SUPPLEMENTARY MATERIAL:
Aid of a machine learning algorithm can improve clinician predictions of patient quality of life during breast cancer treatments

Mikko Nuutinen[a,b], Anna-Maria Hiltunen[a], Sonja Korhonen[a], Ira Haavisto[a,c], Paula Poikonen-Saksela[d], Johanna Mattson[d], Georgios Manikis[e], Haridimos Kondylakis[e], Panagiotis Simos[e,f], Ketti Mazzocco[g,h], Ruth Pat-Horenczyk[i], Berta Sousa[j,k], Fatima Cardoso[j], Isabel Manica[j], Ian Kudel[l] and Riikka-Leena Leskelä[a]

[a]*Nordic Healthcare Group, Helsinki, Finland*
[b]*Haartman Institute, University of Helsinki, Helsinki, Finland*
[c]*Laurea University of Applied Sciences, Sustainable and Versatile Social and Health Care, Vantaa, Finland*
[d]*Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland*
[e]*Computational Biomedicine Laboratory, FORTH-ICS, Heraklion, Greece*
[f]*Medical School, University of Crete, Heraklion, Greece*
[g]*Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy*
[h]*Applied Research Division for Cognitive and Psychological Science, European Institute of Oncology IRCCS, Milan, Italy*
[i]*Paul Baerwald School of Social Work and Social Welfare, The Hebrew University of Jerusalem, Jerusalem, Israel*
[j]*Breast Unit, Champalimaud Clinical Centre, Champalimaud Centre for the Unknown, Champalimaud Foundation, Lisboa, Portugal*
[k]*PhD Programme in Health Data Science Faculty of Medicine of the University of Porto (FMUP)*
[l]*Varian, a Siemens Healthineers Company, Palo Alto, US*

ARTICLE INFO

Keywords:
Clinical decision support system, breast cancer, quality of life, machine learning, user experiment

ABSTRACT

Proper and well-timed interventions may improve breast cancer patient adaptation and quality of life (QoL) through treatment and recovery. The challenge is to identify those patients who would benefit most from a particular intervention. The aim of this study was to measure whether the machine learning prediction incorporated in the clinical decision support system (CDSS) improves clinicians' performance to predict patients' QoL during treatment process. We conducted two user experiments in which clinicians used a CDSS to predict QoL of breast cancer patients. In both experiments each patient was evaluated both with and without the aid of a machine learning (ML) prediction. In Experiment I, 60 breast cancer patients were evaluated by 6 clinicians. In Experiment II, 90 patients were evaluated by 9 clinicians. The task of clinicians was to predict the patient's quality of life at either 6 (Experiment I) or 12 months post-diagnosis (Experiment II). Taking into account input from the machine learning prediction considerably improved clinicians' prediction accuracy. Accuracy of clinicians for predicting QoL of patients at 6 months post-diagnosis was .745 (95% CI .668-.821) with the aid of the prediction provided by the ML model and .696 (95% CI .608-.781) without the aid. Clinicians' prediction accuracy at 12 months was .739 (95% CI .667-.812) with the aid and .709 (95% CI .633-.783) without the aid. When the machine learning model's prediction was correct, the average accuracy of the clinicians for predicting QoL at 6 months was .793 (95% CI .739-.838) with the aid and .720 (95% CI .636-.798) without the aid. Corresponding prediction accuracy of QoL at 12 months was .909 (95% CI .881-.936) and .827 (95% CI .782-.871).

A. Collected variables

The variables of the project were collected during seven waves: baseline (3-4 weeks after the diagnosis) and 3, 6, 9, 12, 15 and 18 months after the baseline. Each collection wave contained a set of specific measures able to capture changes in specific domains. The variables that were not sensitive to changes were collected at only baseline. The collected data is divided into groups of (1) sociodemographic and lifestyle, (2) medical and treatment and (3)

*Corresponding author: Mikko Nuutinen, Nordic Healthcare Group, Vattuniemenranta 2, 00210, Helsinki, Finland, Email: mikko.nuutinen@nhg.fi
ORCID(s):

psychosocial assessment. Patients from Finland, Italy, and Portugal responded to patient-reported questionnaire in their native languages using Noona, an electronic patient-reporting platform. Israeli participants responded to the same questionnaires, translated into Hebrew, using pen-and-paper.

A.1. Sociodemographic and lifestyle variables

The sosiodemographic and lifestyle variables that were collected at the baseline were, e.g., age, education, marital status, number of children, employment status, monthly income, sick leave days, level of religious faith, smoking, drinking habits, use of drug, weight, height, diet and physical exercise. Variables such as employment status, sick leave days, smoking and drinking were collected also from the other waves.

A.2. Medical and treatment variables

Medical and treatment data was collected from the electronic health records of clinical sites. At the baseline, the medical information of cancer stage, comorbidities, genetic risk factor, menopausal status, tumor pathology, psychotropic medication, hormone replacement treatment and laboratory tests were retrieved. The information of ongoing oncological therapy of surgery, chemotherapy or radiotherapy was collected at the time point of 6 months after the baseline. The ongoing oncological therapy of endocrine therapy or anti HER2 was collected at the time point of 12 months after the baseline. Table A.1 presents a descriptive analysis of sociodemographic and lifestyle and medical and treatment variables for the patients in the cohort of this study.

A.3. Psychosocial assessment variables

Several psychosocial assessment tools for measuring domains of personality, meaning, coping, social support, resilience, illness perception and behaviors, quality of life and distress were used and measured at different waves. Personality was measured by TIPI (Ten Item Personality Measure [1]) and LOT-R (Optimism/Pessimism) [2] questionnaires. Meaning was measured by SOC-13 (Sense of Coherence [3]). Coping was measured by PACT (The Perceived Ability to Cope With Trauma [4]), CERQ short (Cognitive Emotion Regulation Questionnaire [5]) and MAAS (Mindful Attention Awareness Scale [6]) questionnaires. Social Support was measured by mMOS-SS (modified Medical Outcomes Study Social Support Survey [7]) questionnaires. Resilience was measured by CD-RISC (Connor-Davidson Resilience Scale [8]) questionnaires. Illness Perception and behaviors was measured by CBI-B (Cancer Behavior Inventory [9]), IPQ (Illness Perception Questionnaire [10]), MAC (Mental Adjustment to Cancer [11]) and PTGI (The Posttraumatic Growth Inventory [12]) questionnaires. Quality of life was measured by QLQ-C30 (EORTC quality of life questionnaire [13]) and QLQ-BR23 (EORTC quality of life questionnaire breast cancer module [14]) questionnaires. Distress was measured by FCRI-SF (Fear of Recurrence - short form [15]), HADS (Hospital Anxiety and Depression Scale [16]) and PANAS (Positive and Negative affectivity [17]) questionnaires. Table A.2 presents the mean values of psychosocial variables for the cohort measured at the baseline and 6 and 12 months after baseline.

