08	0.791 ± 0.05	0.849 ± 0.04
F1	0.357 ± 0.01	0.455 ± 0.05	0.531 ± 0.02	0.592 ± 0.07

Note: Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results.

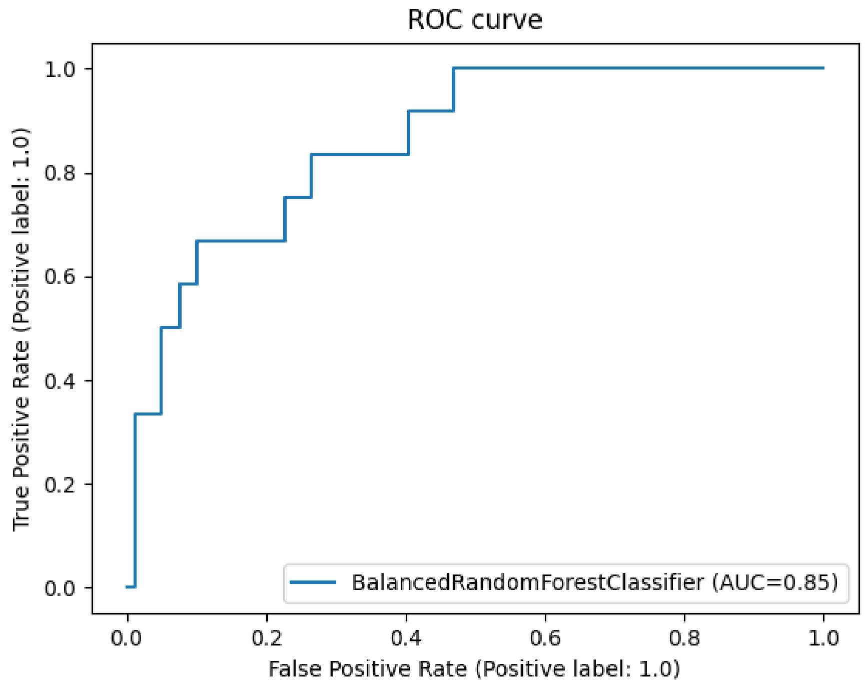


Figure F6. Receiver Operating Characteristic Curve for discriminating between Deteriorating and Stable-Good Mental health groups through M18 using all available variables registered within 3 months post-diagnosis (Model 6 in Table FV).

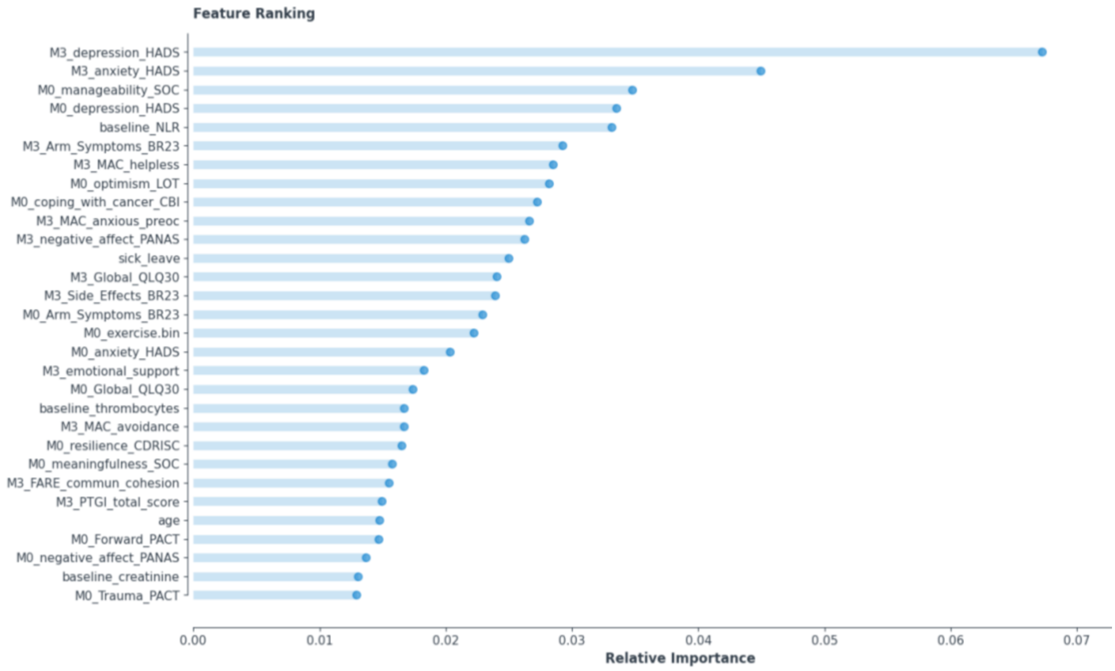


Figure F7. The selected features for Model 6 in Table FV ranked according to their relative importance for prediction of mental health change between M0 and M18. M0 indicates variables assessed at baseline and M3 variables assessed at M3.

1.2.3.2.2. Prediction of QoL deterioration

As shown in Table FV, Model 8 correctly predicted one-year global QoL deterioration for 77% of patients. Moreover, the model identified the patients who had stable good QoL at M18 with approximately 78% certainty (AUC=0.849; see Figure F8).

Both M0 and M3 variables featured among the highest-ranking ones (see Figure F9) as was the case in predicting M12 QoL. With the exception of trait resilience, psychological characteristics presumed to be associated with illness adaptation (such as mindfulness, emotion regulation strategies, perceived social support, and coping styles) as well as emotional status and subjective QoL of the patient (particularly on month 3). In addition, treatment side effects and physical symptoms and age emerged as important predictors of QoL deterioration. Interestingly, the two biological indices which were among the top-ranking variables in predicting M12 and M18 mental health deterioration, also feature among the most important variables in Model 8 (NLR and platelet count at M0).

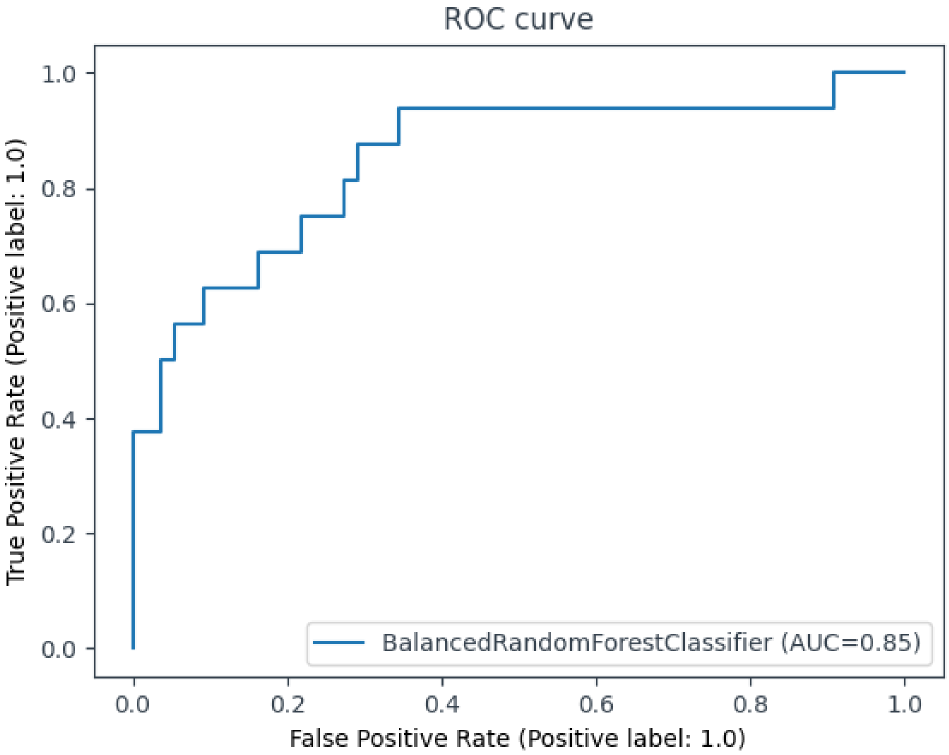


Figure F8. Receiver Operating Characteristic Curve for discriminating between Deteriorating and Stable-Good QoL groups through M18 using all available variables registered within 3 months post-diagnosis (Model 8 in Table FV).

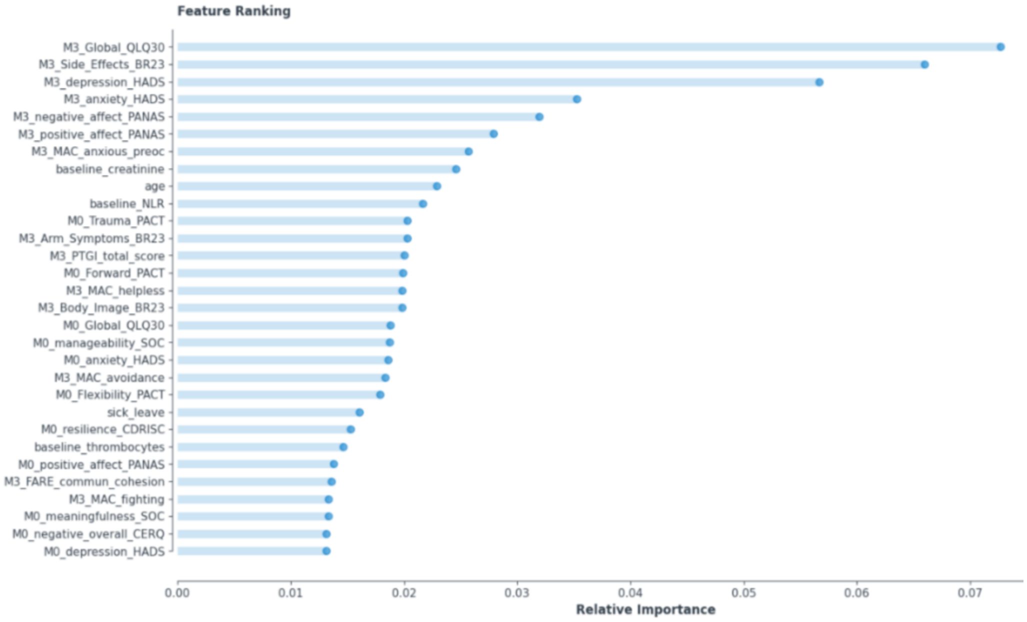


Figure F9. The selected features for Model 8 in Table FV ranked according to their relative importance for prediction of global QoL change between M0 and M18. M0 indicates variables assessed at baseline and M3 variables assessed at M3.

1.2.3.3. Supplementary models predicting mental health and QoL at M18 based on M0 and M3 data

Extending the work aiming to develop models that predict mental health and global QoL status at M12, we implemented models that use clinical, biological, psychological and lifestyle variables to predict overall mental health and global QoL status at M18, respectively. This work extends the prediction models 2 and 4 in Table FIII of D4.3a, respectively. Performance was somewhat inferior as indicated by Accuracy= 0.76 ± 0.04, AUC=0.74 ± 0.05, Sensitivity=0.71 ± 0.10, and Specificity= 0.76 ± 0.04 for mental health and Accuracy= 0.65 ± 0.05, AUC=0.64 ± 0.06, Sensitivity=0.62 ± 0.11, and Specificity= 0.66 ± 0.06.

1.2.3.4. Supplementary models predicting mental health and QoL at M12 based on M0 and/M6 data

Additional supervised prediction models were implemented toward predicting overall mental health outcomes (identical to those described in previous sections) based on psychological and lifestyle variables measured at M6 rather than M3 (in addition to clinical variables). The aim of these models was to enhance the flexibility of the BOUNCE Decision Support Tool toward accommodating alternative clinical scenario (i.e., risk assessment of patients who are assessed at the time of diagnosis and 6 months later while cancer treatments are well underway). These models showed very comparable performance to those relying on M0 and M3 measurements as shown in Tables FVI and FVII.

Table FVI. Model performance predicting M12 mental health and QoL **general status ("good" vs "poor")** based on M0 and/or M6 measurements (values are means ± SD).

	Model 9	Model 10	Model 11	Model 12
Outcome	M12 Mental Health	M12 Mental Health	M12 Global QoL	M12 Global QoL
Predictors	All M0 clinical and all M6 variables	All M0 & M6 variables	All M0 clinical and all M6 variables	All M0 & M6 variables
Accuracy	0.78 ± 0.04	0.78 ± 0.04	0.75 ± 0.04	0.74 ± 0.04
Balanced accuracy	0.79 ± 0.05	0.80 ± 0.06	0.78 ± 0.06	0.77 ± 0.05
Sensitivity	0.78 ± 0.10	0.77 ± 0.10	0.72 ± 0.10	0.73 ± 0.11
Specificity	0.77 ± 0.05	0.78 ± 0.05	0.75 ± 0.05	0.74 ± 0.05
AUC	0.78 ± 0.05	0.77 ± 0.05	0.74 ± 0.05	0.74 ± 0.05
F1	0.56 ± 0.06	0.56 ± 0.06	0.52 ± 0.06	0.52 ± 0.06

Note: Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results. All models include HADS and QoL indices measured at M0 and/or M6.

Table FVII. Model performance predicting M12 mental health and QoL **deterioration** based on M0 and/or M6 measurements (values are means ± SD).

	Model 13	Model 14	Model 15	Model 16
Outcome	M12 Mental Health	M12 Mental Health	M12 Global QoL	M12 Global QoL
Predictors	All M0 clinical and all M6 variables	All M0 & M6 variables	All M0 clinical and all M6 variables	All M0 & M6 variables
Accuracy	0.81 ± 0.04	0.82 ± 0.04	0.80 ± 0.05	0.80 ± 0.05
Balanced accuracy	0.83 ± 0.03	0.83 ± 0.05	0.82 ± 0.08	0.81 ± 0.06
Sensitivity	0.75 ± 0.13	0.74 ± 0.14	0.78 ± 0.12	0.78 ± 0.12
Specificity	0.83 ± 0.08	0.83 ± 0.05	0.81 ± 0.06	0.81 ± 0.06
AUC	0.79 ± 0.07	0.78 ± 0.07	0.80 ± 0.06	0.79 ± 0.06
F1	0.52 ± 0.07	0.51 ± 0.09	0.62 ± 0.06	0.61 ± 0.08

Note: Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results. All models include HADS and QoL indices measured at M0 and/or M6.

1.2.3.5. Supplementary models predicting mental health and QoL at M18 based on M0 and/M6 data

Additional supervised prediction models were implemented toward predicting Global QoL outcomes (identical to those described in previous sections) based on psychological and lifestyle variables measured at M6 rather than M3. These models displayed comparable performance to those relying on M0 and M3 measurements as shown in Tables FVIII and FIX (with the exception of the models predicting overall Global QoL at M18 (Models 19-20).

Table FVIII. Model performance predicting M18 mental health and QoL **general status ("good" vs "poor")** based on M0 and/or M6 measurements (values are means ± SD).

	Model 17	Model 18	Model 19	Model 20
Outcome	M18 Mental Health	M18 Mental Health	M18 Global QoL	M18 Global QoL
Predictors	All M0 clinical and all M6 variables	All M0 & M6 variables	All M0 clinical and all M6 variables	All M0 & M6 variables
Accuracy	0.79 ± 0.04	0.79 ± 0.04	0.70 ± 0.05	0.71 ± 0.04
Balanced accuracy	0.80 ± 0.05	0.82 ± 0.06	0.73± 0.06	0.73 ± 0.05
Sensitivity	0.79 ± 0.10	0.78 ± 0.11	0.63 ± 0.10	0.66 ± 0.11
Specificity	0.79 ± 0.05	0.79 ± 0.05	0.71 ± 0.06	0.72 ± 0.06
AUC	0.79 ± 0.05	0.78 ± 0.06	0.67 ± 0.05	0.69 ± 0.06
F1	0.54 ± 0.07	0.53 ± 0.07	0.43 ± 0.06	0.47 ± 0.06

Note: Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results. All models include HADS and QoL indices measured at M0 and/or M6.

Table FIX. Model performance predicting M18 mental health and QoL **deterioration** based on M0 and/or M6 measurements (values are means ± SD).

	Model 21	Model 22	Model 23	Model 24
Outcome	M18 Mental Health	M18 Mental Health	M18 Global QoL	M18 Global QoL
Predictors	All M0 clinical and all M6 variables	All M0 & M6 variables	All M0 clinical and all M6 variables	All M0 & M6 variables
Accuracy	0.83 ± 0.04	0.84 ± 0.04	0.82 ± 0.06	0.82 ± 0.05
Balanced accuracy	0.84 ± 0.04	0.86 ± 0.05	0.83 ± 0.06	0.84 ± 0.07
Sensitivity	0.80 ± 0.13	0.83 ± 0.13	0.80 ± 0.12	0.81 ± 0.11
Specificity	0.84 ± 0.05	0.84 ± 0.05	0.83 ± 0.07	0.83 ± 0.07
AUC	0.82 ± 0.07	0.83 ± 0.06	0.81 ± 0.07	0.82 ± 0.06
F1	0.58 ± 0.09	0.58 ± 0.08	0.67 ± 0.09	0.69 ± 0.08

Note: Cross validation using grid search optimization against Area Under the ROC Curve (AUC) was applied following a 3-fold data division scheme for better and more generalizable results. All models include HADS and QoL indices measured at M0 and/or M6.

