

Original Article

Comorbid insomnia among breast cancer survivors and its prediction using machine learning: a nationwide study in Japan

Taro Ueno¹, Daisuke Ichikawa¹, Yoichi Shimizu^{2,3}, Tomomi Narisawa², Katsunori Tsuji², Eisuke Ochi⁴, Naomi Sakurai⁵, Hiroji Iwata⁶, and Yutaka J. Matsuoka^{ID 2,*}

¹SUSMED, Inc, Tokyo, Japan, ²Division of Health Care Research, Behavioral Science and Survivorship Research Group, Center for Public Health Sciences, National Cancer Center Japan, Tokyo, Japan, ³Division of Nursing, National Cancer Center Hospital, Tokyo, Japan, ⁴Faculty of Bioscience and Applied Chemistry, Hosei University, Koganei, Tokyo, Japan, ⁵Cancer Solutions, Tokyo, Japan and ⁶Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, Japan

*For reprints and all correspondence: Division of Health Care Research, Behavioral Science and Survivorship Research Group, Center for Public Health Sciences, National Cancer Center Japan, 104-0045 5-1-1 Tsukiji, Chuo-ku, Tokyo, Japan. E-mail: matsuoka-psy@umin.ac.jp

Received 10 August 2021; Editorial Decision 6 October 2021; Accepted 12 October 2021

Abstract

Objective: Insomnia is an increasingly recognized major symptom of breast cancer which can seriously disrupt the quality of life during and many years after treatment. Sleep problems have also been linked with survival in women with breast cancer. The aims of this study were to estimate the prevalence of insomnia in breast cancer survivors, clarify the clinical characteristics of their sleep difficulties and use machine learning techniques to explore clinical insights.

Methods: Our analysis of data, obtained in a nationwide questionnaire survey of breast cancer survivors in Japan, revealed a prevalence of suspected insomnia of 37.5%. With the clinical data obtained, we then used machine learning algorithms to develop a classifier that predicts comorbid insomnia. The performance of the prediction model was evaluated using 8-fold cross-validation.

Results: When using optimal hyperparameters, the L2 penalized logistic regression model and the XGBoost model provided predictive accuracy of 71.5 and 70.6% for the presence of suspected insomnia, with areas under the curve of 0.76 and 0.75, respectively. Population segments with high risk of insomnia were also extracted using the RuleFit algorithm. We found that cancer-related fatigue is a predictor of insomnia in breast cancer survivors.

Conclusions: The high prevalence of sleep problems and its link with mortality warrants routine screening. Our novel predictive model using a machine learning approach offers clinically important insights for the early detection of comorbid insomnia and intervention in breast cancer survivors.

Key words: insomnia, breast cancer, machine learning

Introduction

Breast cancer is the leading cancer affecting women worldwide. In 2009, 60 000–70 000 new cases of breast cancer were reported in Japan (1). Advances in screening and treatment are improving survival times. With the 5-year survival rate as high as 93%, many Japanese women are now living as survivors of breast cancer. Longer survival times are drawing increasing attention to the impact of the disease and its treatment on long-term outcomes and health-related quality of life.

Insomnia is one of the most prevalent symptoms experienced by cancer patients (2). In Japan, the prevalence of insomnia is 14.6–22.3% among women in the general population (3–5), and prevalence is known to be higher in breast cancer survivors than in the general population (2,6–8). The pooled estimate for the prevalence of sleep disturbance is ~40%. In addition, insomnia can be a significant independent prognostic factor in breast cancer survivors (9–11). Despite this evidence, no studies have systematically investigated the prevalence of comorbid insomnia among breast cancer survivors in Japan.

The purpose of this study was to clarify the prevalence, severity and characteristics of insomnia in breast cancer survivors and to determine the clinical characteristics associated with the comorbidity. We also developed a classifier that predicts the comorbid insomnia using two machine learning algorithms, namely, the L2 penalized logistic regression model and the XGBoost model (12). Furthermore, we used the RuleFit algorithm to extract the hidden rules for segments at high risk of comorbid insomnia.