B. Open-ended interview

After the second session of Experiment I, an open-ended interview was conducted for the participants. The interview data was analyzed following thematic analysis and the approach identified by [18]. The aim of the interview was to determine how this kind of decision support tool could be used and who would benefit from it and how. Furthermore, the results were utilized for updating the experimental setup of Experiment II. The interview included the following questions:

- Could you make use of this kind of decision support tool when taking care of a patient and how?
- How would you envision it to be used in your organization / department?
- Who (what role/s) in your organization would use such a tool?
- Who (what role/s) in your organization would make use of the information?
- How might the predicted score affect the patient care processes from your perspective / in your organisation?
- Do you think the patients could benefit from this kind of prediction? Under which conditions?
- What aspects should take in consideration when further developing the decision support tool?

Table A.1
Sociodemographic and lifestyle and medical and treatment variables of the patients in the cohort of this study

	Variable	Value
Sociodemographic and lifestyle	Age, mean (std)	54.69 (8.99)
	BMI, mean (std)	25.72 (4.75)
	Education High	65.3%
	Part-time or unemployed	27.4%
	Low income	30.9%
	No exercise	30.9%
	Living alone	23.2%
	Number of children, mean (std)	1.97 (1.41)
	Ever smoked	32.3%
	Sick leave days, mean (std)	12.23 (19.42)
Medical and treatment	Cancer family history	40.3%
	Cancer stage = 1	47.7%
	Cancer stage = 2	42.6%
	Cancer stage = 3	9.6%
	Cancer grade = 1	17.4%
	Cancer grade = 2	51.2%
	Cancer grade = 3	31.2%
	Cancer grade = 4	0.2%
	Performance status = 0	88.9%
	Performance status = 1	10.9%
	Performance status = 2	0.2%
	Performance status = 3	0.0%
	Radiotherapy treatment	79.7%
	Chemotherapy treatment	52.6%
	Endocrine treatment	84.4%
	Antiher treatment	16.3%
	COPD	0.5%
	Hypertension	16.5%
	Asthma	5.3%
	Cardiac disease	0.7%
	Diabetes	3.4%
	Hypothyreosis	6.4%
	Arthrosis	2.7%
	Depression	2.2%

std = Standard deviation
Education high = Bachelor, high school, postgraduate school or vocational non academic diploma
Low income = Net monthly income 0 -1500€

B.1. Results

According to the results, all clinicians found the CDSS to be useful if incorporated into the care of breast cancer patients. According to participants, the information provided by the CDSS would not likely affect the actual breast cancer treatment of patients or the choice of therapies, but rather influence the psychosocial support and other possible interventions offered to patients. However, there was a consensus that for the quality of life prediction to be valuable it must lead to an actual intervention for the patient. The usefulness of the tool is therefore affected by the availability of interventions to improve quality of life and adaptation. Furthermore, one participant thought that the prediction would be most useful and informative in cases where the predicted quality of life is lower than the clinician's intuitive prediction. Several participants of Experiment I viewed that the CDSS would be most useful if it could identify the patients with weak quality of life 12 months after the end of treatment, at which point in time a portion of patients are generally less vigorous than the majority. The quality of life prediction could then be used to target specific individually planned interventions and a higher level of support for this group of patients. The optimal timing for the use of the CDSS is thought to differ between patients, varying from the time of planning adjuvant treatment to the post-treatment period.

Most participants thought that both doctors and nurses may be possible users of the CDSS and could make use of prediction information. However, the suitable user depends on which interventions would follow from the prediction, as offering certain interventions may require a referral from a doctor. However, the likelihood and motivation of clinicians to use the CDSS is generally believed to be significantly affected by the ease of use and convenience of the tool. With regards to breast cancer patients, there were conflicting views on whether the information provided by the tool would be useful to be shared with patients. While some participants viewed that patients learning their quality of life prediction may motivate and encourage them through their treatment and rehabilitation process, some participants worried that a poor prediction may cause discouragement and increase stress. Therefore, if the prediction is shared with patients the manner in which the information is communicated must be paid attention to.

In further development of the CDSS, one participant highlighted the importance of the incorporation of more parameters concerning breast cancer treatment and possible comorbidities into the CDSS, while another hoped for more detailed information of the mental health and possible medications of the patient. Furthermore, one participant also hoped for the patient perspective in terms of their feelings towards learning their predicted quality of life to be explored further.

Table A.2
Baseline and 6 and 12 months after the baseline values of psychosocial scale values of the patients in the cohort of this study, mean (std). The psychosocial scale values are calculated by combining several raw values. Raw values are the patients' answers for the individual psychosocial questions.