1.2.4. Discussion

The main goal of the analyses outlined in Section 1.2 was to develop a framework that would allow accurate identification of BC patients at risk for significant decline in mental health or subjective QoL among those who report relatively good levels of well-being at the time of diagnosis. The design of the present analyses was guided primarily by the potential future clinical utility of the model results. Thus, the supervised learning models included variables that could be readily available to clinicians in future practice, namely medical, sociodemographic, and lifestyle variables integrated with a select set of psychosocial patient characteristics.

For each classification problem defined in the current analysis (i.e. mental health and global QoL prediction at M12 and M18), mental health and subjective QoL ratings were incorporated into the respective classification model (i.e. Model 2, Model 4, Model 6 and Model 8) toward optimizing the prediction of one-year and one year and half adverse mental health and QoL outcomes.

A rigorous analytic approach was employed to mitigate some of the commonly observed pitfalls of machine learning approaches, namely overfitting and poor model generalizability. To address these issues our pipeline entailed feature selection, model training, validation and testing based on the four clinical sites that contributed to the BOUNCE prospective study. Further, the generalizability of the prediction model was assessed on cases that were not considered during the training phase avoiding thereby overfitting while ensuring the minimization of classification errors. We increased the generalizability of predictions and the accuracy of the final prediction model when applied to new unseen data (i.e., test subsets).

1.2.4.1. Mental health prediction

The prediction model for mental health correctly classified 84% and 76% of the cases that displayed clinically significant mental health change at M12 or M18, respectively. Moreover, the model identified the patients who had stable-good mental health status at M12 or M18 with approximately 85% and 77% certainty, respectively. Hence, our results inspire optimism on transferring our prediction model to relevant clinical settings.

The variables that emerged as significant predictors of changes in mental health change over one year and one year and a half (i.e. stable-good or deterioration) correspond to certain “clusters” of factors, namely a) negative affect; b) coping with cancer responses and self-efficacy to cope with cancer; c) a sense of control/positive expectations (i.e., sense of coherence; optimism); d) social and family support; e) certain lifestyle factors (i.e., exercise) and, (f) certain symptoms (e.g., arm symptoms). These findings are in accordance with the major psychological theories about adaptation to severe illness, including BC, such as the Common Sense Model[8] or the Transactional Stress Model[9]. According to these models,

[8] Leventhal, H., Halm, E., Horowitz, C., Leventhal, E.A., & Ozakinci, G. (2005). Living with chronic illness: A contextualized, self–regulation approach. In S. Sutton, A. Baum & M. Johnston (Eds.), *The SAGE handbook of health psychology* (pp. 197-240). London: Sage.

[9] Lazarus, R., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

adaptation to a severe health crisis is a complex process which is determined by: (a) a variety of personal and interpersonal resources, such as expectations, lifestyle, or social support, which may buffer the negative impact of the situation and facilitate adaptation; (b) cognitive-emotional processes, such as affect and emotion regulation and self-efficacy to cope with cancer, that guide behaviors such as preoccupation and helplessness; (c) contextual and specific stressor-related factors that may impact adaptation directly or indirectly, such as physical symptoms. In addition, the findings pinpoint those early factors, coming from a large array of sociopsychological, medical, and lifestyle variables, that are significant predictors of the outcome and, in this way, will guide the efforts to develop appropriate clinical recommendations which may guide informed and shared decision making between health professionals and patients, to promote adapted mental health status.

1.2.4.2. Global QoL prediction

Overall, prediction of mental health was somewhat more accurate than prediction of global QoL despite the substantial overlap between groupings according to mental health symptoms and global QoL (68%). Specifically, the prediction models correctly classified 77% and 78% of the cases that displayed clinically significant decline in global QoL at M12 or M18, respectively. Moreover, the model identified the patients who maintained adequate global QoL at M12 or M18 with approximately 74% and 77% certainty, respectively. A potential reason for this might be that global QoL is a general, overarching concept which encompasses a variety of personal evaluations regarding physical, psychological, and social well-being, as well as self-rated health and patient perspective of their future, which very often change over the course of illness. In this regard, global QoL is indeed a very useful but also a volatile and thus difficult to predict variable. An additional reason may refer to the way that global QoL was assessed. That is, with two general items, whereas mental health was evaluated with a set of concrete and easy to understand and report symptoms.

Moreover, prediction accuracy of M12 outcomes is only slightly higher than prediction accuracy of M18 outcomes. As trajectory analyses suggest (see Section 1.5) both overall mental health and global QoL has largely reached plateau by month 12 post-diagnosis for the majority of participants. This notion is supported by the fact that the overlap between patient subgroups formed on the basis of M12 outcomes and those formed by also considering M18 outcomes was very high (95 for mental health and 84 for global QoL).

Finally, it is worth noting that few medical variables emerged as significant predictors of the M12 (or M18) overall mental health or QoL change. These included specific symptoms related to BC treatments[10,11], indices of immune response and even mild thrombocytopenia[12,13]. The role of BC treatments or subsequent mental and functional well-being has been highlighted by

[10] Grusdat NP, Stäuber A, Tolkmitt M, Schnabel J, Schubotz B, Wright PR, Schulz H. Routine cancer treatments and their impact on physical function, symptoms of cancer-related fatigue, anxiety, and depression. Support Care Cancer. 2022 Jan 11. doi: 10.1007/s00520-021-06787-5.

[11] Sebri, V., Triberti, S., & Pravettoni, G. (2020). Injured self: autobiographical memory, self-concept, and mental health risk in breast cancer survivors. Frontiers in Psychology, 11, 2962

[12] Ehrlich, D., & Humpel, C. (2012). Platelets in psychiatric disorders. World Journal of Psychiatry, 2(6), 91.)

[13] Scharinger, C., Rabl, U., Kasess, C. H., Meyer, B. M., Hofmaier, T., Diers, K., ... & Pezawas, L. (2014). Platelet serotonin transporter function predicts default-mode network activity. PloS one, 9(3), e92543.

recent studies[10,11]. At present we may only surmise on the mechanisms responsible for the impact of biological measures. For instance, low platelet count may contribute to serotonergic dysfunction[12]. At the systems level, low platelet count has been found to be associated with increased functional connectivity (FC) within the Default Mode Network and reduced FC within the Executive and Salience Networks[13]. In this framework, low platelet count could bias brain function toward reduced efficiency in shifting attention to external stimuli and in refocusing from negative thoughts (i.e., toward increased thought rumination). At the same time reduced efficiency of the Executive Network may negatively impact on the capacity to allocate cognitive resources necessary to support cognitive reappraisal of stressful events. Combined, these functional brain states could render patients more susceptible to the development of persistent symptoms of anxiety and depression. Further, NLR, as an index of systemic inflammation in primary emotional disorders, may be both directly and indirectly involved in the pathways leading to increased mental health symptomatology[14]. Thus systemic inflammation may impact cognitive functions, as recently shown among cancer patients[15] impacting the patient's capacity to cope with everyday tasks and also indirectly by reducing her capacity to engage cognitive resources required for active coping (such as for engaging in cognitive reappraisal.

Other medical factors (such as cancer state and histological tumor characteristics) may exert their influence on mental health and QoL indirectly; that is, through the cognitive-emotional, behavioral, and situation specific variables described above. Future analyses will examine the potential indirect impact of medical factors on the outcomes. As more measurement waves become available from the BOUNCE study, advanced statistical and computational models can be applied to explore the complex interplay of treatments side effects, life events, emotional, behavioral, and cognitive processes over time.

The most significant limitation of these analyses refers to the fact that the outcomes (i.e., mental health and global QoL) were based on self-report scales. Although these are very reliable, valid, and widely-used, they still reflect only patients' perception of their condition. Moreover, the set of variables selected by the ML as important predictors of mental health (or QoL) deterioration depend to some extent upon the characteristics of the comparison group which included only those patients who exhibited consistently high levels of psychological resilience, in an attempt to optimize group separation (and consequently model performance).

Although a very large number of variables was included in the model, still not everything could be assessed. For example, the impact of factors related to the health care system or the delivery of health services that are particular in each participant country, was not examined. In this regard, future similar studies could focus on such variables. In the same line, only mental health as a specific "binary" outcome was examined here. Other, equally important, outcomes, including physical or social functioning, as well as alternative classifications schemes (e.g.,

[14] Mazza MG, Lucchi S, Tringali AGM, Rossetti A, Botti ER, Clerici M. Neutrophil/lymphocyte ratio and platelet/lymphocyte ratio in mood disorders: A meta-analysis. Prog Neuropsychopharmacol Biol Psychiatry. 2018 Jun 8;84(Pt A):229-236. doi: 10.1016/j.pnpbp.2018.03.012.

[15] Vivek S, Nelson HH, Prizment AE, Faul J, Crimmins EM, Thyagarajan B. Cross sectional association between cytomegalovirus seropositivity, inflammation and cognitive impairment in elderly cancer survivors. Cancer Causes Control. 2022 Jan;33(1):81-90. doi: 10.1007/s10552-021-01504-3.

stable poor vs. improved mental health at 12 and 18 months after diagnosis) should also be the focus of future work.

1.3. Identifying potentially modifiable predictors of resilience outcomes: Supervised ML models and Model-Agnostic Analysis

To further support the clinical utility of the assessment and modelling framework developed within BOUNCE we conducted additional supervised analyses predicting mental health (or QoL) decline at one or one and half year post diagnosis. These models specifically aimed at identifying the most critical, yet modifiable, variables which as measured within the initial phase of BC diagnosis and treatment could help maintain adequate levels of wellbeing among BC survivors. To address this aim we run the analysis pipeline described in the previous section *omitting HADS Anxiety, HADS Depression, and Global QoL measured at M0 in order to focus on potential modifiable factors*. These results are presented in Section 1.3.1.

Additionally, we applied a model-agnostic analysis approach[16,17] to aid interpretation of prediction results at the patient level (i.e., identify predictor variables that emerge as key contributors to a given classification result after statistically controlling for all other predictors in the model) and provide reasoning behind the resilience outcome predictions. This analysis was applied on the set of variables that emerged as significant features to identify predictor variables of primary importance for a particular mental health prediction. In view of the lack of precedence in the literature we selected mathematical models that made no assumptions about data structure. An explainer object was designed on the exploratory set and personalised model predictions were decomposed into contributions of individual features and displayed using break-down profile plots. Examples of patient-level profiling developed from models designed to assess prediction of M12 deterioration of mental health are presented in Section 1.3.2.

1.3.1. Identification of modifiable risk factors for one-year decline in mental health: Group level analyses

Table FI lists performance measures for predicting mental health deterioration at 12 months post-diagnosis according to Model 1. Correct prediction was achieved for 74% of patients. Moreover, the model identified the patients who had stable good mental health status at M12 with 76% certainty (AUC=0.789; see Figure F10).

As shown in Figure F11, important predictors of 12-month patient's mental health status in the total sample included variables measured shortly after disease diagnosis as well as variables reported at the three-month follow-up (that is, during treatment). Overall emotional state (negative affect) measured at baseline and M3, certain coping reactions at M3 (i.e., anxiety preoccupation, avoidance, and helplessness), a sense of control over adversities (i.e. sense of

[16] Biecek, Przemysław. "DALEX: explainers for complex predictive models in R." *The Journal of Machine Learning Research* 19.1 (2018): 3245-3249.

[17] Baniecki, Hubert, et al. "dalex: Responsible Machine Learning with Interactive Explainability and Fairness in Python." *arXiv preprint arXiv:2012.14406* (2020).

coherence) and social support featured strongly among the most highly ranked predictors of mental health deterioration. These factors could become the focus of systematic psychological interventions in order to enhance patients' well-being and adaptation to BC. In addition, certain other variables emerged as significant predictors of mental health, such as self-efficacy to cope with cancer, resilience as trait, the ability to cope with trauma, family cohesion, future perspective, optimism, and some specific symptoms (e.g., arm symptoms, side-effects), which might also be considered as potential targets of appropriate clinical interventions. Importantly, the three biological variables highlighted by Model 1 also featured in the list of important predictors for Model 2 (thrombocyte count, NLR, and serum creatinine levels).

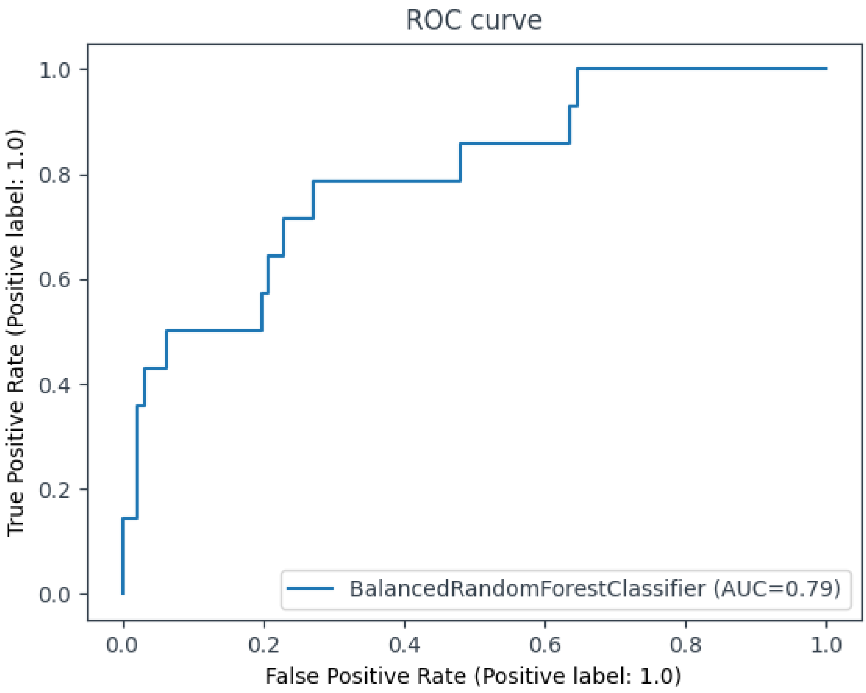


Figure F10. Receiver Operating Characteristic Curve for discriminating between Deteriorating and Stable-Good Mental health groups according to the cross-validated Model 1 (Table FI; excluding HADS and QoL scores from the set of potential predictors).

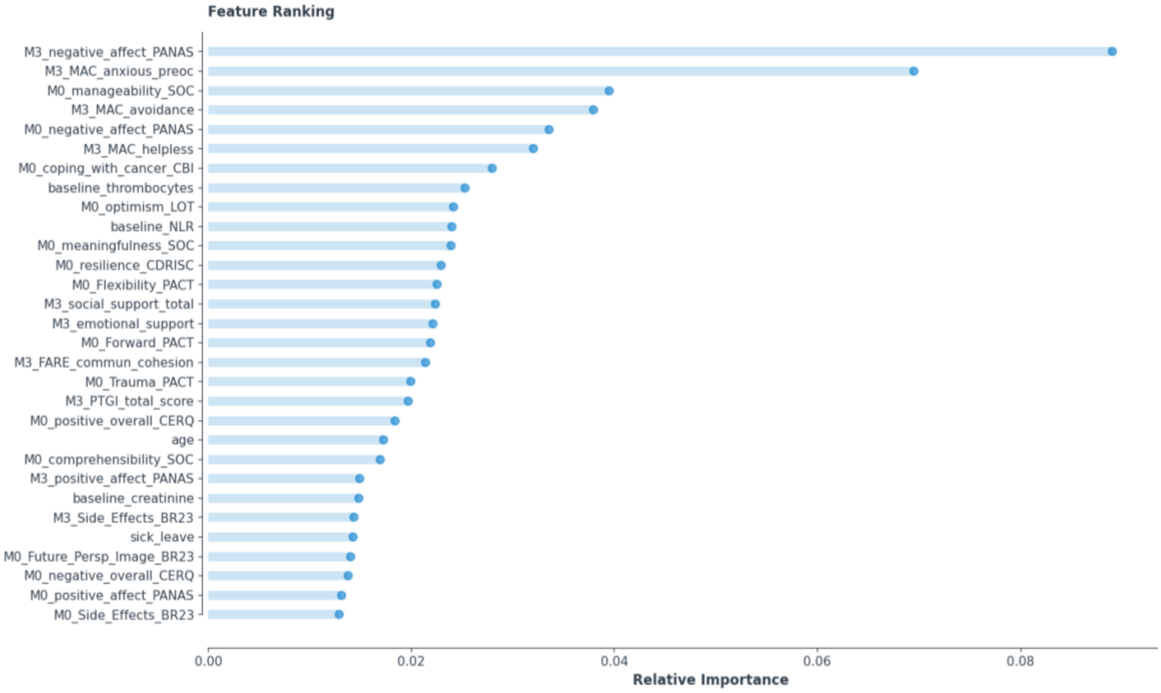


Figure F11. The selected features for Model 1 (Table FI) ranked according to their relative importance for prediction of mental health change between M0 and M12. M0 indicates variables assessed at baseline and M3 variables assessed at M3.