Materials and methods

Ethical considerations

In this study, we analyzed responses to the Athens Insomnia Scale (AIS) as part of a nationwide survey of Japanese breast cancer survivors (13). The study was approved by the institutional review board of the National Cancer Center, Japan (ID: 2018-295) and by the ethics committees of all 34 participating hospitals. Data were collected anonymously, and care was taken not to collect any identifying information.

Study design

Attending physicians in the outpatient clinics of the facilities handed out a set of materials containing explanatory documents and a survey form to eligible participants, who completed the survey independently and returned it by mail to the research office. Measurement items are described in the protocol paper (13). The following items were collected: background information, the Global Physical Activity Questionnaire, EuroQol 5 Dimension, the Japanese equivalent of WHO Health and Work Performance Questionnaire Short Form, the Cancer Fatigue Scale (CFS), the Concerns about Recurrence Scale, the AIS, the Common Terminology Criteria for Adverse Events (PRO-CTCAE) and the Resilience Scale. The total scores of each scale were used to develop the prediction models.

Participants and recruitment

Eligibility criteria were as follows: (i) diagnosis of primary breast cancer without distant metastasis, (ii) no recurrence, (iii) age ≥ 20 years, (iv) completion of initial treatments with curative intent aside from hormone therapy and (v) already informed of the diagnosis of breast

cancer. Participants who could not complete the self-reported questionnaire (written in Japanese) unaided were excluded. Participants who did not complete the AIS were excluded.

For participant recruitment, we selected 52 facilities accredited by the Japanese Breast Cancer Society that conducted more than 100 breast cancer surgeries in the period from April 2016 to March 2017. Based on the sampling method used in the public opinion survey on cancer countermeasures conducted by the Japan Cabinet Office, we set 22 stratified categories according to 11 districts and population size ($>200\ 000$ people and $<200\ 000$ people).

Comorbidity of insomnia

We used the AIS to assess the comorbidity of insomnia. The scale, which was created by the World Health Organization as part of the ‘World Sleep and Health Project’ (14,15), is an eight-item, self-administered psychometric instrument, with a total score of 24 points. An AIS score of 6 is the optimum cutoff based on the balance between sensitivity and specificity (16), and we considered a score of ≥ 6 to indicate suspected comorbid insomnia in this study.

Missing data imputation

The missing values were imputed using a regression model in which variables other than those to be imputed were used as variables (17). Logistic regression analysis was used for categorical variables, and multiple regression analysis was used for continuous variables.

Prediction models

All the data except for those with $>20\%$ of missing values were used to build the prediction models.

We used the L2 penalized logistic regression model (18) to realize increased stability while overcoming logistic regression’s shortcoming of degraded performance when features are strongly correlated (19).

Along with the baseline model, we used models obtained by machine learning using a gradient-boosting decision tree (GBDT) approach, which is an ensemble learning algorithm that combines base learners, such as decision trees and linear classifiers (20). We adopted the GBDT approach because of its superior predictive ability (21). GBDT gives a predictive model as an ensemble of decision trees and achieves high predictive ability with a differentiable loss function. GBDT requires tuning of parameter, such as the number of trees, shrinkage parameter and interaction depth, which was done by a grid search.

Segment extraction

By using the RuleFit algorithm, sparse linear models are learned, which include automatically detected interaction effects in the form of decision rules (22). The first step of the algorithm is extracting decision rules from original features, which is done by using random forests in the present study (23). The second step is learning a sparse linear model with the original features and new features based on the decision rules. In the study, we used the L2 penalized logistic regression model as a sparse linear model (24). These new features are reflected in the interactions between the original features.