Variable	Baseline	Month 6	Month 12
TIPI extraversion	4.67 (1.70)		
TIPI aggreableness	5.50 (0.99)		
TIPI conscientiousness	5.87 (1.03)		
TIPI emot stab	4.62 (1.31)		
TIPI openess	5.07 (1.29)		
LOT optimism	2.72 (0.64)		
SOC comprehensibility	20.61 (4.34)		
SOC manageability	19.95 (4.23)		
SOC meaningfulness	22.98 (3.72)		
PACT Forward	5.20 (0.96)		
PACT Trauma	5.31 (0.83)		
PACT Flexibility	9.89 (1.82)		
CERQ self blame	2.01 (0.81)		
CERQ other blame	1.46 (0.60)		
CERQ rumination	2.98 (0.92)		
CERQ catastrophizing	1.97 (0.84)		
CERQ perspective	3.37 (0.97)		
CERQ pos refus	3.02 (0.95)		
CERQ pos reapp	3.66 (1.00)		
CERQ acceptance	3.72 (0.98)		
CERQ planning	3.44 (0.98)		
CERQ negative overall	2.11 (0.53)		
CERQ positive overall	3.44 (0.67)		
MAAS mindfulness	4.36 (0.71)		
CDRISC resilience	2.85 (0.65)		
CBI coping with cancer	7.26 (1.11)	7.41 (1.09)	7.17 (1.55)
QLQ30 Global	73.32 (18.65)	75.59 (16.98)	75.39 (18.67)
QLQ30 Phys Fun	86.70 (14.90)	83.50 (15.34)	84.34 (15.95)
QLQ30 Role Fun	80.19 (23.64)	83.09 (20.27)	86.09 (21.11)
QLQ30 Emot Fun	74.68 (19.53)	76.72 (17.16)	77.27 (19.03)
QLQ30 Cogn Fun	80.25 (18.74)	82.09 (18.53)	80.22 (20.54)
QLQ30 Soc Fun	81.22 (22.82)	81.82 (22.15)	86.24 (22.47)
QLQ30 Fatigue	22.22 (21.34)	23.88 (20.13)	22.14 (20.64)
QLQ30 Nausea	7.50 (12.11)	9.13 (14.47)	8.47 (14.56)
QLQ30 Pain	25.08 (20.79)	29.19 (21.66)	25.74 (22.41)
QLQ30 Dyspnoea	6.41 (14.61)	10.04 (17.75)	8.77 (18.14)
QLQ30 Insomnia	11.74 (23.26)	15.12 (25.29)	11.43 (24.63)
QLQ30 Apetite	5.63 (17.25)	6.23 (16.31)	3.63 (14.03)
QLQ30 Constipation	28.01 (25.18)	31.22 (23.65)	28.43 (24.83)
QLQ30 Diarrhoea	10.95 (19.15)	14.82 (21.84)	12.52 (21.33)
QLQ30 Financial	12.58 (23.25)	15.73 (23.75)	10.95 (21.91)
BR23 Body Image	85.50 (17.91)	76.80 (21.88)	77.90 (22.66)
BR23 Side Effects	13.37 (12.99)	20.52 (14.62)	18.22 (13.89)
BR23 Breast Symptoms	19.95 (16.09)	22.87 (17.01)	16.49 (14.89)
BR23 Arm Symptoms	15.06 (16.26)	18.59 (17.72)	17.60 (20.29)
BR23 Upset Hair Image	60.31 (20.65)	34.48 (20.32)	32.85 (19.89)
BR23 Future Persp Image	53.78 (27.73)	61.52 (24.78)	62.73 (25.40)
BR23 Sex Funct	27.86 (21.97)	27.07 (19.76)	27.62 (22.94)
BR23 Sex Enjoy	61.77 (19.75)	60.25 (21.06)	59.10 (21.30)
FCRI fear of recur	1.70 (0.62)	1.69 (0.67)	1.68 (0.69)
HADS anxiety	0.96 (0.56)	0.72 (0.49)	0.76 (0.53)
HADS depression	0.52 (0.47)	0.52 (0.47)	0.54 (0.52)
HADS mental health total	0.74 (0.47)	0.62 (0.43)	0.65 (0.48)
PANAS negative affect	1.90 (0.77)	1.63 (0.66)	1.67 (0.70)
PANAS positive affect	3.50 (0.70)	3.42 (0.68)	3.45 (0.71)
IPQ timeline		2.81 (0.81)	
IPQ timecycl		2.50 (0.74)	
IPQ conseque		2.97 (0.68)	
IPQ perscon		3.33 (0.71)	
IPQ treatcon		3.99 (0.59)	
IPQ illcoher		3.45 (0.80)	
IPQ emotrepr		2.73 (0.76)	
PTGI relating to others			2.66 (1.43)
PTGI new possibilities			2.36 (1.44)
PTGI personal strength			3.01 (1.43)
PTGI spiritual change			1.70 (1.54)
PTGI appreciation of life			2.98 (1.38)
PTGI total score			2.54 (1.19)

LOT = Optimism/Pessimism [2], SOC = Sense of Coherence [3], PACT = The Perceived Ability to Cope With Trauma [4], CERQ = Cognitive Emotion Regulation Questionnaire [5], MAAS = Mindful Attention Awareness Scale [6], CBI = Cancer Behavior Inventory [9], QLQ-C30 = EORTC quality of life questionnaire [13], HADS = Hospital Anxiety and Depression Scale [16], PANAS = Positive and Negative affectivity [17], TIPI = Ten Item Personality Measure [1], QLQ-BR23 = EORTC quality of life questionnaire breast cancer module [14], FCRI = Fear of Recurrence - short form [15], IPQ = Illness Perception Questionnaire [10]

C. Machine learning models

C.1. Data preprocessing

The framework for data integration and homogenization from different clinical sites is presented in [19]. All the collected variables are presented in details in [20]. Data preprocessing step of this study detected outliers, replaced possible missing values and processed categorical variables after test patients were separated from data set but prior to machine learning model training. Outliers were detected as cases of out of the possible value ranges. The value ranges of important variables of the models of this study are presented in Tables C.5 and C.8. Cases and variables with more than 10% of missingness were excluded from the dataset. Remaining missing values were replaced by the global

median value. No missing values were detected from the important variables (Tables C.5 and C.8). Categorical nominal variables were encoded to binary variables (marital status -> living alone; education -> education high; employment status -> part-time or unemployment, physical exercise practises -> no exercise). Categorical ordinal variables were encoded to continuous integer and binary variables (net monthly income category -> low income).

The target variable of the trained machine learning models was the binarized EORTC QLQ30 Global QoL scale. That is, the trained task for the models was binary classification. In Experiment I, the machine learning model was trained for predicting QoL class after six months from the baseline (Figure C.1a). All predictor variables were collected from the baseline (Month 0). In Experiment II, the machine learning model was trained for predicting QoL after 12 months from the baseline (Figure C.1b). The predictor variables were collected from the timeline between the baseline and three months after the baseline.

In both user experiments, the predicted probability of high QoL value was displayed on the user interface. The threshold of the binarization was the value of 75. The patients with QoL value > 75 were encoded in the high QoL class. The patients with QoL value ≤ 75 were encoded in the low QoL class. Figure C.2 presents the distributions of EORTC QLQ30 Global QoL scale values for month 6 (target of Experiment I) and month 12 (target of Experiment II). The threshold of the classification task is presented as a red line on the figures. After binarization, the target variable of Experiment II is balance (Low/High = 46%/54%) and of Experiment I pretty close to balance (Low/High = 62%/38%).

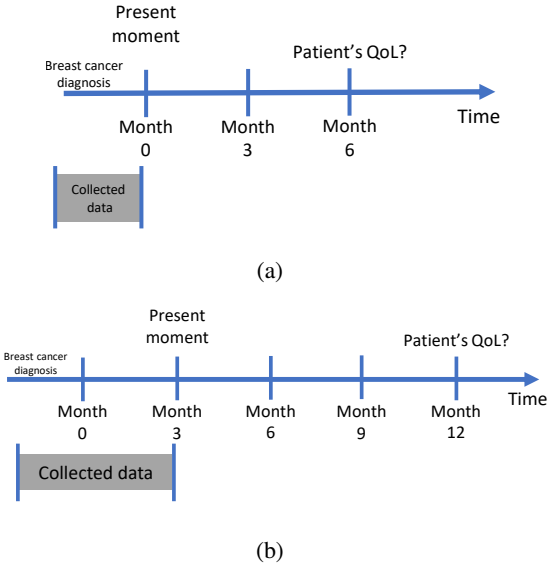
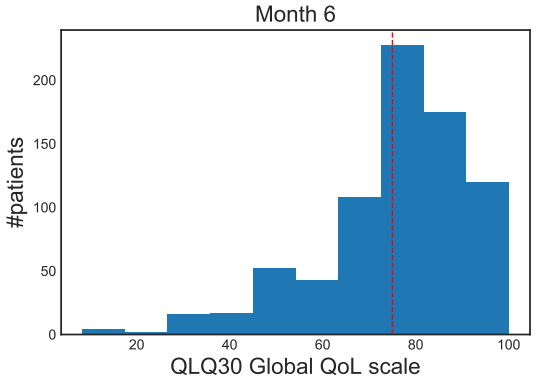