1.3.2. Personalized risk profiles for deteriorated mental health through M12

Examples of Break Down profile plots for randomly selected patients from each group are presented in Figures F12 and F13. The estimated contribution of a subset of variables toward a correct prediction of Deteriorated Mental Health for a randomly selected participant is shown in Figure F12. The 15 most highly ranked features selected by the RF model for this specific instance-level prediction are displayed for demonstration purposes. Nine variables predominantly “facilitate” the adverse mental health outcome for this patient: relatively high scores on Negative Affectivity (M0 and M3), anxious preoccupation (M3), helplessness (M3), combined with relatively low scores on the Future Perspectives (M0) and Resilience scale (M0). Conversely, relatively low scores on treatment side effects and negative affectivity (M3), combined with high levels of body image and mindfulness at the time of diagnosis, appear to exert a protective role for this patient by reducing the probability for an adverse mental health outcome.

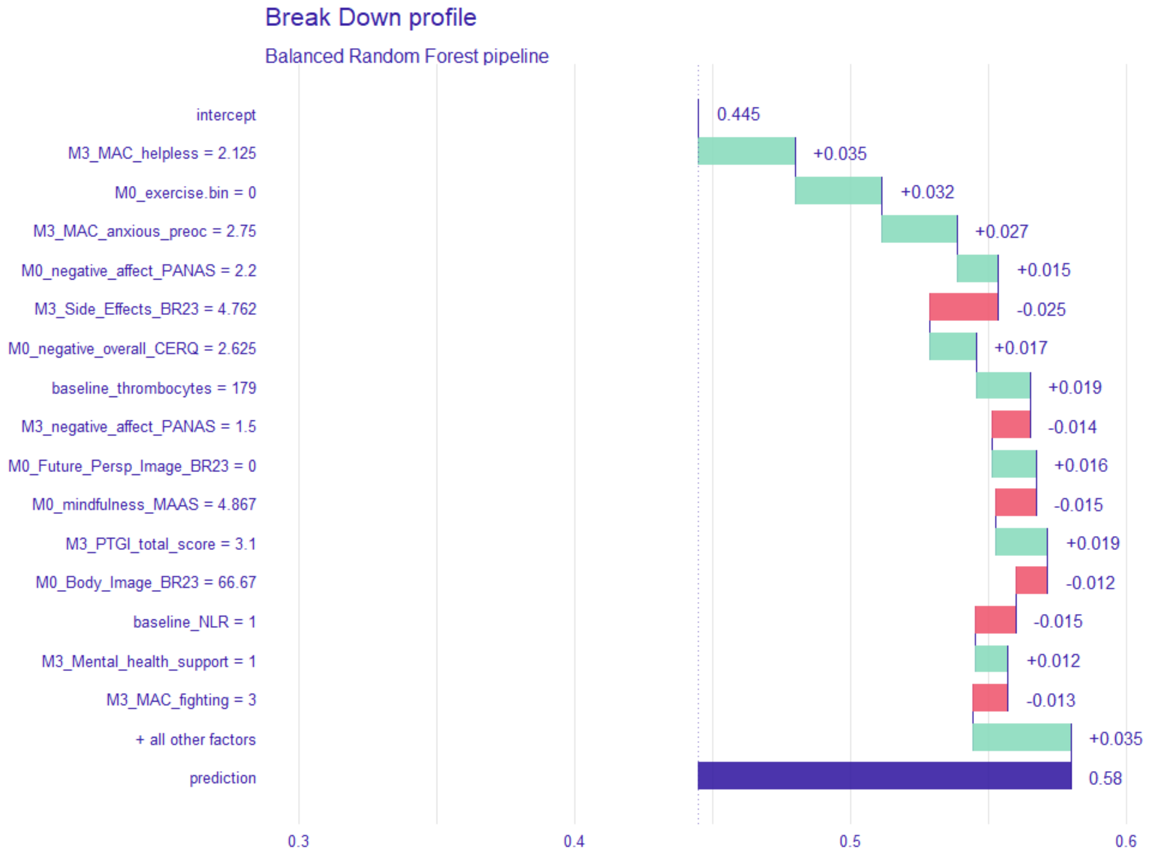


Figure F12. BD profile of a patient who displayed deteriorated mental health status at M12. Prediction probability is shown on the horizontal axis (Stable-Good mental health = 0, Deteriorated Mental Health =1). This patient was given a high probability of membership in the Deteriorated Mental Health class (0.58). Actual patient scores on each predictor variable are shown on the left-hand side. A positive value assigned to a given score (green bars) indicates the degree of its contribution toward a prediction of Deteriorated Mental Health. A negative value (red bars) indicates the degree of a given score's contribution away from a prediction of Deteriorated Mental Health (i.e., increasing the probability of assigning this patient to the Stable-Good Mental Health class).

Figure F13 illustrates the BD profile for a patient who was correctly predicted to have displayed stable-good mental health status over time. Individual patient scores that reduced the probability of deteriorated mental health (i.e., contributed toward a prediction of stable-good mental health) include: moderate coping reactions (anxious preoccupation at M3), the low negative affectivity at M0 and M3, high levels of coping and emotion regulation (M0) combined with very low scores on post traumatic growth (M3). Certain patient scores, however, emerged as risk factors increasing, albeit modestly, the probability of an adverse mental health outcome, namely and relatively low scores on communication cohesion (M3) and family coping reactions (M3). Surprisingly, reporting adherence to a balanced diet combined with relatively high levels of platelet counts at the time of diagnosis also increased the probability of predicting deteriorated mental health.

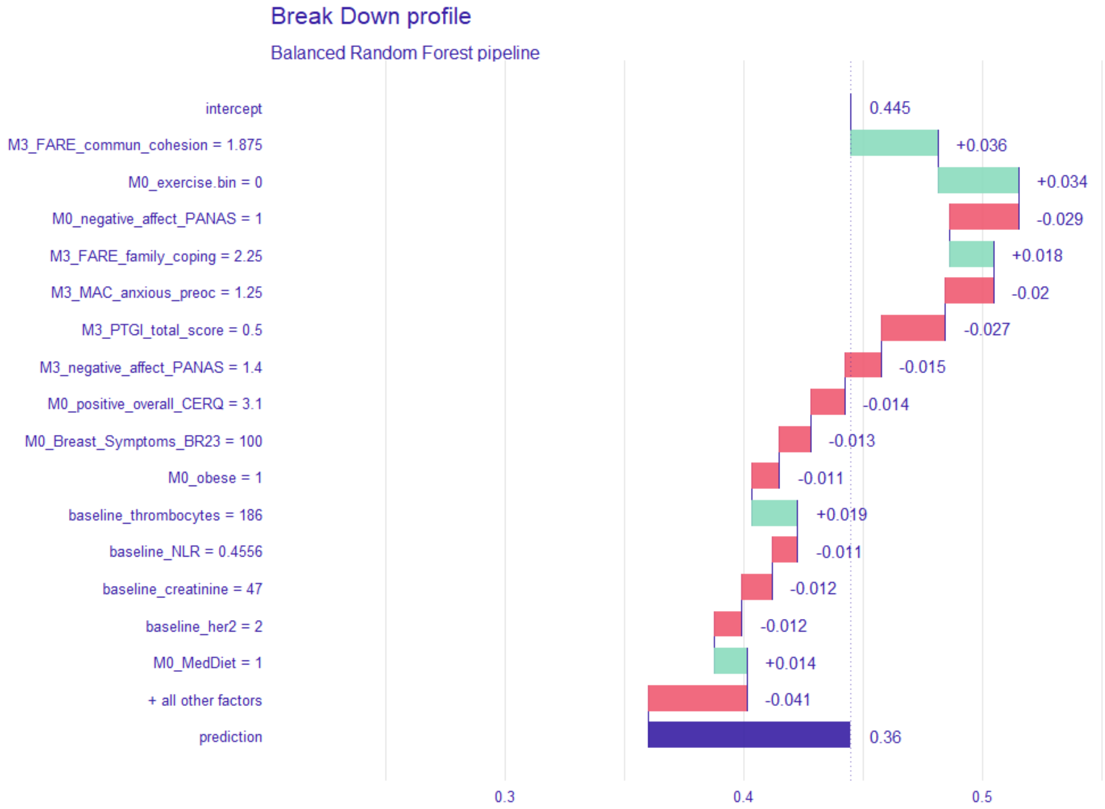


Figure F13. BD profile of a patient who displayed stable good mental health status as indicated by a low probability of membership in the Deteriorated Mental Health class (0.36). All other figure elements are the same as in Figure F12.

1.3.2. Discussion

Transparency and trustworthiness of the ML-based models assisting decision-making processes are critical in the medical field, given the impact on patient safety and social costs that these computational models may cause. On this basis, we employed patient-level interpretation techniques to identify key variables that contribute to the successful predictions for individual patients (Figures F12-F13). By design, we did not include mental health and/or QoL measurements collected during the first three months post-diagnosis so as to enhance model sensitivity toward more clinically useful psychological characteristics and behaviors given the prediction model of one-year mental health deterioration for personalized risk profiles identification.

Pending future validation, these results reflect the impact of each predictor variable to the estimated probability that a given patient is assigned to the low- or high-risk class (i.e., Stable-Good and Deteriorated Mental Health groups, respectively). In principle, these indices of relative variable impact could help clinicians identify patient characteristics which may predispose toward, or have a protective role against, adverse mental health outcomes. If such predictions are validated through programmatic clinical research, they could be used in the future to guide personalized planning of psychological interventions.

1.4. Resilience as Process: Predicting individual differences in the course of psychological adaptation

The goal of these analyses was two-fold: Firstly, to identify temporal patterns of change in well-being (mental health symptoms and global QoL) over the first 18 months post-diagnosis. Secondly, to identify predictors of systematic patterns of change in wellbeing indicators based on medical, biological, lifestyle, sociodemographic, and psychological patient characterises registered soon after diagnosis and close to the onset of cancer treatments.

1.4.1. Dataset description

The data set used to model mental health (MH) trajectories of HADS total scores over 7 measurement points (M0 to M18) comprised 474 patients who had no more than 1 missing value. Imputation, was implemented using the Multiple Imputation through Chained Equations (MICE) algorithm from the R package 'MICE' for 1, 14, 7, 12, 18, 17, and 19 observations, respectively for data at M0, M3, M6, M9, M12, M15, and M18.

For modelling of global QoL trajectories across the 7 measurement points the available data set comprised 472 patients who had no more than 1 missing value. Imputation, was implemented for 9, 15, 8, 13, 21, 17, and 24 observations, respectively for data at M0, M3, M6, M9, M12, M15, and M18.

1.4.2. Unsupervised clustering

1.4.2.1. Trajectory Analysis

The endpoints, MH and QLQ, were treated as continuous variables and unsupervised trajectory clustering was performed by means of a modified k-means algorithm. Specifically, the kmlShape R package was used to cluster individual patient trajectories accounting for the shape of each trajectory using a shape-respecting distance metric. Options regarding the random initialization points and expectation–maximization were kept as default, while the time-slice was set to 0.01. Overall, kmlShape is a variant of k-means where the Fréchet distance, associated with trajectory shape, is used as the distance metric. The Fréchet distance is defined on a continuous interval, so that the real Fréchet distance cannot be obtained in discrete cases, but can be infinitely approximated. In brief, a curve P can be regarded as the mobile trajectory that travels at a speed α. Then, the Fréchet distance between the curve P and another Q considered as a mobile trajectory with speed β, is the smallest possible maximum distance between the two curves after reparameterization of P and Q by α and β, respectively: DistFrechet(P, Q) = $d_{\alpha,\beta}(P, Q)$. With appropriate approximation, this distance can account for the different number or location of measurement points and missing values in patients. In the implementation of kmlShape, Fréchet distance is also used to determine the cluster centers.

A known limitation of k-means algorithm is that the resulting clusters depend on the initial random assignments and, thus, each run with the same number of clusters k, might yield slightly different results. To mitigate this dependence, for each k value, we run the algorithm

10 times with different initial values so as to pick the best result in terms of within-cluster compactness. This was defined as the clustering solution with the lowest within-cluster sum of squares (WSS). The WSS is a commonly used measure of cluster compactness and is defined as the sum of distances between the points and the corresponding centroids for each cluster:

$$WSS = \sum_{i=1}^{N_C} \sum_{\forall x \in C_i} \frac{1}{2n_{C_i}} d\left(x, \bar{x}_{C_i}\right)^2 \tag{F1}$$

Where N_C is the number of clusters, C_i is the i^{th} cluster, $\bar{x}_{C_i}$ is the centroid of the cluster C_i and n_{C_i} is the number of data points in the cluster C_i.

1.4.2.2. Optimal cluster selection

The best clustering solution was chosen based the following criteria: 1) Having a meaningful number of clusters as with too few clusters, data with significant differences may be grouped together while with too many clusters we might overfit the data; 2) Having clustering patterns that are interpretable and clinically meaningful. To aid the selections of the number of clusters the elbow method was also used. The objective function for the elbow method was the total within-cluster sum of squares (WSS) computed for each clustering solution. This is a function monotonically decreasing with increasing number of groups, but often with a turning point (the elbow point) indicating the optimal number of clusters in terms of cluster compactness.

1.4.2.3. Clustering results

Figure F14 depicts the WSS plot for the clustering solutions from 2 to 10 clusters for the studied outcomes. Inspection of the WSS plots suggests an optimal number of 7 (for mental health) and 6 for QoL. However, inspection of the cluster trajectories revealed that solutions with more than 5 clusters included one or more very small subgroups of patients (i.e. with <40 patients) and at least 2 clusters representing very similar time courses. Therefore, we opted for the 5 cluster solutions which represented time courses that have been described previously in longitudinal studies of cancer patients. The corresponding trajectory patterns are presented in Figure F15, depicting individual trajectories belonging to each group along with the mean cluster trajectory.

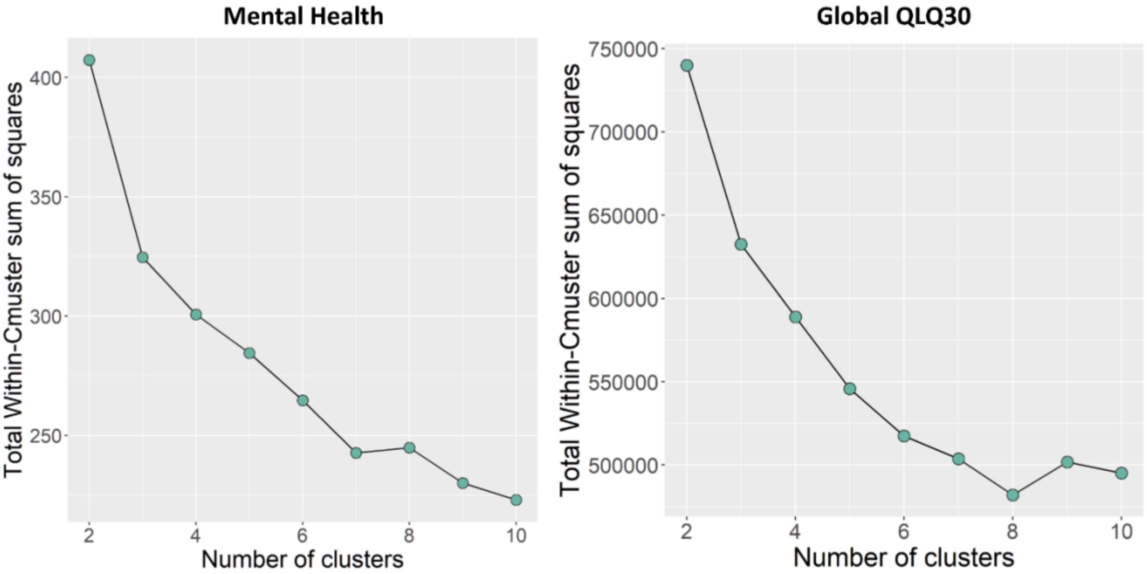


Figure 14. Within-cluster Sum of Squares (WSS) plot for M0-M18 MH and QLQ models considering 2 to 10 clusters.

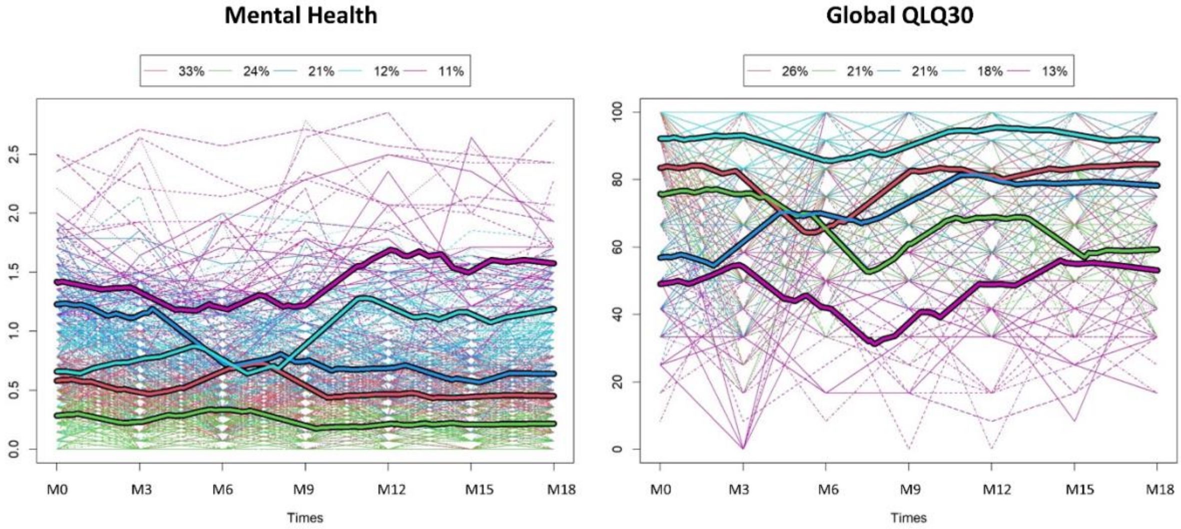


Figure15. Trajectory patterns for M0-M18 MH and QLQ models according to the 5-clusters solution. Thin lines represent individual patient trajectories and bold lines represent the average trajectory for each identified cluster. The proportion of the sample belonging to each cluster is given on top of the corresponding plot.