Table 1. Demographic and medical characteristics of the breast cancer survivors in this study (N = 759 from 34 hospitals)

Characteristics	Responses, <i>n</i>	Missing
Age, mean (SD), y	59 (12)	12
Height, mean (SD), cm	156.6 (5.8)	21
Weight, mean (SD), kg	55.9 (9.4)	23
Highest education level, <i>n</i> (%)		7
Junior high school	47 (6)	
High school	308 (41)	
College or more	397 (53)	
Employment status, <i>n</i> (%)		8
Full- or part-time worker	419 (56)	
Not working or housewife	332 (44)	
Breast cancer stage		76
0	127 (19)	
I	286 (42)	
II	149 (22)	
III	69 (10)	
Other	52 (7)	
Treatment, <i>n</i> (%)		
Surgery	746 (99)	5
Radiotherapy	424 (58)	29
Chemotherapy	253 (37)	75
Hormone therapy	557 (78)	48
Time since surgery, <i>n</i> (%)		142
<6 months	42 (7)	
0.5–1.5 years	143 (23)	
1.5–3 years	217 (35)	
3–5 years	202 (33)	
5–10 years	13 (2)	
Medication for insomnia, <i>n</i> (%)		4
Everyday	45 (6)	
Sometimes	38 (5)	
None	672 (89)	

SD, standard deviation.

Statistical analysis

To assess the predictive ability of each machine learning algorithm, we used an 8-fold external cross-validation procedure. For cross-validation, we used subject-wise data splitting rather than record-wise data splitting because identity confounding has been reported with the latter (25,26). Then, to assess the predictive accuracy of our developed classifier, we used receiver-operating characteristic curve analysis, where the area under the curve (AUC) represented the ability to predict comorbid insomnia (27,28). In the medical field, AUC of a prediction model is considered to have high accuracy when it is ≥ 0.9 , moderate accuracy when it is ≥ 0.7 but < 0.9 and low accuracy when it is ≥ 0.5 but < 0.7 . In this study, moderate accuracy or higher was considered to indicate a suitable model. The importance of variable in class discrimination in the predictive model was assessed using the mean decrease in gain. The Python libraries pandas (version 1.0.5), xgboost (version 0.9) and scikit-learn (version 0.21.2) were used for data handling and building and evaluating the prediction models.

Results

Prevalence of insomnia in breast cancer survivors

In total, 791 individuals from 34 hospitals participated in the nationwide survey, and we collected data from 759 participants who

returned the AIS questionnaire. Characteristics of these participants are shown in Table 1.

Overall, 284 participants (37.4%) were assessed as having suspected comorbid insomnia, 83 of whom (11%) took medication for insomnia. An AIS score of 6 is the optimum cutoff based on the balance between sensitivity and specificity (16) and we considered a score of ≥ 6 to indicate suspected comorbid insomnia in this study. Characteristics of each questionnaire in AIS are shown in Table 2.

Generation of the predictive model for comorbid insomnia by machine learning

The L2 penalized logistic regression model (Fig. 1) and the XGBoost model (Fig. 2) were used to develop classifiers for comorbid insomnia in breast cancer survivors based on questionnaire surveys and had predictive accuracy of 71.5 and 70.6% for the presence of suspected insomnia, giving AUCs of 0.76 and 0.75, respectively.

Importance of variables in the predictive model

We then investigated the importance of variables in the optimal predictive model obtained by machine learning. Figure 3 shows the ranking of variable importance in the L2 penalized logistic regression model. General fatigue determined on the CFS (29) was the most important variable for predicting comorbid insomnia, followed

Table 2. Results for the clinical characteristics assessed using the Athens Insomnia Scale