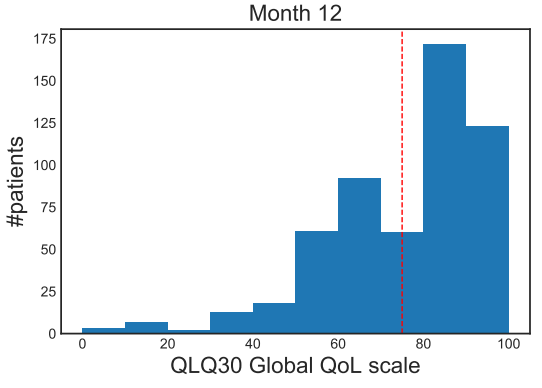
Figure C.1: Timeline of data collection for machine learning model training in Experiment I (a) and Experiment II (b)

C.2. Experiment I

Figure C.3 shows the data flow of machine learning model training pipeline for Experiment I. First, the test patients ($n = 60$) were separated from the data set. The model was trained (variable selection, hyperparameter searching and model fitting) with the train patients ($n = 548$). 40 variables from four variable groups were selected for model training (Step 1). The principle was that 10 variables were selected from each variable group. The groups were sociodemographic and lifestyle, medical and treatment and psychosocial assessment (raw and scale variables). Psychosocial assessment was divided into two sub-groups: raw values and scale values. Raw values were the patients' answers for the individual psychosocial questions. Scale values were calculated by combining several raw values. The sub-sets of 10 variables from each variable group were selected by the sequential forward variable selection (SFS) method. SFS starts with an empty set and adds one variable at a time from the original group for classifier by maximizing a performance measure. For the variable selection we used logistic regression classifier. That is, variables were selected by training logistic regression classifier by searching the most efficient sub-sets of 10 variables for each variable group at a time to predict the class of QoL. For the selected 40 variables (4x10 variables from the SFS) we trained random forest classifier. A grid-search algorithm was used to tune the hyperparameters of the classifier (Step



(a)



(b)

Figure C.2: Distributions of EORTC QLQ30 Global QoL scale values for Month 6 (target of Experiment I) (a) and Month 12 (target of Experiment II) (b). Red dashed lines indicate the threshold value for the patient classification of low and high QoL class.

2). The hyperparameters for grid-search were as follows: max depth 2, 3, 4, 6, 8, min samples leaf 1, 2, 3, 4, min samples split 1, 2, 3, 4, number of estimators 20, 25, 30, 35, 40. The hyperparameters determined by grid-search of the classifier were [max depth 4, min samples leaf 3, min samples split 2, number of estimators 30]. The final random forest classifier was trained using the hyperparameters and all training data (Step 3). The ranges of hyperparameters of decision trees of the trained random forest classifier were [number of estimators 30, depth of trees 4, number of leaves 12-16].

Model evaluation

Table C.3 presents the confusion matrix of the machine learning model for the test data set. The model classified 6/38 low QoL patients in the group of high QoL (false positives). 6 patients from the group of high QoL were classified in the group of low QoL (false negatives). The *AU ROC* value of the machine learning model for the test data set was .832 (95% CI .757-.900). Recall and precision values were .727 (95% CI .583-.857) and .727 (95% CI .589-.854) when the threshold value of the model was .60. Figure C.4 shows the calibration properties of the machine learning model. The Brier score was .185. As can be seen from the figure, higher predicted probabilities present some overestimations. Table C.4 lists the 10 most important variables of the machine learning model according to the random forest feature importance values. The variables of QLQ-Global QoL, HADS-mental health and Distress thermometer at the baseline (Month 0) were important psychosocial factors. Age, BMI and monthly income were important sociodemographic and lifestyle factors.

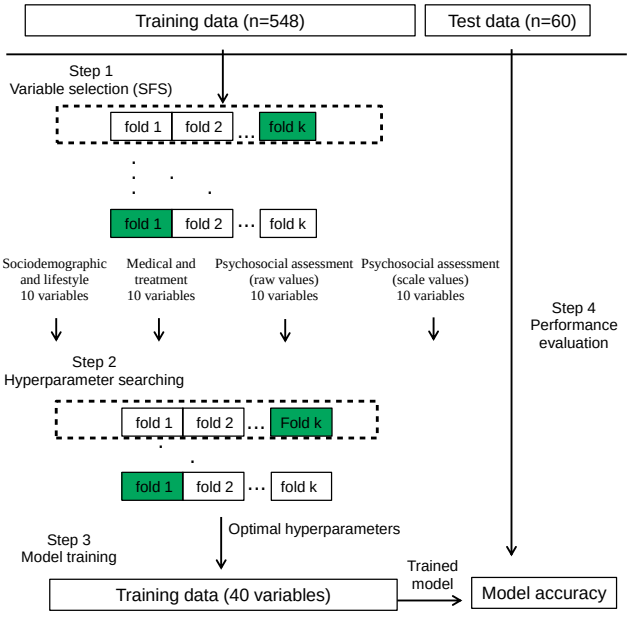


Figure C.3: Experiment I. Data flow of machine learning model training pipeline: variable selection (Step 1), hyperparameter searching (Step 2), model training (Step 3) and performance evaluation (Step 4). SFS = Sequential forward variable selection.

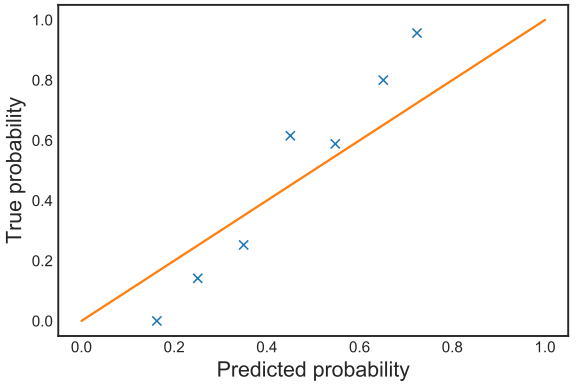


Figure C.4: Experiment I. Calibration of the model of Experiment I.

Shapley values

Figure C.5 and Table C.5 list 15 variables of the trained random forest classifier sorted by the highest sum of absolute SHAP values (SHapley Additive exPlanations [21, 22]) over all patients of the training set. Table C.5 presents also the scale, type and minimum and maximum values for the important variables. The distributions of the data points on the plots of Figure C.5 show the impacts of each variable for the classifier output. The red points indicate higher patient specific variable values than the average value of the variable and, the blue points lower patient specific variable values than the average value of the variable. Longer distance between red and blue points indicates better capacity of a variable to predict high QoL. SHAP values show that the strongest impact on the prediction from the psychosocial assessment scale values were the variables of QLQ-Global QoL, CBI-Coping with cancer, HADS-Anxiety and HADS-Mental health. The strongest impact from the psychosocial assessment raw values were the variables of QLQ-C30-29,

Table C.3
Model of Experiment I. Confusion matrix for the trained machine learning model for the test data set when the classification threshold was .60.