Regarding overall mental health, the majority of patients displayed stable good mental health (24% of the total data set) or generally low symptom severity with a slight increase at M6 (33%). A sizable proportion of patients (21%) displayed declining symptomatology after M3. The smallest groups comprised patients who demonstrated a notable rise in symptom severity (12%, primarily after M6) and patients who maintained high levels of symptoms throughout the study period (11%).

Regarding global QoL, the majority of patients displayed stable good levels (18% of the total data set) or generally good QoL with a slight decline at M6 (26%). Sizable proportions of patients displayed improving (21%) or worsening QoL after M3 (21%). The smallest group comprised patients who reported poor QoL throughout the study period with further decline between M6 and M9 (13%).

Figure F16 depicts the distribution of the clusters in each clinical center. Cluster size was not identical across centers with HUS contributing the highest number of patients with good progression of MH and global QLQ30 scores compared to other centers.

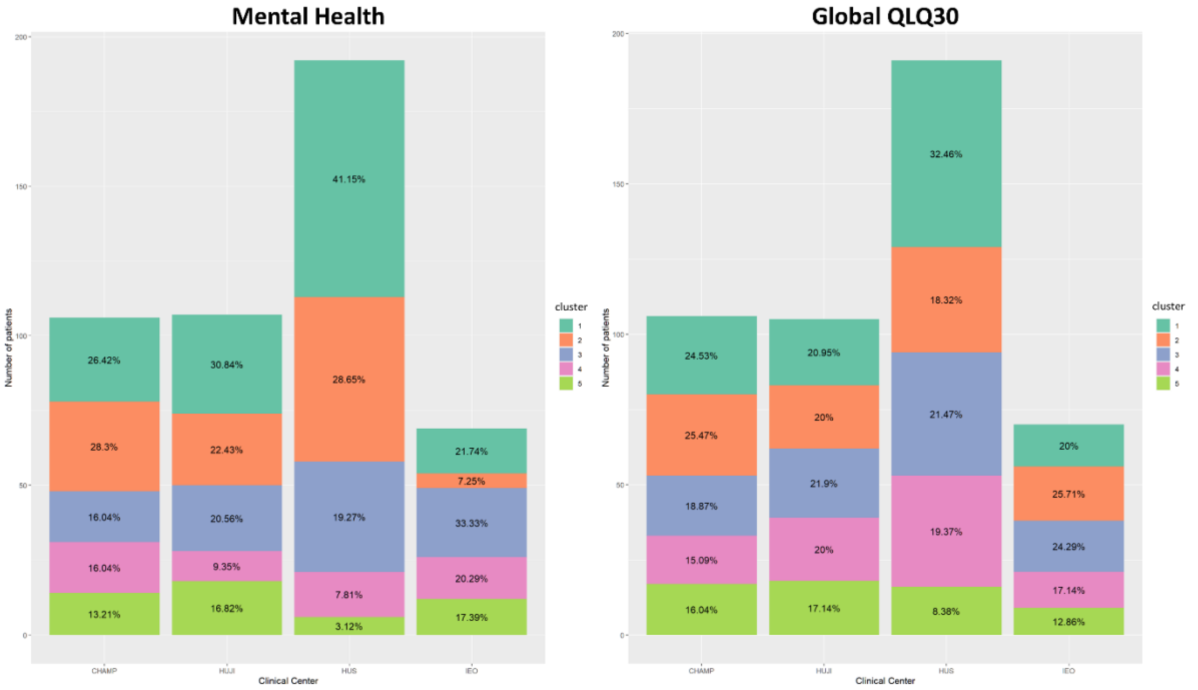
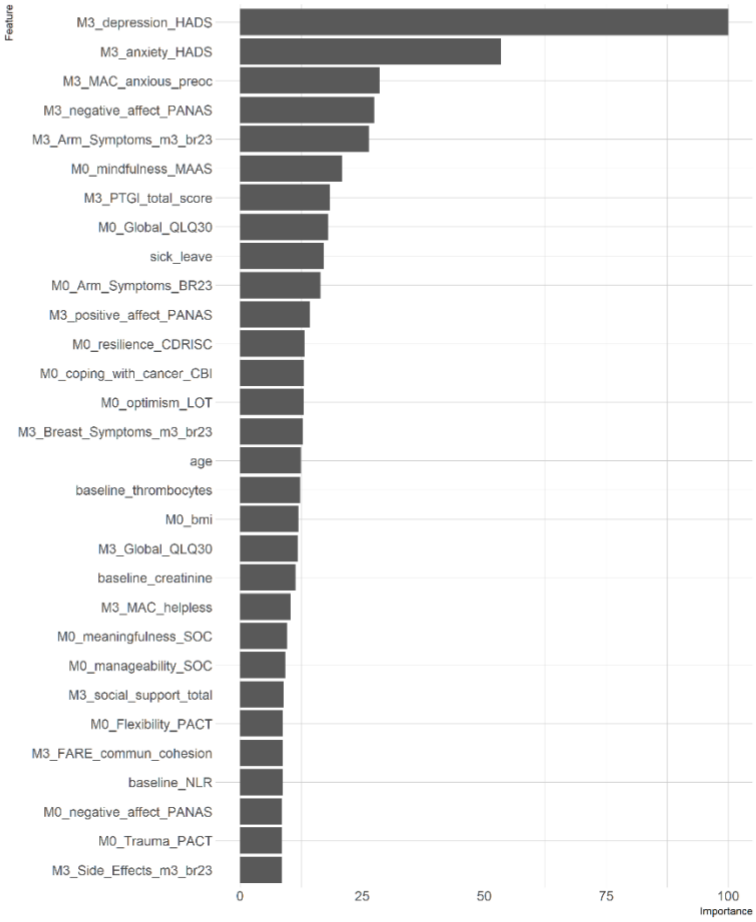


Figure F16. Distribution of clusters across clinical centers.

1.4.3. Predictors of individual differences in resilience as process

Mental health trajectories. In the main modeling scheme, subgroups of patients were formed according to 18-month trajectories as described in Section 1.4.2. The main comparison of interest was between patients who maintained subclinical symptomatology throughout the study period (green and red lines in Figure F15) or initially reported relatively high levels of symptoms with subsequent improvement (blue line) vs patients who either maintained high levels of symptoms throughout or displayed a notable increase in symptomatology during the study period (magenta and cyan lines). In Model 1 predictors included all available variables registered during the first 3 months of the study period (i.e., at M0 and M3, including HADS anxiety and depression scores and EORTC global QoL scores). This model performed modestly as indicated by AUC=0.82, balanced accuracy=72%, Sensitivity=68%, Specificity=76%, and F1=0.54. The complementary model (Model 2) which did not include HADS scores in the set of potential predictors displayed slightly lower performance (AUC=0.81, balanced accuracy=74%, Sensitivity=77%, Specificity=72%, and F1=0.55). As shown in Figure F17, among the top-ranking 30 predictors were symptoms of anxiety and depression and negative affect at M3, physical symptoms throughout the first 3 months post diagnosis, and psychological measures of adaptive processes such as trait resilience, coping strategies, and mindfulness. Biological variables (NLR, creatinine) ranked lowest. With the exception of HADS scores these variables featured among the top-ranking predictors highlighted by Model 2.


Feature
M3_depression_HADS
M3_anxiety_HADS
M3_MAC_anxious_preoc
M3_negative_affect_PANAS
M3_Arm_Symptoms_m3_br23
M0_mindfulness_MAAS
M3_PTGI_total_score
M0_Global_QLQ30
sick_leave
M0_Arm_Symptoms_BR23
M3_positive_affect_PANAS
M0_resilience_CDRISC
M0_coping_with_cancer_CBI
M0_optimism_LOT
M3_Breast_Symptoms_m3_br23
age
baseline_thrombocytes
M0_bmi
M3_Global_QLQ30
baseline_creatinine
M3_MAC_helpless
M0_meaningfulness_SOC
M0_manageability_SOC
M3_social_support_total
M0_Flexibility_PACT
M3_FARE_commun_cohesion
baseline_NLR
M0_negative_affect_PANAS
M0_Trauma_PACT
M3_Side_Effects_m3_br23
0
25
50
75
100
Importance

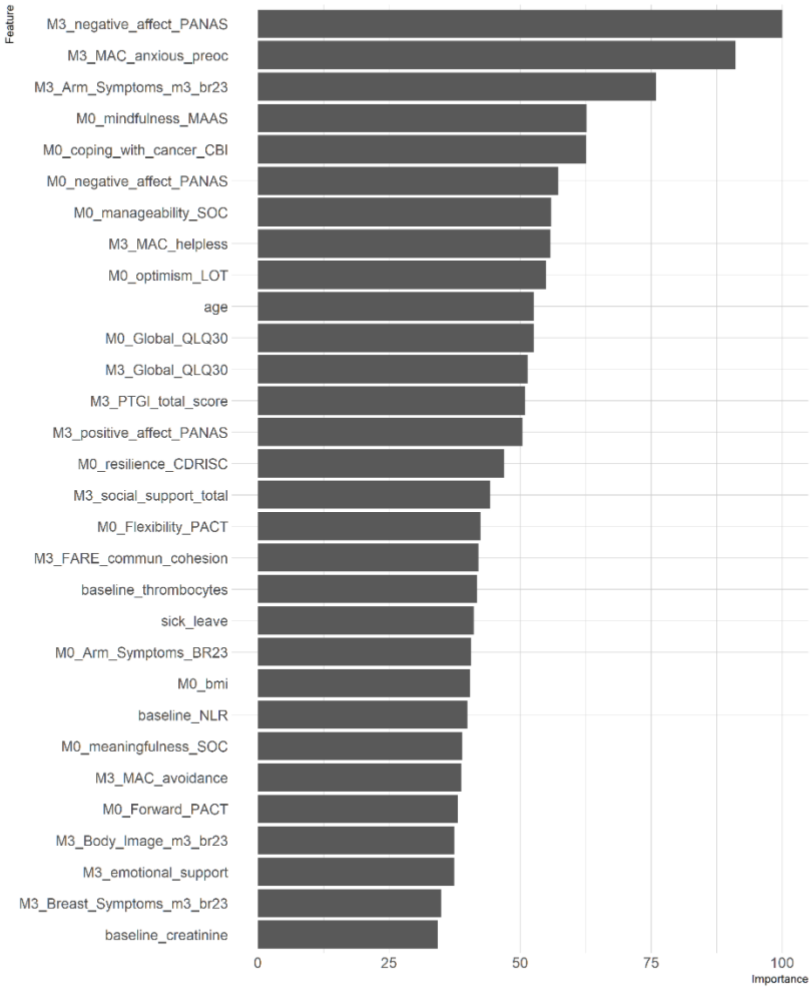


Figure F17. Variable Importance for the top 20 predictors of MH progression over the entire study follow-up period according to Model 1 (upper panel) and Model 2 (lower panel).

In supplementary analyses, we tested the capacity of ML models to predict **future course** of mental health symptomatology (i.e., between **M3 and M18**) based on individual differences on variables obtained at the time of diagnosis (M0). In this framework we obtained patient clusters based on M3-M18 trajectories comprised of 6 time points (see Figure F18).

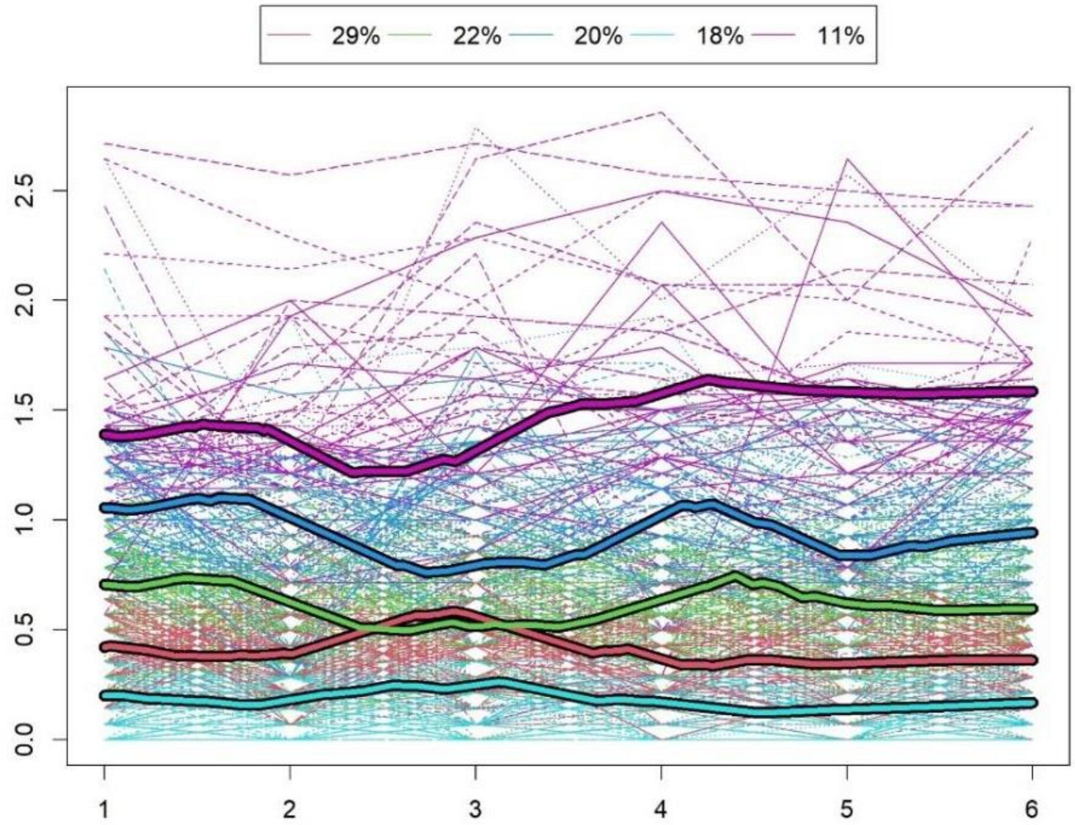


Figure F18. Trajectory patterns for M3-M18 MH model considering the 5-clusters solution. Thin lines represent individual patient trajectories and bold lines represent the average trajectory for each identified cluster. The proportion of the population belonging to each cluster is given on top of the corresponding plot.

Subsequently we performed supervised clustering using the same pipeline as described above. The main comparison of interest was between patients who maintained subclinical symptomatology throughout the study period (cyan and red lines in Figure F18) vs patients who either maintained high levels of symptoms (magenta line) or scored near the clinical threshold on HADS total scores (blue line). In Model 3 predictors included all available variables registered at M0 (i.e., including HADS anxiety and depression scores and EORTC global QoL scores). This model performed fairly well as indicated by AUC=0.89, balanced accuracy=82%, Sensitivity=82%, Specificity=81%, and F1=0.78. The complementary model (Model 4) which did not include HADS scores in the set of potential predictors displayed only slightly lower performance (AUC=0.87, balanced accuracy=77%, Sensitivity=81%, Specificity=74%, and F1=0.73). As shown in Figure F19, among the top-ranking 20 predictors were symptoms of anxiety and depression and negative affect, treatment side effects, NLR, and psychological measures of adaptive processes such as trait resilience, coping strategies, and mindfulness. With the exception of HADS scores these variables featured among the top-ranking predictors highlighted by Model 2 as well.