Sleep induction, <i>n</i> (%)	
No problem	459 (60.5)
Slightly delayed	217 (28.6)
Markedly delayed	60 (7.9)
Very delayed or did not sleep at all	23 (3.0)
Awakening during the night, <i>n</i> (%)	
No problem	466 (61.4)
Minor problem	243 (32.0)
Considerable problem	44 (5.8)
Serious problem or did not sleep at all	6 (0.8)
Final awakenings earlier than desired, <i>n</i> (%)	
Not earlier	390 (51.4)
A little earlier	297 (39.1)
Markedly earlier	59 (7.8)
Much earlier or did not sleep at all	13 (1.7)
Total sleep duration, <i>n</i> (%)	
Sufficient	306 (40.3)
Slightly insufficient	362 (47.7)
Markedly insufficient	83 (10.9)
Very unsatisfactory or did not sleep at all	8 (1.1)
Overall quality of sleep, <i>n</i> (%)	
Satisfactory	299 (39.4)
Slightly unsatisfactory	364 (48.0)
Markedly unsatisfactory	86 (11.3)
Very unsatisfactory or did not sleep at all	10 (1.3)
Sense of well-being during the day, <i>n</i> (%)	
Normal	512 (67.5)
Slightly decreased	213 (28.1)
Markedly decreased	30 (3.9)
Very decreased	4 (0.5)
Functioning during the day, <i>n</i> (%)	
Normal	473 (62.3)
Slightly decreased	231 (30.5)
Markedly decreased	51 (6.7)
Very decreased	4 (0.5)
Sleepiness during the day, <i>n</i> (%)	
None	147 (19.4)
Mild	547 (72.1)
Considerable	62 (8.1)
Intense	3 (0.4)
Total, <i>n</i> (%)	
AIS < 6	475 (62.6)
AIS ≥ 6	284 (37.4)

by physical fatigue and cognitive fatigue. In addition, high QOL measured on the EuroQol Five-Dimensional Questionnaire (30) and resilience measured on the 14-item Resilience Scale (31,32) less strongly related to comorbid insomnia. Figure 4 shows the ranking of variable importance in the XGBoost model, where general fatigue was again ranked as the most important variable for prediction of comorbid insomnia.

Segment extraction of patients with insomnia using the RuleFit algorithm

We further used the RuleFit algorithm (22) to discover the hidden rules that may be predictive of the risk of comorbid insomnia from among a number of potential candidates. Figure 5 shows results for the population segment classified by the RuleFit algorithm as having high risk based on the following rules: presence of depressive

symptoms on the Patient-Reported Outcomes version of the PRO-CTCAE, cognitive fatigue score > 4.5 on the CFS, and resilience score < 78.5 on the 14-item Resilience Scale Short Version (RS14). The prevalence of insomnia in this segment was high at 73%.

Figure 6 shows the results for the population segment classified by the RuleFit algorithm as having low risk of comorbid insomnia based on the following rules: RS14 score > 62.5, a higher EQ5D score indicating higher QOL and no decrease in physical activity. The prevalence of insomnia in the segment was only 20%.

Discussion

Comorbid insomnia is known to affect quality of life and prognosis in breast cancer survivors. In this study, using nationwide survey data, we have shown for the first time a high prevalence of insomnia in