	Predicted Low QoL	Predicted High QoL
Low QoL	32	6
High QoL	6	16

Table C.4
Model of Experiment I. Variable importance values of the trained model (Random Forest feature importance values).

Variable	Importance value
QLQ-Global QoL	0.106
HADS-Mental health	0.079
Age	0.072
NCCN-Distress thermometer	0.071
QLQ-C30 - 30	0.057
QLQ-C30 - 29	0.048
Panas 5	0.044
Monthly income	0.043
CBI-Coping with cancer	0.041
BMI	0.040

QLQ-C30 = EORTC quality of life questionnaire [13]
HADS = Hospital Anxiety and Depression Scale [16]
PANAS = Positive and Negative affectivity [17]
CBI = Cancer Behavior Inventory [9]
BMI = Body Mass Index

Table C.5
Model of Experiment I. Name and mean, standard deviation, minimum and maximum values of the variables

Name	Mean value	STD	Min value	Max value	Type	Range
QLQ-Global QoL	71.33	19.32	0.00	100.00	Continuous	0-100
QLQ-C30 - 29	5.24	1.20	1.00	7.00	Ordinal	1-7
Distress thermometer	3.93	2.82	0.00	10.00	Ordinal	0-10
Age	54.05	9.14	41.00	72.00	Continuous	40-100
CBI-Coping with cancer	7.14	1.17	3.25	9.00	Continuous	1-9
Panas 2	3.10	1.02	1.00	5.00	Ordinal	1-5
BMI	25.51	4.69	17.10	42.46	Continuous	15-50
QLQ-C30 - 30	5.27	1.26	1.00	7.00	Ordinal	1-7
QLQ-C30 - 25	1.85	0.73	1.00	4.00	Ordinal	1-4
HADS-Anxiety	0.95	0.56	0.00	2.71	Continuous	0-3
Panas 5	2.10	1.16	1.00	5.00	Ordinal	1-5
QLQ-C30 - 19	1.94	0.75	1.00	4.00	Ordinal	1-4
HADS-Mental health	0.77	0.49	0.00	2.50	Continuous	0-3
PANAS-Negative affect	2.00	0.81	1.00	5.00	Ordinal	1-5
Panas 3	2.45	1.20	1.00	5.00	Ordinal	1-5

QLQ-C30-30, Panas 2 and Distress thermometer questions. The strongest impact from the sociodemographic and lifestyle variables were the variables of Age and BMI. Any variable from the group of Medical and treatment did not rise to the group of top 15 variables.

Partial dependence plots

Figure C.6 shows partial dependence plots (PDP, [23]) for the 12 variables with the highest absolute sum of SHAP values. The PDP plots present how variables affect QoL predictions. We found that the plots of the following variables show the largest probability scales (>.04): QLQ-Global QoL, QLQ-C30-29, Distress thermometer, Age, CBI-Coping

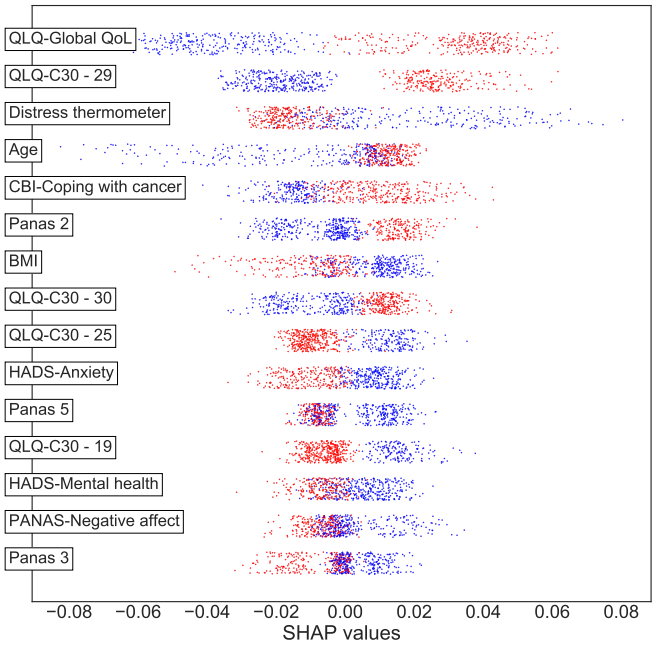


Figure C.5: Model of Experiment I. SHAP values (SHapley Additive exPlanations) for the 15 most important variables predicting high QoL. Random forest was used as a classifier. The definition of the 15 most important variables is based on the sum of the absolute SHAP values. QLQ-C30 = EORTC quality of life questionnaire [13], CBI = Cancer Behavior Inventory [9], PANAS = Positive and Negative affectivity [17], HADS = Hospital Anxiety and Depression Scale [16], BMI = Body Mass Index.

with cancer and BMI. That is, the average predicted high QoL probability varied more than .04 units between the low and high value of these variables, whereas for the other variables the probability varied less than .04 units. For example, the PDP of QLQ-Global QoL showed a large scale of the probability of high QoL ranging from value of .36 for patients with smaller baseline than 60, up to a value of about .44 for patients with higher baseline than 80 (Figure C.6a). Similarly, if patient's QLQ-C30-29 value was smaller than 5, average probability of high QoL was about .37 and if higher than 6, average probability was about .42 (Figure C.6b). The plot of age showed a sharp drop the probability of high QoL when the patient's age was less than 50 years (Figure C.6d). The probability was over .4 when the patient was over 50 years old. When patient was about 45 years old, the probability decreases to .35. The probability value was .41 for patients with BMI less than 25 (Figure C.6g). When BMI was about 25 and more the probability decreased to .375.

C.3. Experiment II

Figure C.7 shows the data flow of machine learning model training pipeline of Experiment II. First, the test patients ($n = 90$) were separated from the data set. The model was trained (variable selection, hyperparameter searching and model fitting) with the train patients ($n = 397$). The hyperparameters and 15 most significant variables were searched via nested cross validation feature selection, using consensus-based feature importance ranking (Step 1). The final random forest classifier was trained using the optimal hyperparameters and 15 variables and all training data (Step 2). The ranges of hyperparameters of decision trees of the trained random forest classifier were [number of estimators 100, depth of trees 8-18, number of leaves 47-63].