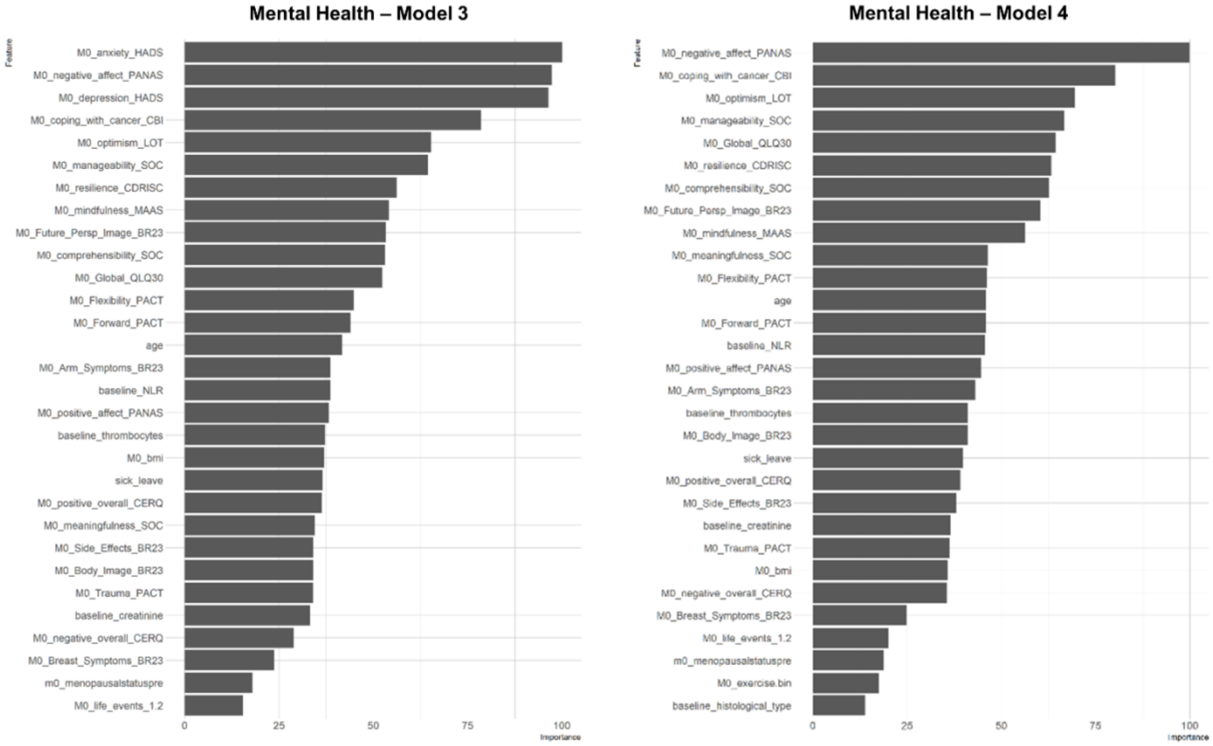


Figure F19. Variable importance for the top 20 predictors of MH progression (Models 3 and 4).

Global QoL trajectories. Prediction models of cluster groups characterizing 18-month trajectories (M0-M18) were associated with overall worse performance than corresponding models for mental health. Advance prediction models for QoL trajectories displayed poorer performance as indicated by balanced accuracy<62% and AUC<0.65.

1.4.4. Discussion

Five distinct trajectories of change over time were identified for mental health and global quality of life. Previous analyses in smaller samples[18, 19] had showed four trajectories. That is, a resilient trajectory (i.e., patients with high levels of well-being across time), a chronic distress (i.e., patients with low levels of well-being across time), a deteriorating as well as a recovering one. Our analyses revealed an additional fifth trajectory. In both mental health and global quality, this trajectory represented patients with initially good levels of well-being which deteriorated around Month 6 (probably near the end of the active treatment phase and the beginning of the re-entry phase), but rapidly returned to the initial good levels. This finding shows that resilience as a process may take different forms and it probably does not refer only to patients with stable high levels of well-being. Also, it is worthy note that, despite the same

[18] Helgeson, V. S., Snyder, P., & Seltman, H. (2004). Psychological and physical adjustment to breast cancer over 4 years: identifying distinct trajectories of change. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, *23*(1), 3–15. https://doi.org/10.1037/0278-6133.23.1.3

[19] Henselmans, I., Helgeson, V. S., Seltman, H., de Vries, J., Sanderman, R., & Ranchor, A. V. (2010). Identification and prediction of distress trajectories in the first year after a breast cancer diagnosis. *Health psychology*, *29*(2), 160–168. https://doi.org/10.1037/a0017806

number and outlook of the trajectories in mental health and global quality of life, their shape was dissimilar, revealing differences in the ways that the different aspects of well-being progress over time.

A series of mostly psychological, but also disease-related (e.g., felt symptoms), medical, and lifestyle factors differentiated the trajectories. Only a few differences were identified in the sets of predictors for mental health and global quality. However, the strength of each predictor was diverse across outcomes. These findings provide critical information regarding the mechanism of patient resilience and adaptation to breast cancer, and they bear significant practical implications.

Finally, it should be pointed out that, despite the efforts to develop and use a single, composite indicator of patients' resilience-as-outcome (or well-being), we finally abandoned this approach. Mainly the different shapes of the trajectories of change over time in psychological symptoms and global health status/quality of life, but also the differences in the final set of predictors for each outcome demonstrated that the use of a single indicator probably is not clinically relevant.

1.5. BOUNCE Decision Support module

The working version of the BOUNCE Decision Support Module incorporates the final models described in Section 1.3. For demonstration purposes, the module uses only the first 15 variables selected (i.e., most highly ranked) by the best-performing prediction model to compute a set of risk scores (and accompanying uncertainty for each patient). There will be one risk score for each of the main outcomes (overall mental health, QoL). Prediction uncertainty will also be provided to the clinician who, as a matter of fact, will be asked to use their clinical judgment in evaluating the need for specific actions on a case-by-case basis.

1.6. Summary

The present document constitutes the final version of D4.3 concerning the computational approaches that have been adopted for modelling resilience status *(resilience as outcome*) through supervised learning techniques, and of resilience as process through unsupervised clustering techniques. Both pooled data across sites and generalizability analyses have been conducted to establish a robust framework that could be exploited to other clinically relevant problems. The present work extends the explainability domain by adopting model-agnostic analysis in our models.

Part 2 (ICCS)

ICCS Models

2.1. Introduction (Aim of the work)

The aim of the present work is to predict the future trajectory of key mental health and quality of life indices for a new patient. The work is divided into two main parts.

Part I1- Trajectory clustering: The first step is to identify clusters of patients that follow similar patterns of change in psychological or behavioral outcomes across multiple time points (trajectories). In the present analysis eight time points are considered (month 0, 3, 6, 9, 12, 15, 18). To this end, latent-class mixed-effects regression analysis has been used. A classical statistical analysis is subsequently performed to identify features that explain class membership.

Part I2- Detecting trajectory clusters: The second step is a classification problem where the outcome variable (classes to predict) are the trajectory clusters identified in Part I1. The aim is twofold: a) to identify features of importance in explaining/predicting each trajectory at each time point, b) to construct models which are able to predict/detect the developmental trajectory for a new patient, as early as possible. To this end supervised learning methods are used.

The key psychological outcomes that have been considered are HADS depression, HADS Anxiety and C30 general health/QoL.

2.2. Methods

2.2.1. Participants included in the analysis

Through BOUNCE infrastructure, records of 790 women diagnosed with breast cancer have been collected. From these patients, 58 did not meet the inclusion criteria (age out of range 40-70 or time interval between diagnosis/surgery and baseline psychological records more than 6 months) and were excluded from the analysis. Participants with missing values from a certain point onward (drop outs) were also excluded (namely patients missing months 3, 6, 9, 12, 15, 18 or months 6, 9, 12, 15, 18 or months 9, 12, 15, 18 or months 12, 15, 18 or months 15, 18). The analysis refers to a subgroup of 513 patients that have HADS Depression and C30 QoL scores at, at least four time points, over the 18 months period. Participants that miss month 0 or have HADS Depression or C30 QoL score at one, two or three time points over the 18 months period have been removed.

2.2.2. Trajectory clustering

2.2.2.1. Regression analysis

A latent-class mixed-effects regression analysis is performed in order to identify sub-groups of patients with distinct trajectory patterns of the psychological variable considered. The analysis is performed using the lcmm package of R. Linear, quadratic and cubic models of the change across time are considered. Models with one to six latent growth classes are fit to the data. Models with different latent processes are also produced. Each model runs several times from different sets of initial values (from a grid of 80 initial values) to avoid convergence to local minima. The number of latent growth classes that best fit the data is assessed by identifying the model with the lowest Akaike information criterion (AIC), Bayesian information criterion (BIC) and sample size–adjusted BIC (SABIC). The average posterior probability of class membership should be above 0.7. Finally, the minimal class size should be at least 5% of the total number of patients.

2.2.2.2. Variable Impact on cluster membership

After identifying the latent class solution that best fits the data, differences among the predicted classes were examined for important covariates and concurrent outcomes *outside the models.* The aim is to assess for differences in demographic and clinical characteristics, symptom severity scores, functioning scales and other psychological scales and outcomes of interest among the growth mixture model latent classes at each time point, and find determinants that explain the divergence between latent trajectory clusters. This is done here using classical statistical analysis, e.g. Kruskal-Wallis test, chi-squared test, Repeated Anova, post hoc analysis and time series plots. The results are in the process of analysis and clinical interpretation. Indicative results are provided here.

2.2.3. ML approach for trajectory cluster detection

2.2.3.1. Preprocessing of predictor data and Handling of missing values

The dataset is screened for highly redundant variables, i.e. variables that carry the same or nearly the same information, that have derived after transformation. For example, if A is a categorical variable of N levels (e.g. 'marital status'), a redundant variable may be the one that derives after the merging of the N levels into two (e.g. 'alone'). Only one relevant variable is kept. Furthermore, competing predictors are combined to form new variables. The dataset is subsequently checked for the existence of very high correlations among predictors (>0.8).

Certain machine learning algorithms cannot operate on categorical data directly. They require all input and output variables to be numeric (e.g. XGBoost, svm etc). An integer encoding had already been applied to the categorical variables of the dataset. However, for categorical variables with more than two categories, such an action imposes an ordinal relationship where no such relationship exist. To overcome this problem, we apply the dummy encoding technique, which represents N categories/levels with N-1 binary variables (the method avoids redundancy).

Predictors that only have a single unique value ("zero-variance predictor") or a number of unique values but with the second more frequent value occurring with a very low frequency ("near-zero-variance predictor") are removed. The reason is to avoid unstable fitting of the classifier, or an undue influence that few samples may have on the final classification model. Pre-processing is performed in R using the caret package.

Missing values may occur due to skip pattern in survey and data collection design, e.g. certain questions are only asked to respondents who have given a certain answer to a previous question. This type of missing is treated before applying any imputation technique. Patients and predictors with high percentage of missing values (>30%) are removed. For the results presented here single imputation has been applied using missForest package of R (Nonparametric Missing Value Imputation using Random Forest). The developed framework also supports multiple imputation with mice and kml packages of R. Next step planning includes the examination of the effect of different/multiple imputations on the classification results. Multiple imputation allows for missing rate of this magnitude [I1]. Furthermore, missForest can successfully handle missing values up to 30%, in datasets including different types of variables [I2].

2.2.3.2. Classifier training - Feature selection

Model parameters are tuned using grid search with 4-fold cross validation. Class imbalance was handled using smote, down or up subsampling, performed inside the cv resampling. Downsampling was chosen for the final models. A number of classifiers was initially tested (random forest, xgboost, partial least squares (pls), naïve bayes, etc.). Random forest, consistently showed good performance and was finally chosen for the development of the models. Feature selection was performed based on recursive feature elimination (RFE). Random forest specific variable importance was utilized during the RFE procedure. For random forest classifiers, there is a plateau of good performance after a specific number of features. The optimal number of predictors is chosen by taken into account this performance profile. We pick a small subset size that yields a performance score within tolerance of the best (but more complex) model, without sacrificing too much performance. It is noted that feature selection is not a deterministic procedure and results may vary due to differences in resampling, splitting and, overall, due to the stochastic nature of the algorithms used. RFE is repeated a number of times to evaluate the variability of the selected features. The most frequently selected features are finally chosen. The caret package of R is used.

2.2.3.3. Performance evaluation

Classifier performance is evaluated by means of nested cross validation. The inner loop of nested cv is responsible for model selection/hyperparameter tuning (validation set), while the outer loop is responsible for error estimation (test set). A stratified, equally-sized four-fold resampling is used for the inner and outer loop. Model parameters are tuned using grid search against AUROC (area under the ROC curve) metric with 4-fold cross validation.

2.3. Results

2.3.1. Trajectory clusters

2.3.1.1. HADS Depression: M18-Trajectory model

Trajectory clustering is applied to imputed dataset. For depression, the model that best describes the data is a quadratic one with four latent classes (Fig. I1) (lowest BIC). Figure I2 shows the resulting trajectories for the model and the actual measurements of the patients that compose each class.

The largest trajectory group accounted for 56% of the patients (labeled ‘Low’) and was composed of patients who generally had relatively low levels of depression throughout the one-year period. The trajectory labeled ‘High’, estimated to account for 27% of the patients, reported initial moderate/high depression score that remained moderate/high over the one year. The group labeled ‘Decreasing’, accounting for 11% of the patients, began with a moderate/high depression score that improved over time. The final group, labeled ‘Increasing’ and accounting for 6% of the population, started with an initial low score but their score varied substantially during the observation period and are characterized by an increasing mean trajectory.

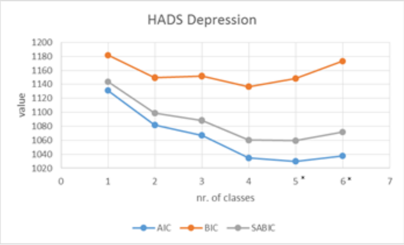
Also shown in Figure I2 are 95% confidence intervals around each trajectory. The fact that the confidence intervals are tight around each trajectory indicates the adequacy of the model.

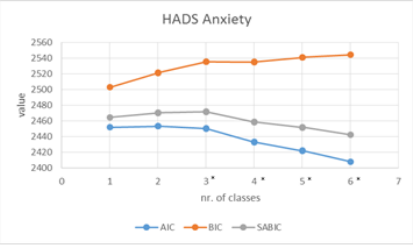
2.3.1.2. HADS Anxiety: M18-Trajectory model

Results suggest (Fig. I1) that the null model (no latent clusters) is the most adequate to describe HADS Anxiety longitudinal data. More precisely, BIC metric is minimum for the null models, whereas the suggested best model by AIC/SABIC contains clusters with too few patients (2% or below) and are characterized wide confidence intervals around the mean trajectories.

2.3.1.3. C30 Global health status/QoL: M18-Trajectory model

For C30 *Global quality of life (QoL) variable*, the best model is an unconditional one with quadratic time and three classes (Fig. I1). Figure I3 shows the resulting trajectories for the three-group model and the actual measurements of the patients that compose each class. We can identify a high increasing class (16%) and one stable mean trajectory at moderate (77%) levels of Qol. One class evidenced a decreasing trajectory at low QoL levels (7%).





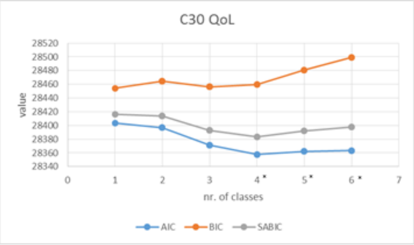


Figure I1. Bayesian information criterion (BIC), Akaike information criterion (AIC) and sample size adjusted Bayesian information criteria (SABIC) values for depression, anxiety and QoL trajectory models as a function of number of latent classes considered. The models with one asterisk indicate that the minimal class size was below 5%.

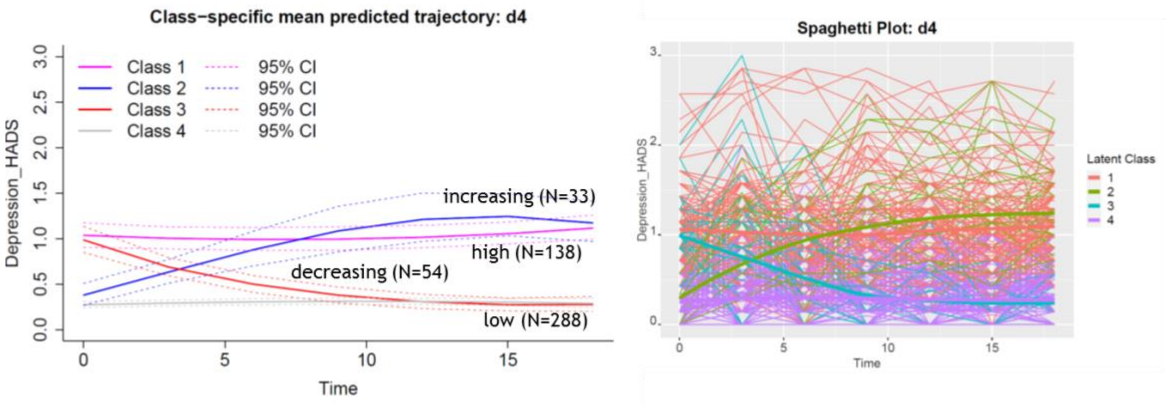


Figure I2: **Left:** Mean Predicted Trajectories over Time (in months) in the four-class linear mixed model. **Right:** spaghetti plot (each thin line connects the responses for the same patient over time) with the mean HADS Depression score at each measurement point for the four classes identified by the latent class model.