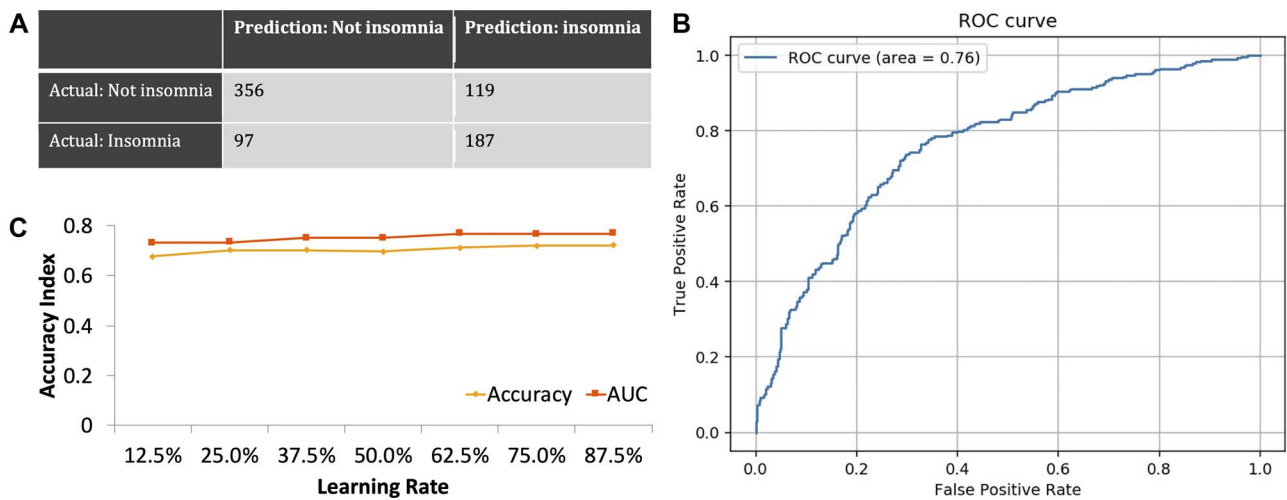


Figure 1. (a) Confusion matrix for the L2 penalized logistic regression model. (b) Receiver-operating characteristic (ROC) curve for predicting the comorbidity of insomnia based on the optimal predictive model developed using the L2 penalized logistic regression model. Area under the curve (AUC) for comorbid insomnia is 0.76. (c) Relationship between accuracy and the amount of training data in the L2 penalized logistic regression model.

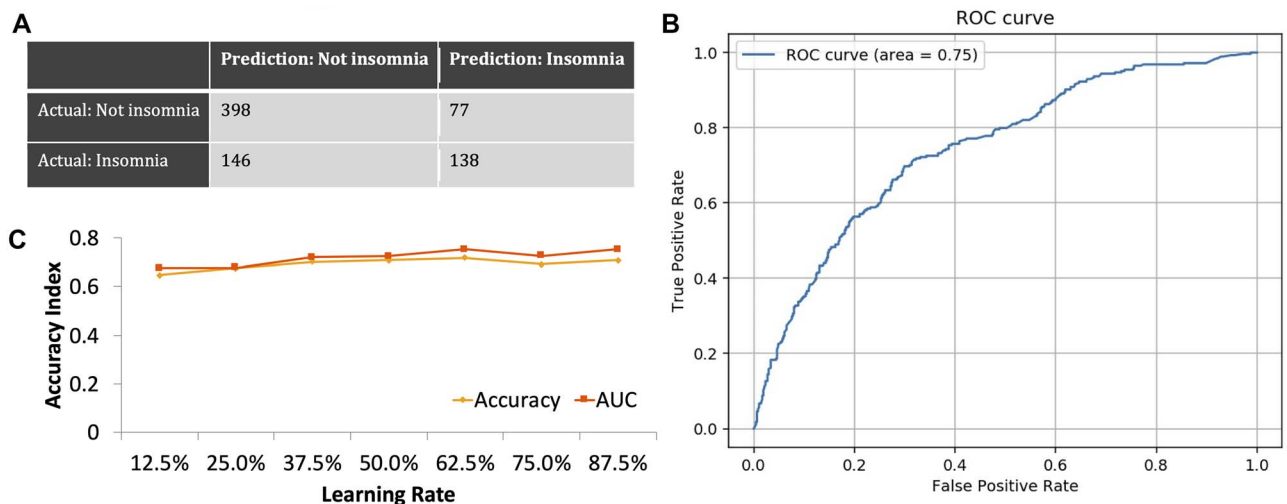


Figure 2. (a) Confusion matrix for the XGBoost model. (b) ROC curve for predicting the comorbidity of insomnia based on the optimal predictive model developed using the XGBoost model. AUC for comorbid insomnia is 0.75. (c) Relationship between accuracy and the amount of training data in the XGBoost model.

breast cancer survivors in Japan. Based on AIS responses, the prevalence of suspected comorbid insomnia was as high as 37.5%. The previous reports showed the prevalence of sleep disturbance in breast cancer is ~40%. The prevalence of insomnia among breast cancer patients in Japan is comparable with the previous reports which is higher than that of general population. The participants tended to complain of symptoms such as daytime sleepiness, insufficient sleep duration and unsatisfactory quality of sleep.