Model evaluation

Table C.6 presents the confusion matrix for the 90 patients (test set). The classifier correctly identified 34/49 patients from the group of high QoL and classified 15 low QoL patients in the group of high QoL (false positives). 15 patients from the group of high QoL were classified in the group of low QoL (false negatives). The *AU ROC* value for the test data set was .712 (95% CI .597-.814). Recall and precision values were .694 (95% CI .556-.822) and .694

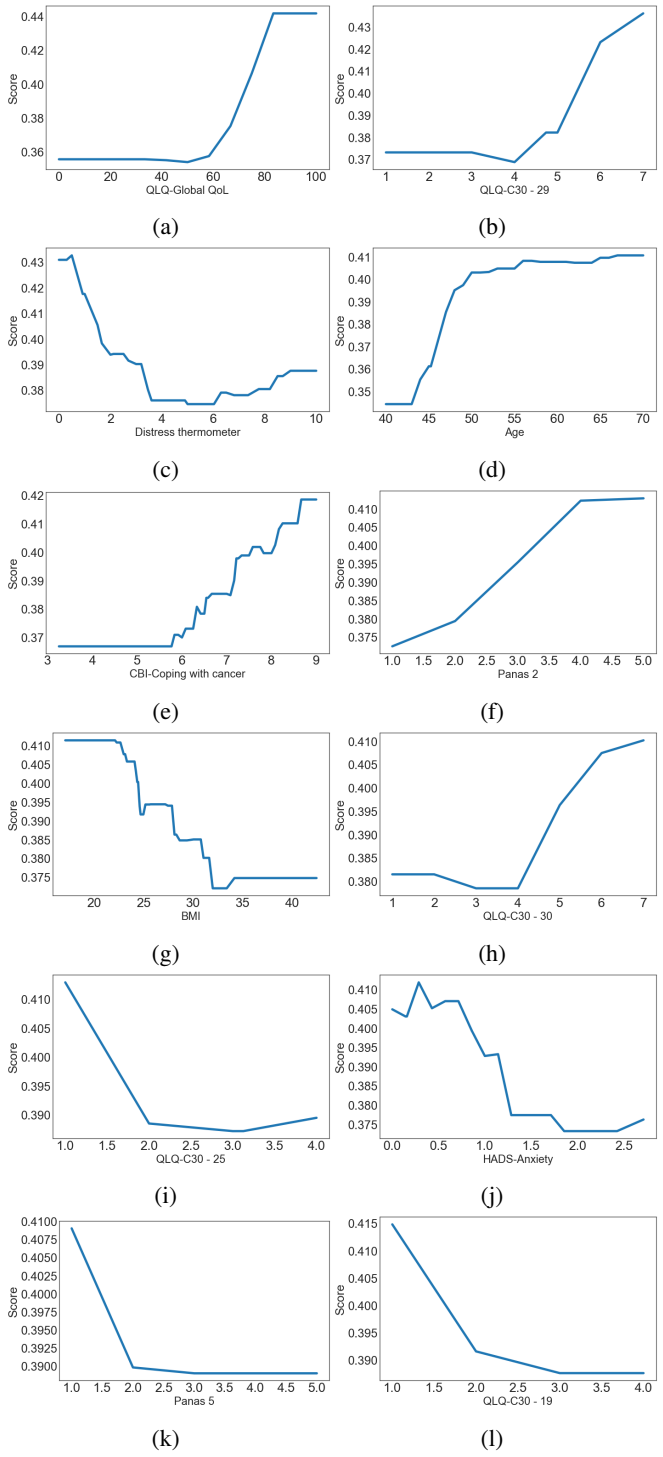


Figure C.6: Model of Experiment I. Partially dependence plots (PDP) for the 12 variables that had the highest SHAP values (SHapley Additive exPlanations). The Figure C.6 is related to the Figure C.5.

(95% CI .569-.820) when the threshold value of the model was .50. Figure C.8 shows the calibration properties of the machine learning model. The Brier score was .340. Table C.7 lists the 10 most important variables of the trained model according to the random forest feature importance values. The variables of QLQ-Role Functioning, HADS-Depression

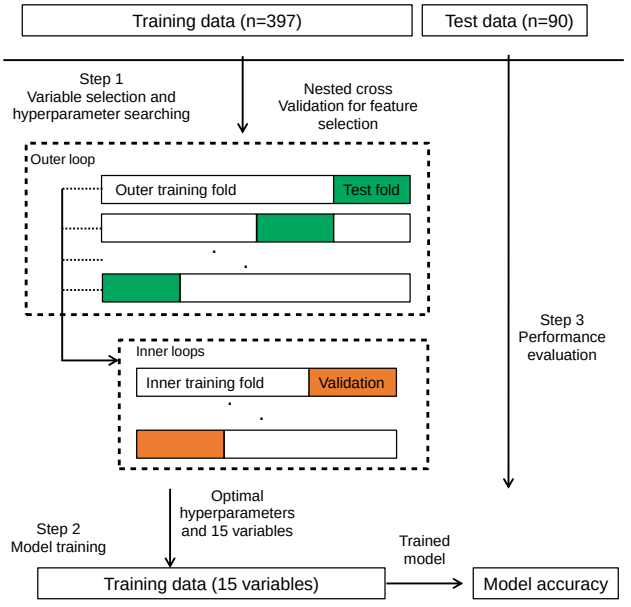


Figure C.7: Experiment II. Data flow of machine learning model training pipeline: Variable selection and hyperparameter searching (Step 1), model training (Step 2) and performance evaluation (Step 3)

and PANAS-Positive affect were the three most important psychosocial factors after three months from the baseline according to the importance values. CBI-Coping with cancer was the most important factor at the baseline.

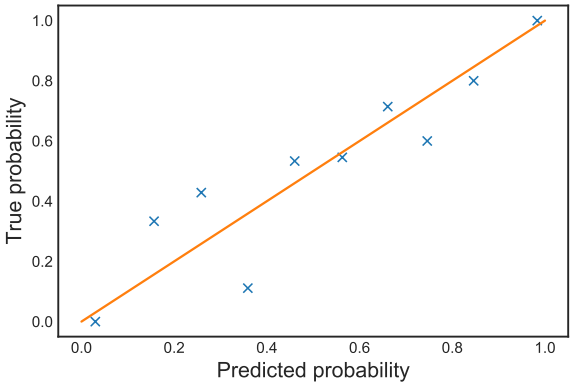


Figure C.8: Experiment II. Calibration of the model of Experiment II.

Shapley values

Figure C.9 and Table C.8 list 15 variables of the trained random forest classifier sorted by the highest sum of absolute SHAP values over all patients of the training set. SHAP values show that the strongest impact on the prediction from the variables were the variables of QLQ-Role Functioning, HADS-Depression and PANAS-Positive affect.

Table C.6
Model of Experiment II. Confusion matrix of the trained random forest classifier for the test data set

	Predicted Low QoL	Predicted High QoL
Low QoL	26	15
High QoL	15	34

Table C.7
Model of Experiment II. Variable importance values of the trained model (Random Forest feature importance values).