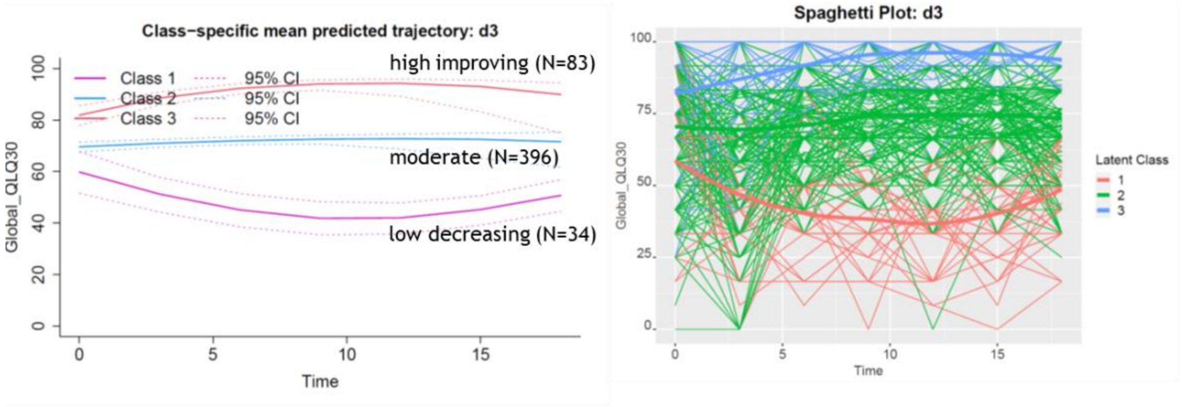


Figure I3: **Left:** Mean Predicted trajectories over Time (in months) in the three-class linear mixed model. **Right:** spaghetti plot (each thin line connects the responses for the same patient over time) with the mean QoL score at each measurement point for the three classes identified by the latent class model.

2.3.2. Association between depression and QoL trajectory clusters

Pearson's chi-squared test showed the classification of *depression and global quality of life* trajectories as significantly associated with each other (χ^2 (6, *N* = 513) = 75.5, *p* <.001). Indicative associations are reported below:

- The vast majority of patients in the 'high improving' QoL group were assigned to the 'low' or 'decreasing' depression group (Fig. I4).
- 'Low decreasing' QoL group are mostly assigned to 'increasing' and 'high' depression groups (Fig. I4).

Taking also into consideration the time course of depression of the QoL trajectory classes (Fig. I4), we can conclude that the 'high improving' QoL group is the most resilient group of patients in terms of both QoL and depression trajectory (significantly lower depression levels than all other groups at all time points based on post hoc analysis). Furthermore, the 'low decreasing' QoL group seems to be the less resilient group of patients in terms of both QoL and depression trajectory (significantly higher depression levels than all other groups based on post hoc analysis).

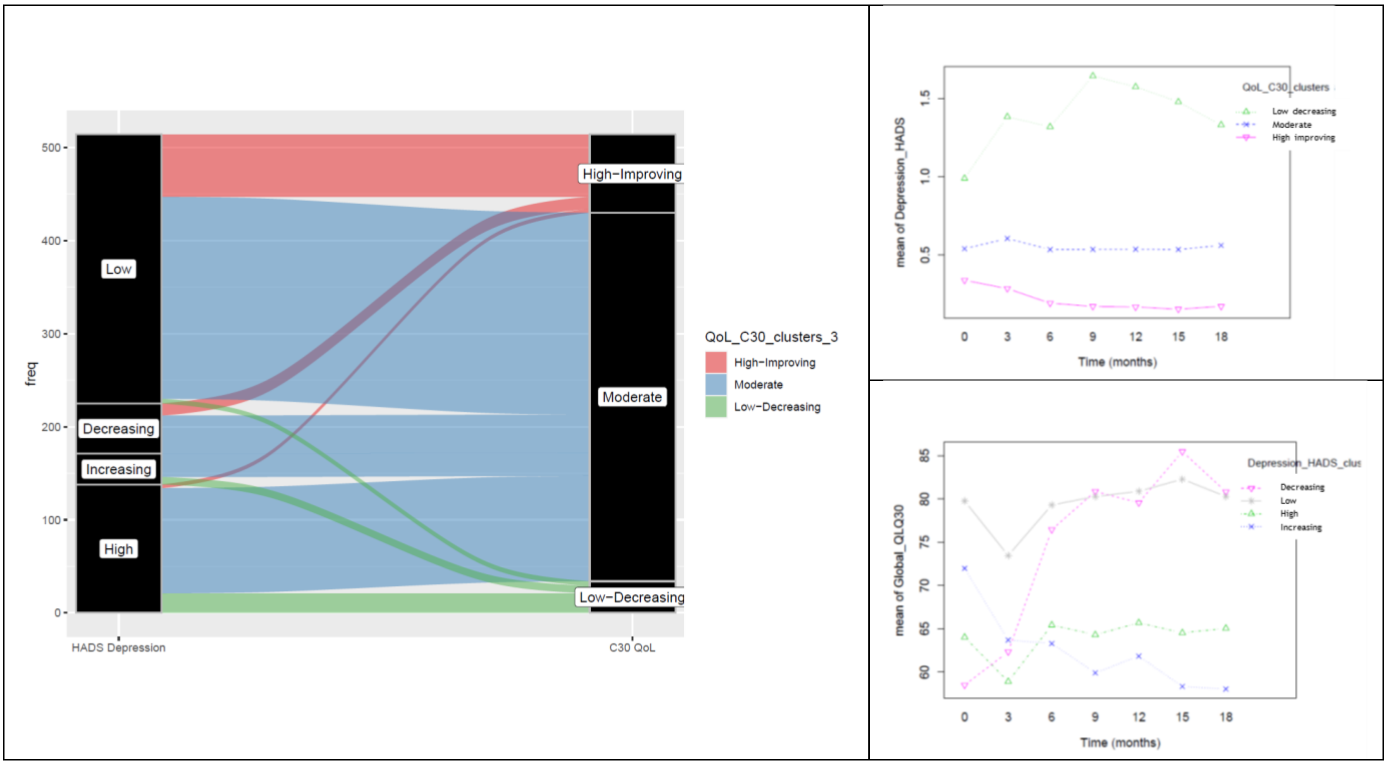
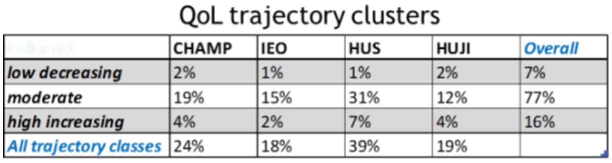


Figure I4. Left: Alluvial diagram between depression and QoL trajectory clusters. **Right up:** Time course of the mean value of depression for each QoL trajectory cluster. **Right down:** Time course of the mean value of QoL for each depression trajectory cluster

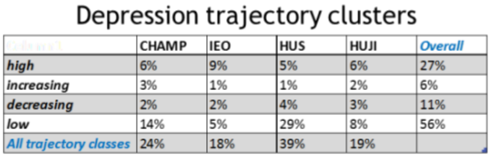
2.3.3. Class distribution in each clinical site

The distribution of the depression and QoL trajectory classes between the clinical sites is reported in Figure I5. Chi-square test of independence showed a significant association between depression trajectories and clinical sites (χ^2 (9, *N* = 513) = 76.94, *p*<.001). In the case of depression, the vast majority of patients in each clinical site is assigned to the 'low' trajectory class, with the exception of IEO where the majority of patients is assigned to the 'high' depression group. Overall, IEO is under-represented in the 'low' depression class, and over-represented in increasing and high trajectory classes, whereas the opposite is the case for HUS. Contrary to depression, the distribution of QoL trajectories is pretty uniform between the clinical sites (χ^2 (6, *N* = 513) = 12.40, p=.054).



QoL trajectory clusters

Column1	CHAMP	IEO	HUS	HUJI	*Overall*
low decreasing	2%	1%	1%	2%	7%
moderate	19%	15%	31%	12%	77%
high increasing	4%	2%	7%	4%	16%
All trajectory classes	24%	18%	39%	19%	



Depression trajectory clusters

Column1	CHAMP	IEO	HUS	HUJI	*Overall*
high	6%	9%	5%	6%	27%
increasing	3%	1%	1%	2%	6%
decreasing	2%	2%	4%	3%	11%
low	14%	5%	29%	8%	56%
All trajectory classes	24%	18%	39%	19%	

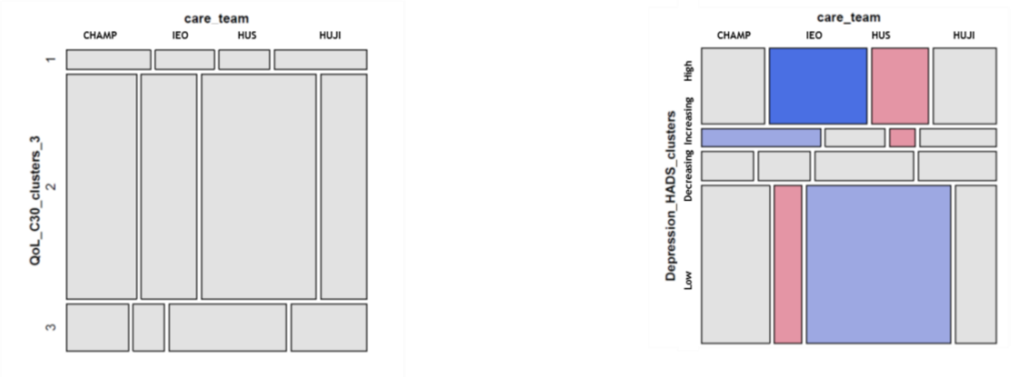


Figure I5. Class distribution of depression and QoL trajectory classes in each clinical site. Blue color indicates that the observed value is higher than the expected value if the data were random. Red color specifies that the observed value is lower than the expected value if the data were random

2.3.4. Variable Impact on cluster membership

Indicative observations are given below for the aforementioned clusters.

2.3.4.1. HADS Depression trajectory clusters

The time course of the mean value of selected scales for each trajectory cluster are presented in Fig. I6. Based on the mean/median scores of most psychological scales, a progressive decrease in emotional wellbeing or Quality of life is evident from low to high trajectories (low-increasing-decreasing-high) at baseline. However the differences are not always statistically significant between the different pairs of trajectory clusters. HADS scales, scales that assess emotional wellbeing (namely, emotional functioning and distress thermometer), PANAS negative affect, CBI coping with cancer, CDRISC resilience, C30 global Qol and self-efficacy are among the variables that significantly differ between the 'low' and 'high' trajectories from baseline up to M18. All functioning scales of C30/B23 questionnaires, such as social functioning, future perspective and symptom scales, such as fatigue, pain etc., also significantly differ between these trajectories during the 18 months period. At M0 mean values of optimism (LOT), manageability (SOC) and meaningfulness (SOC) are significantly higher for 'Low' group.

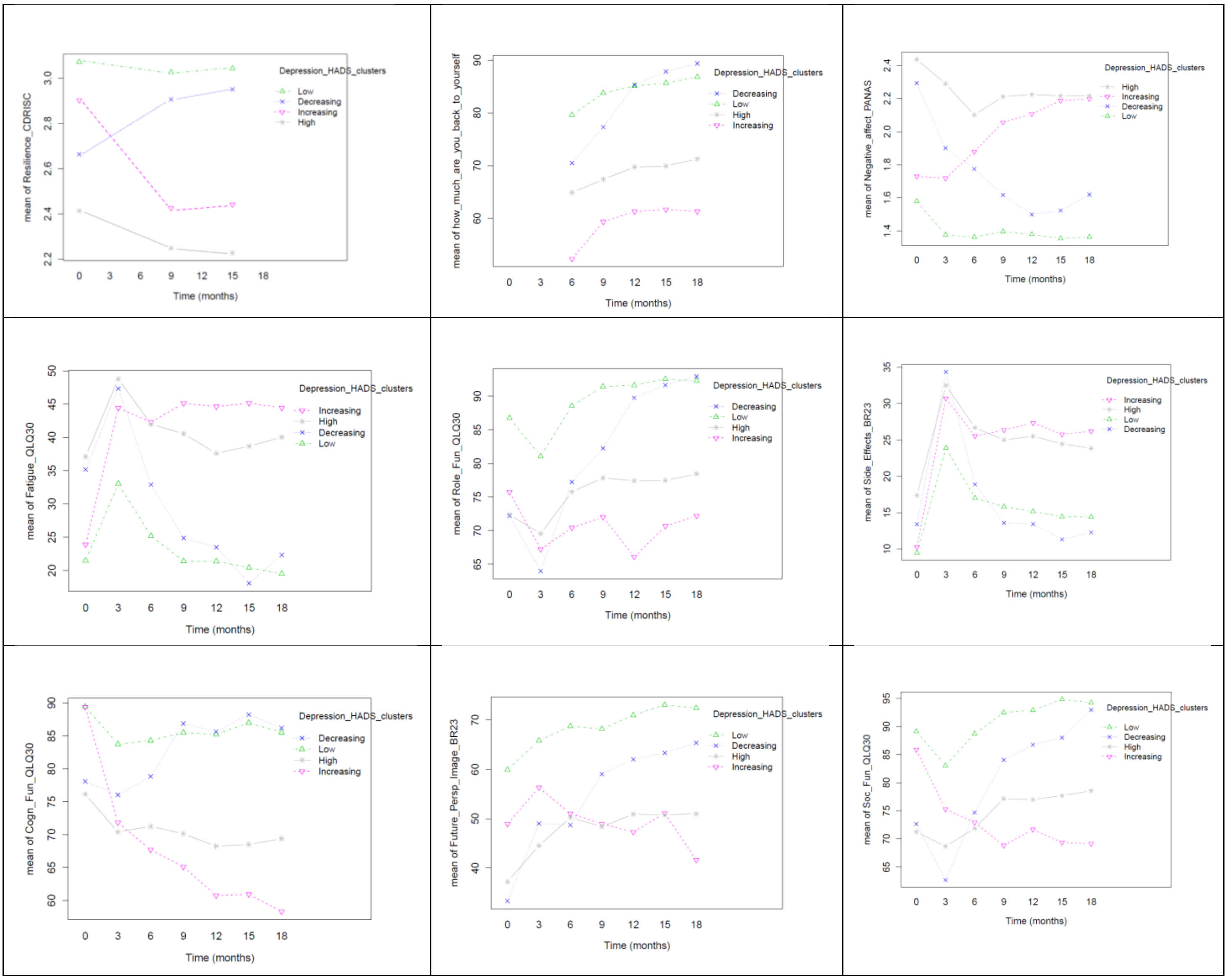


Figure I6: Time course of the mean value of selected scales for each HADS Depression trajectory cluster

2.3.4.2. C30 Global health status/QoL trajectory clusters

Figure I7 depicts, the time course of indicative psychological, symptom and functional scales for each trajectory class. We observe that resilience CDRISC is significantly higher in the high improving group, whereas a statistically significant decline in its value is observed for the 'low decreasing group'. The 'Low decreasing' group, the less resilient group of patents in terms of QoL and based on resilience score, is characterize by prolonged symptoms and significant decrease in function that stays low throughout the 1.5 year. It is characterized by significantly lower scores of 'How much do you think your treatment can help your illness?' and by significantly worse coping with cancer (CBI), manageability (SOC), negative affect (PANAS), fatigue and emotional wellbeing (depression, anxiety, emotional function, cognitive function), illness consequences (IPQ), helpless (MAC) and anxious preoccupation (MAC), post-traumatic stress symptoms (PCL) and others.