Variable	Importance value
QLQ-C30-Role Functioning (Month 3)	0.098
HADS-Depression (Month 3)	0.098
CBI-Coping with cancer (Month 0)	0.083
PANAS-Positive affect (Month 3)	0.079
MAAS-Mindfulness (Month 0)	0.075
HADS-Depression (Month 0)	0.071
PACT-Forward (Month 0)	0.065
CERQ-Positive overall (Month 0)	0.063
SOC-Manageability (Month 0)	0.062
LOT-Optimism (Month 0)	0.061

QLQ-C30 = EORTC quality of life questionnaire [13]
HADS = Hospital Anxiety and Depression Scale [16]
CBI = Cancer Behavior Inventory [9]
PANAS = Positive and Negative affectivity [17]
MAAS = Mindful Attention Awareness Scale [6]
PACT =The Perceived Ability to Cope With Trauma [4]
CERQ = Cognitive Emotion Regulation Questionnaire [5]
SOC = Sense of Coherence [3]
LOT = Optimism/Pessimism [2]

Table C.8
Model of Experiment II. Name and mean, standard deviation, minimum and maximum values of the variables

Name	Mean value	STD	Min value	Max value	Type	Range
QLQ-Role Functioning (Month 3)	75.26	25.91	0.00	100.00	Continuous	0-100
HADS-Depression (Month 3)	0.59	0.53	0.00	3.00	Continuous	0-3
PANAS-Positive affect (Month 3)	3.31	0.75	1.20	5.00	Continuous	1-5
HADS-Depression (Month 0)	0.52	0.47	0.00	2.57	Continuous	0-3
MAC-helpless (Month 3)	1.40	0.44	1.00	3.75	Continuous	1-4
CBI-Coping with cancer (Month 0)	7.26	1.11	3.33	9.00	Continuous	1-9
CERQ-Positive overall (Month 0)	3.44	0.67	1.50	5.00	Continuous	1-5
SOC-Manageability (Month 0)	19.95	4.23	4.00	28.00	Continuous	0-28
QLQ-Cognitive Functioning (Month 3)	79.25	20.27	0.00	100.00	Continuous	0-100
LOT-Optimism (Month 0)	2.72	0.64	0.50	4.00	Continuous	0-4
QLQ-Social Functioning (Month 3)	76.07	24.92	0.00	100.00	Continuous	0-100
PACT-Forward (Month 0)	5.20	0.96	1.56	7.00	Continuous	1-7
CERQ-Acceptance (Month 0)	3.72	0.98	1.00	5.00	Continuous	1-5
MAAS-Mindfulness (Month 0)	4.36	0.71	1.73	5.93	Continuous	0-7
PTGI-total score (Month 3)	2.52	1.13	0.00	4.80	Continuous	0-5

Partial dependency plots

Figure C.10 shows PDP plots for the 12 variables with the highest absolute sum of SHAP values. We found that the following variables showed the largest probability scales (> .15): QLQ-Role Functioning (Month 3), HADS-Depression (Month 3), PANAS-Positive affect (Month 3) and SOC-Manageability (Month 0). That is, the average predicted high probability of low QoL varied more than .15 units between the low and high value of these variables, whereas for the other variables the probability varied less than .15 units.

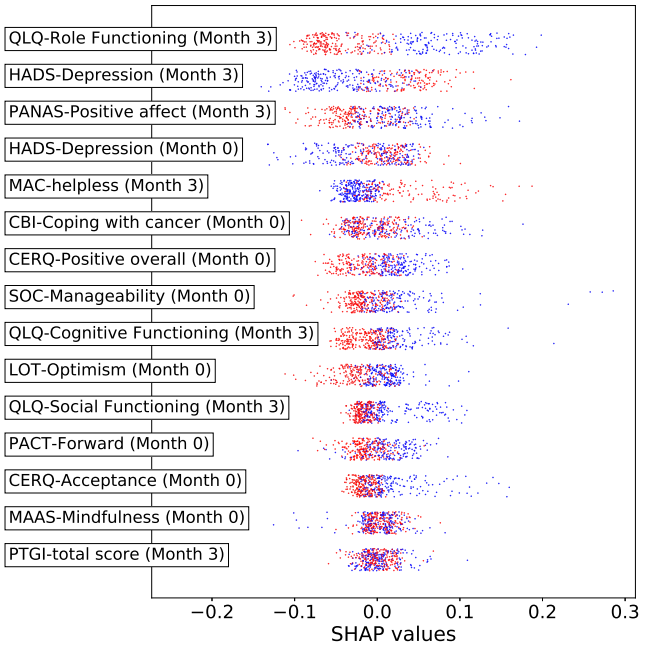


Figure C.9: Model of Experiment II. SHAP values (SHapley Additive exPlanations) for the 15 most important variables predicting low QoL. Random forest was used as a classifier. The definition of the 15 most important variables is based on the sum of the absolute Shapley values. QLQ-C30 = EORTC quality of life questionnaire [13], HADS = Hospital Anxiety and Depression Scale [16], PANAS = Positive and Negative affectivity [17], MAC = Mental Adjustment to Cancer [11], CBI = Cancer Behavior Inventory [9], CERQ = Cognitive Emotion Regulation Questionnaire [5], SOC = Sense of Coherence [3], LOT = Life Orientation Test [2], PACT = Perceived Ability to Cope With Trauma [4], MAAS = Mindful Attention Awareness Scale [6], PTGI = Posttraumatic Growth Inventory [12].

References

[1] S. D. Gosling, P. J. Rentfrow, W. B. Swann Jr, A very brief measure of the big-five personality domains, Journal of Research in personality 37 (2003) 504–528.

[2] M. F. Scheier, C. S. Carver, M. W. Bridges, Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test., Journal of personality and social psychology 67 (1994) 1063–1078.

[3] A. Antonovsky, Unraveling the mystery of health: How people manage stress and stay well., Jossey-bass, 1987.

[4] G. A. Bonanno, R. Pat-Horenczyk, J. Noll, Coping flexibility and trauma: the perceived ability to cope with trauma (pact) scale., Psychological Trauma: Theory, Research, Practice, and Policy 3 (2011) 117.

[5] N. Garnefski, V. Kraaij, Cognitive emotion regulation questionnaire–development of a short 18-item version (cerq-short), Personality and individual differences 41 (2006) 1045–1053.

[6] K. W. Brown, R. M. Ryan, The benefits of being present: mindfulness and its role in psychological well-being., Journal of personality and social psychology 84 (2003) 822.

[7] C. D. Sherbourne, A. L. Stewart, The mos social support survey, Social science & medicine 32 (1991) 705–714.

[8] L. Campbell-Sills, M. B. Stein, Psychometric analysis and refinement of the connor–davidson resilience scale (cd-risc): Validation of a 10-item measure of resilience, Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies 20 (2007) 1019–1028.