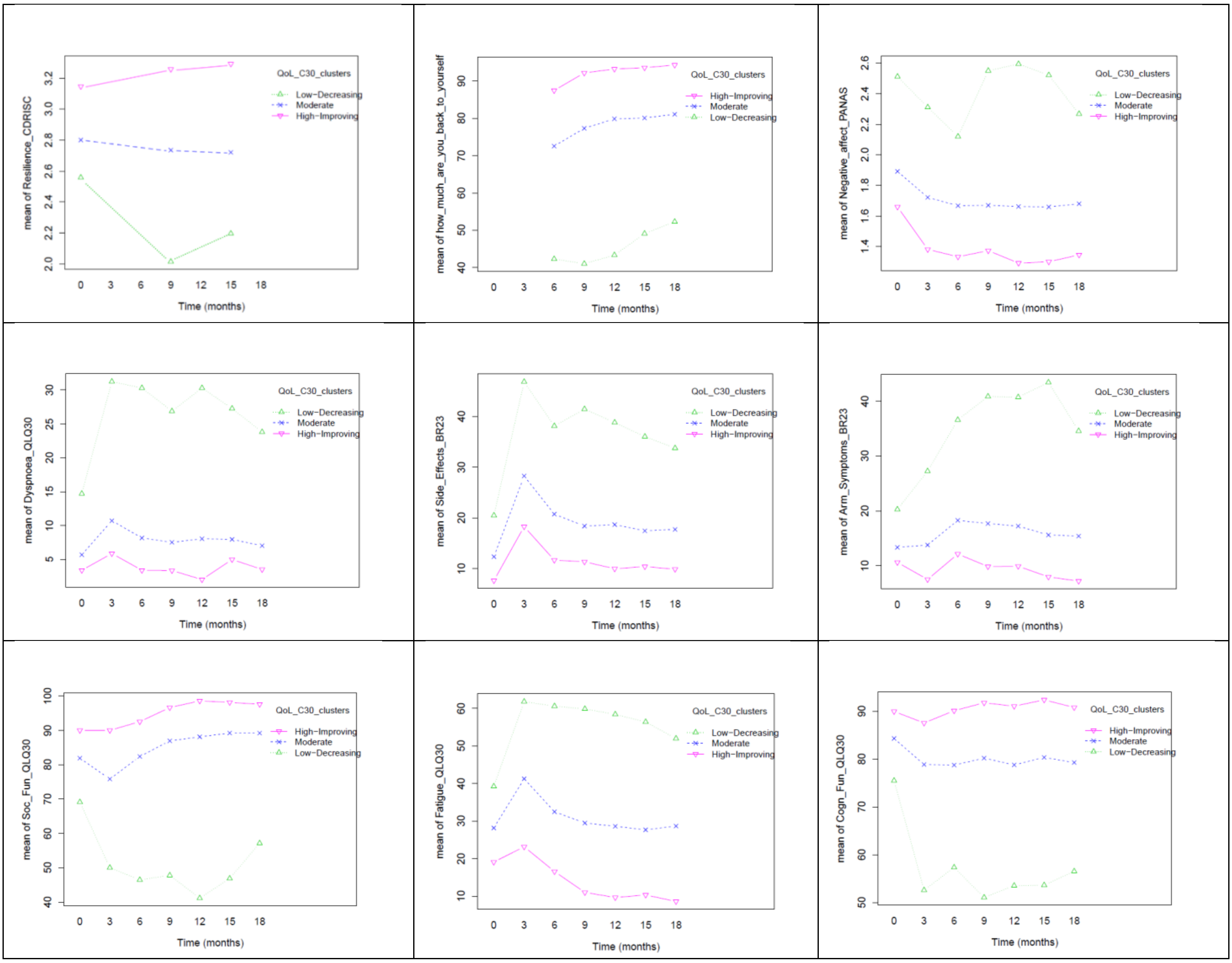


Figure I7: Time course of the mean value of selected scales for each C30 QoL trajectory cluster

2.3.5. ML approach

2.3.5.1. Building a classifier for detecting patients with poor and good depression trajectory

The aim is to build a binary classifier able to detect which patients belong to the 'poor' or 'good' mean trajectory during the 18 months period from baseline. The classifier is trained on 513 patients. The two classes considered emerge from the grouping of the four depression clusters identified in the first part of the analysis (Fig. I8). The 'good' trajectory class comprise the decreasing and low depression trajectory groups and the 'poor' trajectory class comprise the increasing and high depression trajectory groups. Predictors considered are psychometric scales, sociodemographic, clinical, medical and treatment variables from baseline or baseline and M3. The number of predictors is 96 from M0 and 169 from M0 and M3. M6 is also considered to evaluate performance improvement. A random forest classifier was chosen for building the final models.

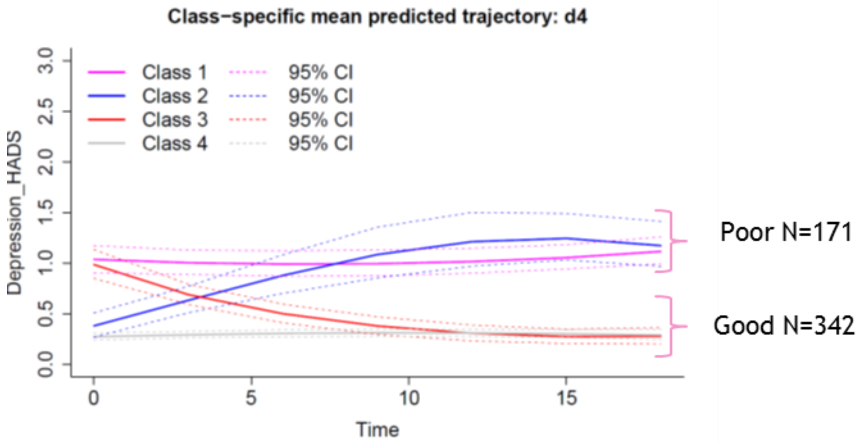


Figure I8: The considered classes for the binary classification. Mean Predicted Trajectories over Time (in months) in the four-class linear mixed model that best described depression. The ‘good’ trajectory class comprise the decreasing and low depression trajectory groups and the ‘poor’ trajectory class comprise the increasing and high depression trajectory groups.

The distribution of the trajectory classes between the clinical sites is reported in Table I1.

Table I1: ***Class distribution*** **(in** ***percentage) in each clinical site (N=513)***

	CHAMP	**IEO**	**HUS**	**HUJI**	
Good	16%	7%	33%	11%	67%
Poor	9%	11%	6%	8%	33%
	24%	18%	39%	19%	

Table I2: Performance of Random Forest Classifier based on nested cross validation with four external folds. Average values with standard deviations are given.

Predictors	**All M0**	**All M0 & M3**	**All M0 & M3 except depression at M3**	**All M0, M3 & M6**	**All M0, M3 & M6 except depression at M3 and M6**
Sensitivity	0.78±0.05	0.82±0.06	0.81±0.04	0.83±0.06	0.82±0.04
Specificity	0.75±0.08	0.79±0.04	0.76±0.06	0.79±0.04	0.77±0.04
ROC	0.84±0.06	0.88±0.05	0.86±0.05	0.89±0.03	0.88±0.04
Balanced Accuracy	0.76±0.07	0.80±0.04	0.79±0.05	0.81±0.03	0.80±0.03
F1	0.68±0.08	0.73±0.05	0.71±0.06	0.74±0.03	0.72±0.03
Pos Pred Value	0.61±0.09	0.66±0.05	0.64±0.07	0.67±0.04	0.64±0.04
Neg Pred Value	0.87±0.04	0.90±0.03	0.89±0.03	0.90±0.03	0.90±0.02

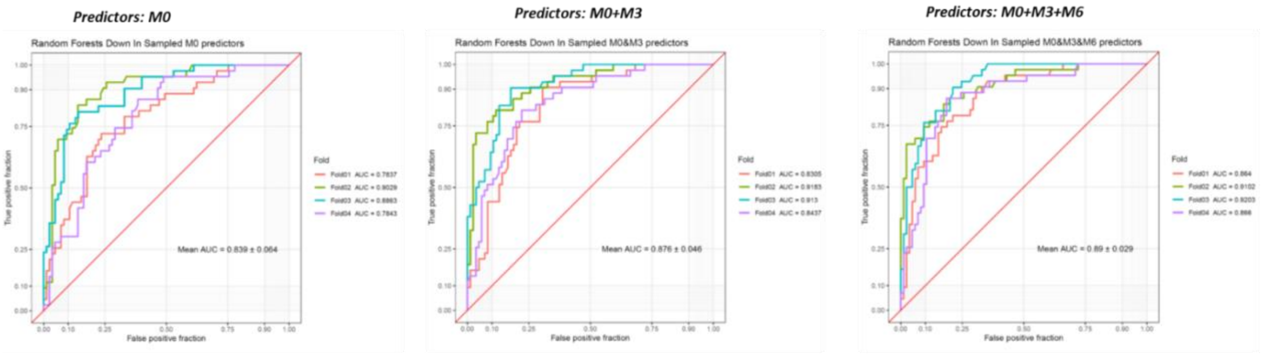


Figure I9: ROC curves for each outer fold of the nested cross validation. Model with all the predictors at M0. Model with all the predictors at M0 and M3. Model with all the predictors at M0, M3 and M6.

A random forest classifier build on M0 predictors

The performance evaluation metrics are depicted in Table I2 & Fig. I9. The performance of the classifier, based on nested cross validation with four external folds, is good (average ROC=0.84, Sensitivity=78%, Specificity=75%) suggesting that a classifier that detects patients with increasing mean depression trajectory can be built at baseline.

Performance profile across different subset sizes for an indicative run of RFE is depicted in Figure I10. We observe that after 10 variables a plateau is reached in model performance. Overall, the results between repetitions of the RFE workflow are consistent. In 70% of the repetitions, the following subset was selected (10 variables):

anxiety HADS (M0), depression HADS (M0), catastrophizing CERQ (M0), coping with cancer CBI (M0), fear of recurrence FCRI (M0), manageability SOC (M0), optimism LOT (M0), resilience CDRISC (M0), negative affect PANAS (M0), self-efficacy (M0)

In 20% of the repetitions two additional variables were also selected: pain QLQ30 (M0), meaningfulness SOC (M0).

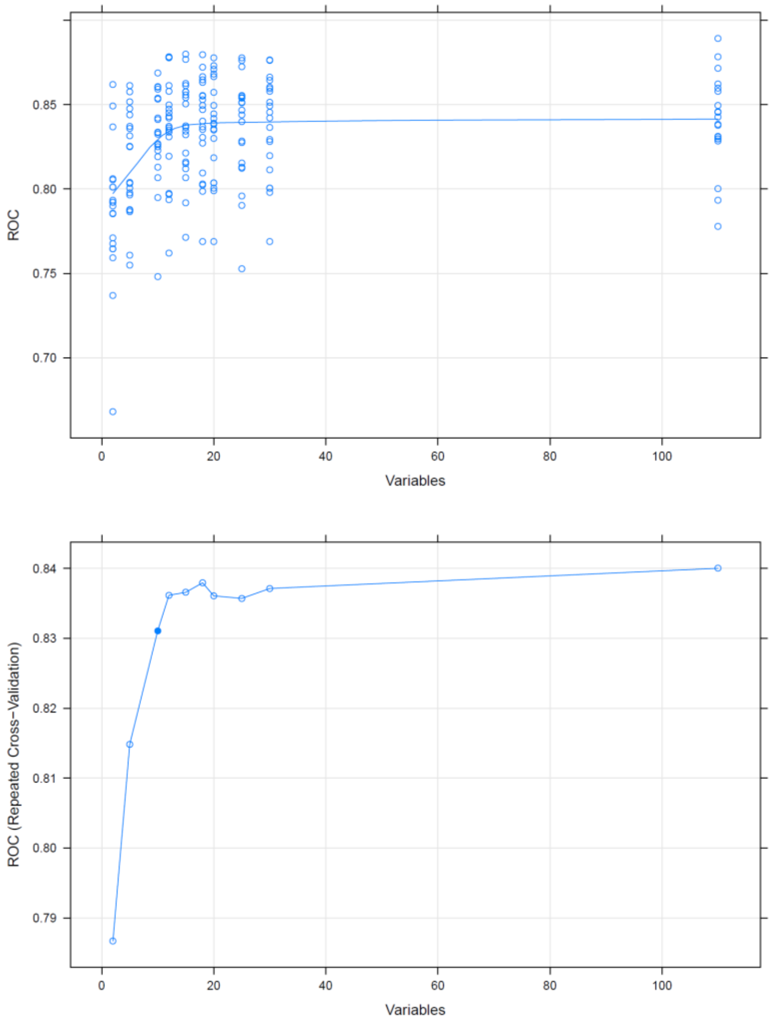


Figure I10. Performance profile of a random forest classifier across different subset sizes. Predictors from M0. Top panel: ROC cv estimates (4 folds repeated 5 times). Bottom panel: average ROC value as a function of the number of features.

A random forest classifier build on M0 and M3 predictors

The performance evaluation metrics are depicted in Table I2 & Figure I9. The performance of the classifier, based on nested cross validation with four external folds, further improves when predictors at M3 are included. Specifically average area under ROC value increases to 0.88, sensitivity to 82% and specificity to 79%. When depression at M3 is omitted the area under ROC value drops to 0.86. Inclusions of M6 variables results in small improvement in performance.

Performance profile across different subset sizes for an indicative run of RFE is depicted in Fig. I11. We observe that after 10 variables a plateau is reached in model performance. Overall, the results between repetitions of the RFE workflow are consistent. In 70% of the repetitions, the following subset was selected (10 variables):

anxiety HADS (M3), depression HADS (M0), depression HADS (M3), coping with cancer CBI (M0), manageability SOC (M0), optimism LOT (M0), resilience CDRISC (M0), anxious preoccupation MAC (M3), negative affect PANAS (M0), negative affect PANAS (M3).

In 20% of the repetitions three additional variables were also selected: MAC helpless (M3), MAC avoidance (M3) and anxiety HADS (M0).

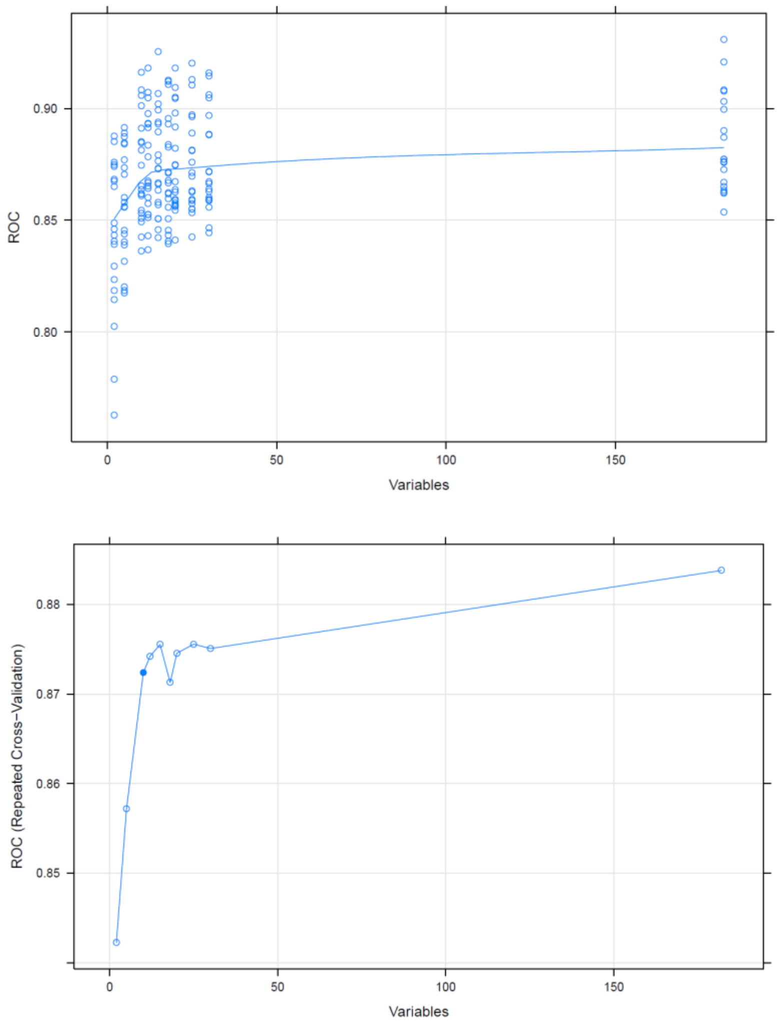


Figure I11. Performance profile of a random forest classifier across different subset sizes. Predictors from M0 and M3. Top panel: ROC cv estimates (4 folds repeated 5 times). Bottom panel: average ROC value as a function of the number of features.

2.3.5.2. Building a classifier for detecting patients with low decreasing QoL trajectory

The aim is to build a binary classifier able to detect patients at risk of low QoL trajectory during the 18 month period from baseline. The positive class is the low decreasing QoL trajectory group identified in the first part of the analysis (Fig. I12). The negative class emerges from the grouping of the high increasing and moderate QoL trajectory, as depicted in Fig. I12. It is noted that the specific group of patients are the least resilient group based on both QoL and depression trajectories. The vast majority of these patients also belong to the "Poor" depression trajectory class and their mean depression levels are relatively high throughout the 18 months after diagnosis (Fig. I4). The classifier is trained on 513 patients. Predictors considered are psychometric scales, sociodemographic, clinical, medical and treatment variables from baseline or baseline and M3. M6 is also considered to evaluate performance improvement. The number of predictors is 96 from M0 and 169 from M0 and M3. A random forest classifier was chosen for building the final models.

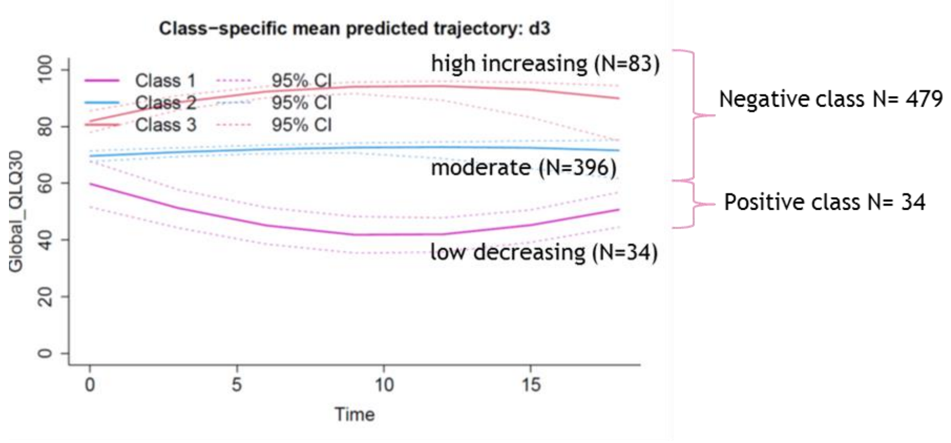


Figure I12: The considered classes

The distribution of the trajectory classes between the clinical sites is reported in Table I3.