[9] C. A. Heitzmann, T. V. Merluzzi, P. Jean-Pierre, J. A. Roscoe, K. L. Kirsh, S. D. Passik, Assessing self-efficacy for coping with cancer: Development and psychometric analysis of the brief version of the cancer behavior inventory (cbi-b), Psycho-Oncology 20 (2011) 302–312.

[10] R. Moss-Morris, J. Weinman, K. Petrie, R. Horne, L. Cameron, D. Buick, The revised illness perception questionnaire (ipq-r), Psychology and health 17 (2002) 1–16.

[11] M. Watson, M. G. Law, M. d. Santos, S. Greer, J. Baruch, J. Bliss, The mini-mac: further development of the mental adjustment to cancer scale, Journal of Psychosocial Oncology 12 (1994) 33–46.

[12] A. Cann, L. G. Calhoun, R. G. Tedeschi, K. Taku, T. Vishnevsky, K. N. Triplett, S. C. Danhauer, A short form of the posttraumatic growth inventory, Anxiety, Stress, & Coping 23 (2010) 127–137.

[13] N. K. Aaronson, S. Ahmedzai, B. Bergman, M. Bullinger, A. Cull, N. J. Duez, A. Filiberti, H. Flechtner, S. B. Fleishman, J. C. de Haes, The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology, J Natl Cancer Inst 85 (1993) 365–376.

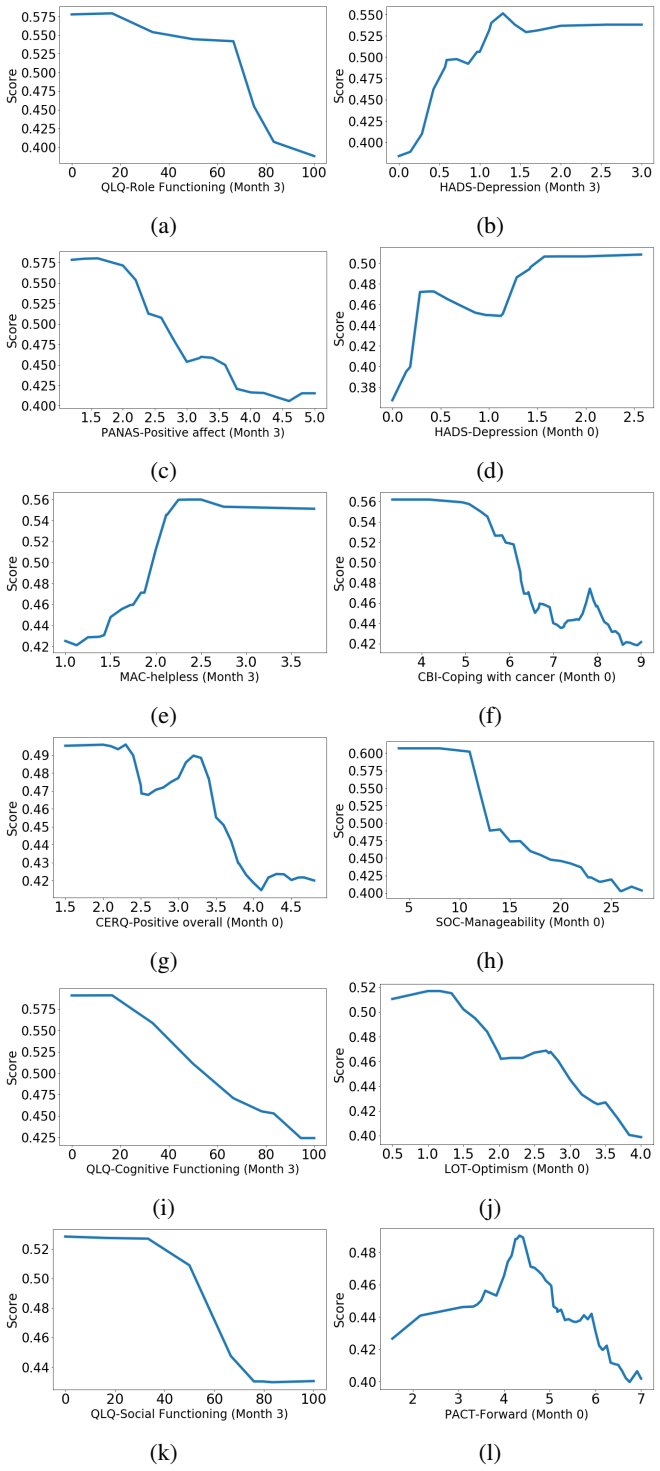


Figure C.10: Model of Experiment II. Partial dependence plots (PDP) for the 12 variables that had the highest SHAP values (SHapley Additive exPlanations). The Figure C.10 is related to the Figure C.9.

[14] M. Sprangers, M. Groenvold, J. I. Arraras, J. Franklin, A. te Velde, M. Muller, L. Franzini, A. Williams, H. De Haes, P. Hopwood, et al., The european organization for research and treatment of cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study., Journal of clinical oncology 14 (1996) 2756–2768.

[15] S. Simard, J. Savard, Screening and comorbidity of clinical levels of fear of cancer recurrence, Journal of Cancer Survivorship 9 (2015) 481–491.

[16] A. S. Zigmond, R. P. Snaith, The hospital anxiety and depression scale, Acta psychiatrica scandinavica 67 (1983) 361–370.

[17] E. R. Thompson, Development and validation of an internationally reliable short-form of the positive and negative affect schedule (panas), Journal of cross-cultural psychology 38 (2007) 227–242.

[18] V. Clarke, V. Braun, Thematic Analysis, volume 3, 2014, pp. 1947–1952. doi:10.1037/13620-004.

[19] H. Kondylakis, L. Koumakis, D. G. Katehakis, A. Kouroubali, K. Marias, M. Tsiknakis, P. G. Simos, E. Karademas, Developing a data infrastructure for enabling breast cancer women to bounce back, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), 2019, pp. 652–657. doi:10.1109/CBMS.2019.00134.

[20] G. Pettini, V. Sanchini, R. Pat-Horenczyk, B. Sousa, M. Masiero, C. Marzorati, V. Galimberti, E. Munzone, J. Mattson, L. Vehmanen, M. Utriainen, I. Roziner, R. Lemos, D. Frasquilho, F. Cardoso, A. J. Oliveira-Maia, E. Kolokotrini, G. S. Stamatakos, R.-L. Leskelä, I. Haavisto, J. Salonen, R. Richter, E. C. Karademas, P. Poikonen-Saksela, M. Ketti, Predicting effective adaptation to breast cancer to help women to bounce back: study protocol for a multicenter clinical pilot, JMIR Preprints. 29/10/2021:34564 (2021).

[21] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[22] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, Nature Machine Intelligence 2 (2020) 2522–5839.

[23] J. H. Friedman, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, The Annals of Statistics 29 (2001) 1189–232.