Table I3. *Class distribution* (in *percentage) in each clinical site (N=513)*

	CHAMP	IEO	HUS	HUJI	
low decreasing	2%	1%	1%	2%	7%
moderate/high	22%	17%	37%	17%	93%
	24%	18%	39%	19%	

A random forest classifier build on M0 & M3 predictors

The performance evaluation metrics are depicted in Table I4 & Fig. I13. When only M0 variables are considered the performance is unstable and the mean average ROC value is acceptable but low (ROC<0.8). The performance improves considerably when M0 and M3 predictors are considered. Specifically average AUC under ROC curve increases to 0.89, sensitivity to 77% and specificity to 81%. However, the performance is quite unstable between folds (relatively large standard deviation). A reason is the very small size of the positive class (N=34). Inclusion of M6 variables further improves performance. Results suggest that a classifier that detects patients with low decreasing QoL trajectory can be built at month 3.

Table I4: Performance of Random Forest Classifier based on nested cross validation with four external folds. Average values with standard deviations are given.

Predictors	All M0	All M0 & M3	All M0 & M3 except depression at M3	All M0, M3 & M6	All M0, M3 & M6 except depression at M3 and M6
Sensitivity	0.69±0.24	0.77±0.20	0.79±0.25	0.82±0.21	0.92±0.12
Specificity	0.76±0.13	0.81±0.10	0.78±0.05	0.87±0.07	0.82±0.04
ROC	0.79±0.11	0.89±0.06	0.87±0.08	0.94±0.05	0.93±0.06
Balanced Accuracy	0.72±0.07	0.79±0.06	0.79±0.11	0.85±0.08	0.87±0.05
F1	0.27±0.12	0.35±0.11	0.31±0.07	0.45±0.12	0.41±0.08
Pos Pred Value	0.18±0.06	0.24±0.11	0.20±0.05	0.33±0.12	0.27±0.06
Neg Pred Value	0.97±0.03	0.98±0.02	0.98±0.02	0.98±0.02	0.99±0.01

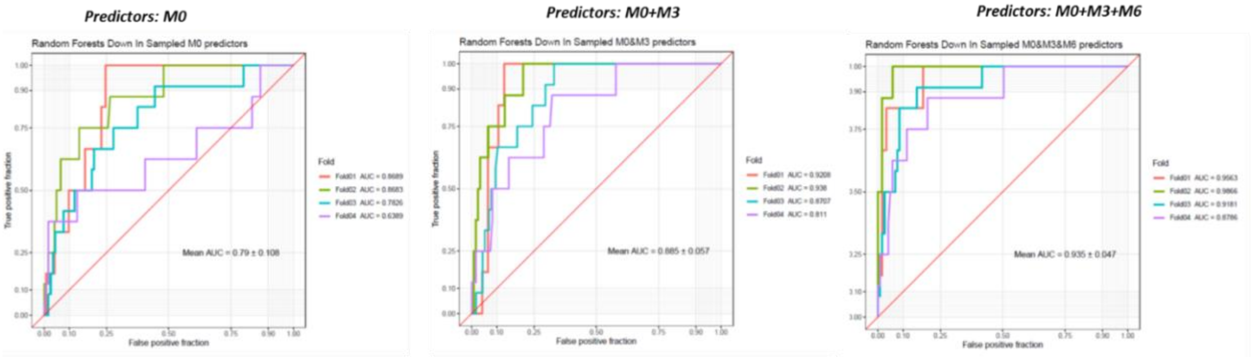


Figure I13: ROC curves for each outer fold of the nested cross validation. Model with all the predictors at M0. Model with all the predictors at M0 and M3. Model with all the predictors at M0, M3 and M6.

Performance profile across different subset sizes for an indicative run of RFE is depicted in Fig. I14. We observe that after approximately 15 variables a plateau is reached in model performance. Overall, the results between repetitions of the RFE workflow are consistent. The variables always selected are the following:

cognitive function C30 (M3), physical function C30 (M3), role function C30 (M3), social function C30 (M3), systemic therapy side effects BR23 (M3), global QoL C30 (M3), pain C30 (M3) and depression HADS (M3).

The following variables were selected in at least 50% of the RFE repetitions:

Treatment control beliefs (M3), fatigue C30 (M3), coping with cancer CBI (M0), anxiety (M3), dyspnoea symptoms C30 (M3) and manageability SOC (M0).
The above 14 variables were selected for the final model.

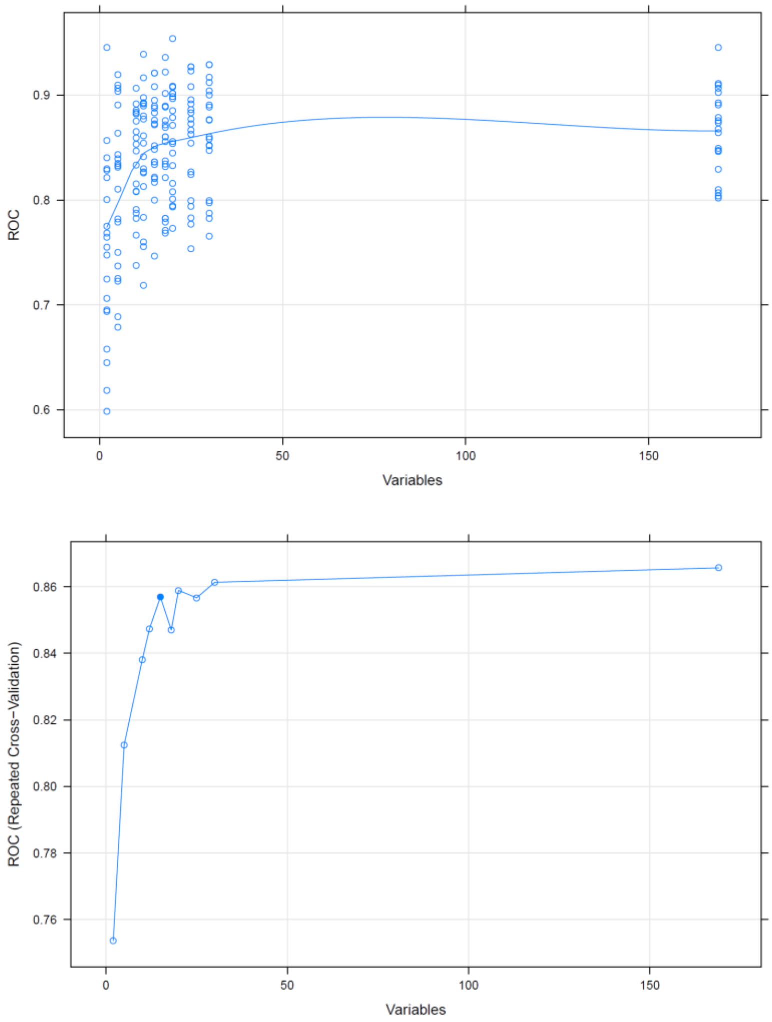


Figure I14. Performance profile of a random forest classifier across different subset sizes. Predictors from M0 and M3. Top panel: ROC cv estimates (4 folds repeated 5 times). Bottom panel: average ROC value as a function of the number of features.

2.3.5.3. Building a classifier for detecting patients with high improving QoL trajectory

The aim is to build a binary classifier able to detect patients of very high QoL trajectory during the 18 month period from baseline. The positive class is the high increasing QoL trajectory group identified in the first part of the analysis (Fig. I12). The negative class emerges from the grouping of the low decreasing and moderate QoL trajectory, as depicted in Fig. I15. It is noted that the specific group of patients are the most resilient group based on both QoL and depression trajectories. The vast majority of these patients also belong to the "good" depression trajectory class and their mean depression levels are very low throughout the 18 months after diagnosis (Fig. I4). The classifier is trained on 513 patients. Predictors considered are psychometric scales, sociodemographic, clinical, medical and treatment variables from baseline or baseline and M3. The number of predictors is 96 from M0 and 169 from M0 and M3. A random forest classifier was chosen for building the final models.

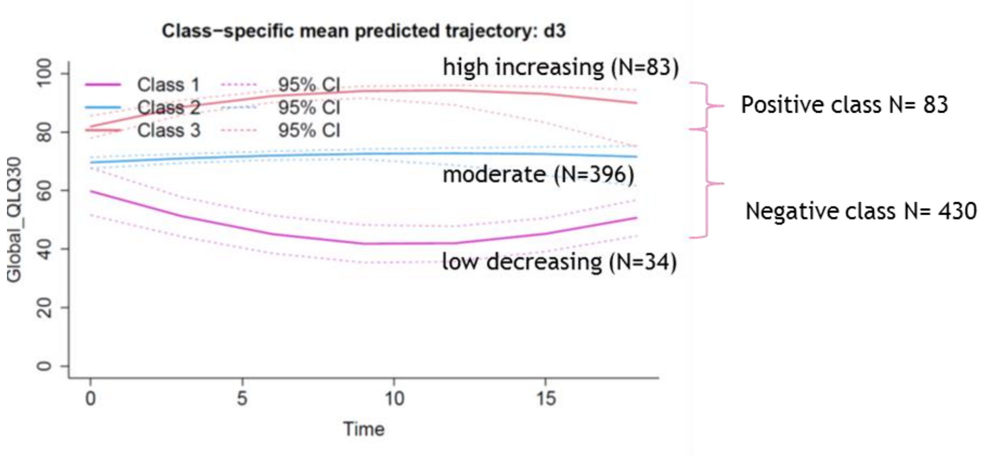


Figure I15. The considered classes

The distribution of the trajectory classes between the clinical sites is reported in Table I5.

Table I5. *Class distribution* (in *percentage) in each clinical site*

	CHAMP	IEO	HUS	HUJI	
moderate/low	21%	16%	32%	15%	84%
high increasing	4%	2%	7%	4%	16%
	24%	18%	39%	19%	

A random forest classifier build on M0 & M3 predictors

The performance evaluation metrics are depicted in Table I6 & Fig. I16. When only M0 variables are considered the performance is acceptable but low (ROC<0.8). The performance improves when M0 and M3 predictors are considered. Specifically average area under ROC value increases to 0.83, sensitivity to 72% and specificity to 75%. Inclusion of M6 variables further improves performance. Results suggest that a classifier that detects patients with low decreasing QoL trajectory can be built at month 3.

Table I6: Performance of Random Forest Classifier based on nested cross validation with four external folds. Average values with standard deviations are given.

Predictors	**All M0**	**All M0 & M3**	**All M0 & M3 except depression at M3**	**All M0, M3 & M6**	**All M0, M3 & M6 except depression at M3 and M6**
Sensitivity	0.69±0.09	0.72±0.11	0.69±0.12	0.78±0.05	0.78±0.03
Specificity	0.71±0.05	0.75±0.05	0.74±0.05	0.80±0.03	0.76±0.03
ROC	0.79±0.05	0.83±0.04	0.80±0.04	0.882±0.002	0.86±0.02
Balanced Accuracy	0.70±0.04	0.73±0.04	0.71±0.05	0.79±0.02	0.77±0.03
F1	0.43±0.08	0.47±0.09	0.45±0.10	0.55±0.09	0.51±0.09
Pos Pred Value	0.31±0.07	0.35±0.08	0.34±0.08	0.43±0.10	0.39±0.09
Neg Pred Value	0.92±0.02	0.94±0.01	0.93±0.01	0.95±0.01	0.95±0.01

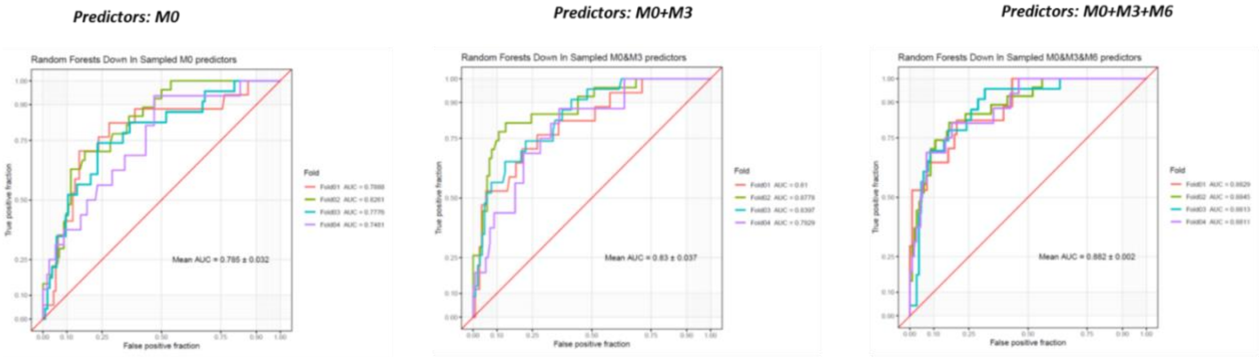


Figure I16: ROC curves for each outer fold of the nested cross validation. Model with all the predictors at M0. Model with all the predictors at M0 and M3. Model with all the predictors at M0, M3 and M6.

Performance profile across different subset sizes for an indicative run of RFE is depicted in Fig. I17. We observe that after approximately 15-20 variables a plateau is reached in model performance. Overall, the results between repetitions of the RFE workflow are consistent. The variables always selected are the following:

depression HADS (M3), fatigue C30 (M3), self-efficacy (M0), global QoL C30 (M3), communication & cohesion FARE (M3), physical function C30 (M3), resilience CDRISC (M0), pain C30 (M3)

The following variables were selected in at least 50% of the RFE repetitions:

future perspective BR23 (M3), role function C30 (M3), cognitive function C30 (M3), social function C30 (M3), positive affect PANAS (M3), personal *control* over the illness (M3), coping with cancer CBI (M0), anxiety (M3), global QoL C30 (M0) and family coping FARE (M3), physical function C30 (M0), perceived support (M3), depression HADS (M0) and helpless MAC (M3).

The above 22 variables were selected for the final model.

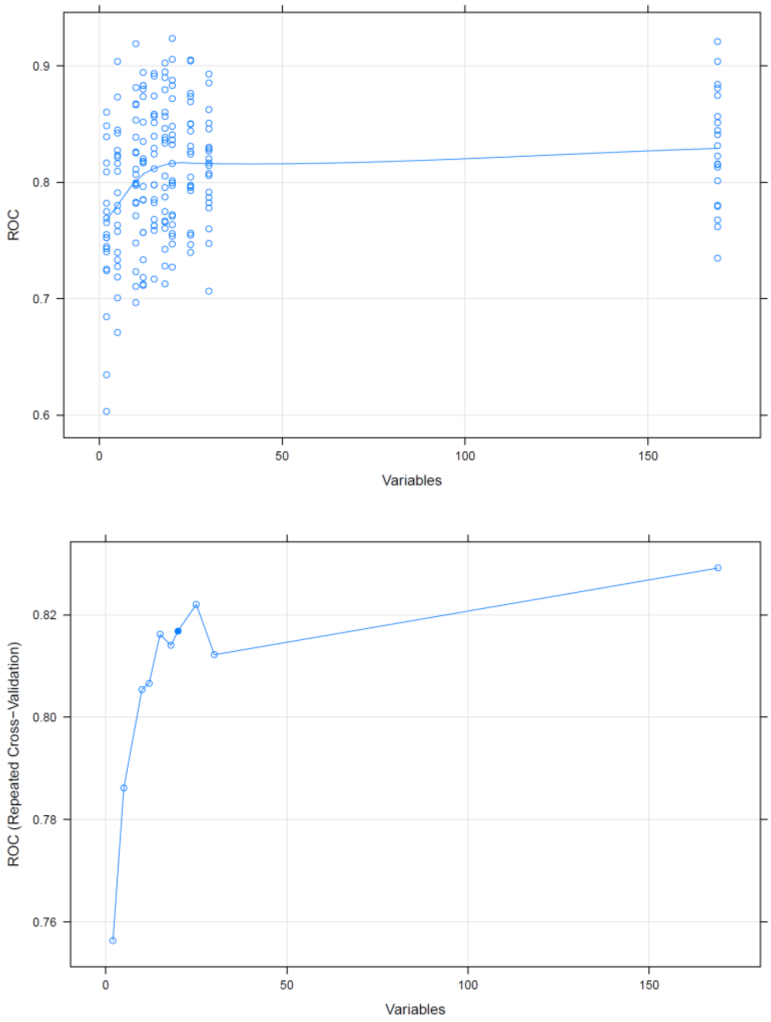


Figure I17. Performance profile of a random forest classifier across different subset sizes. Predictors from M0 and M3. Top panel: ROC cv estimates (4 folds repeated 5 times). Bottom panel: average ROC value as a function of the number of features.

2.4 Conclusion

Both the trajectory clustering analyses and the trajectory classifiers presented in Part 2 of this document appear to have a good performance and reliability. Quantitative measurements through performance evaluation metrics have been provided for several specific cases.

2.5. References

[I1] Katherine J. Lee, Gehan Roberts, Lex W. Doyle, Peter J. Anderson & John B. Carlin (2016) Multiple imputation for missing data in a longitudinal cohort study: a tutorial based on a detailed case study involving imputation of missing outcome data, International Journal of Social Research Methodology, 19:5, 575-591, DOI: 10.1080/13645579.2015.1126486

[I2] Daniel J. Stekhoven, Peter Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data, *Bioinformatics*, Volume 28, Issue 1, 1 January 2012, Pages 112–118, https://doi.org/10.1093/bioinformatics/btr597