Using the survey data collected, we also used machine learning algorithms, namely, the L2 penalized logistic regression model and the XGBoost model, to develop classifiers that successfully predicted comorbid insomnia in breast cancer survivors. The ranking of variable importance revealed that fatigue, QOL and resilience were predictive of both risk of comorbid insomnia and protection against it. These variables were selected in both the logistic regression model and the XGBoost model, so the results were consistent. The impact of insomnia on QOL is widely recognized and the relationship between insomnia, fatigue and resilience has also been reported (33,34).

We further extracted segments with high risk of comorbid insomnia using the RuleFit algorithm. The algorithm identified cancer-related fatigue, health-related QOL and resilience as important rules. These results suggest that, it might be helpful in clinical practice to focus on patients with cancer-related fatigue, low health-related QOL and low resilience for early detection of comorbid insomnia and intervention in breast cancer survivors, and it may be important to assess these factors. In addition, self-reported change in physical activity was also selected as an important rule. Patients whose physical activity does not decrease after diagnosis of breast cancer tend to be less likely to have insomnia. Growing evidence suggests that exercise may play a role in maintaining and improving common cancer-related health outcomes, and multiple international organizations have issued guidelines recommending high levels of physical activity in cancer survivors (35,36). In addition, the adverse effects of decreased physical activity during the COVID-19 pandemic on long-term outcomes in cancer patients have been noted (37). The need for recently developed home-based exercise programs for

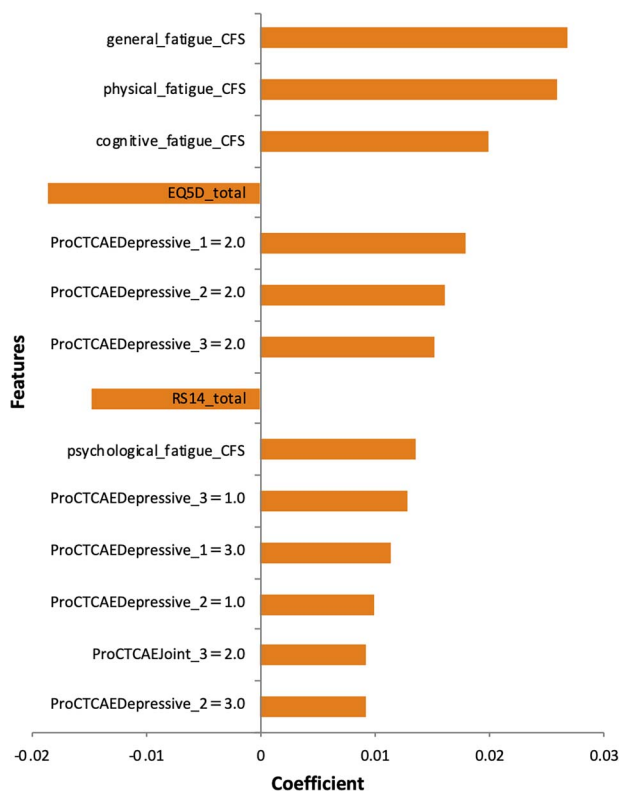


Figure 3. Ranking of variable importance for predicting the comorbid insomnia based on the optimal predictive model developed using the L2 penalized logistic regression model. Important variables are fatigue scores assessed by the Cancer Fatigue Scale (CFS), total QOL score assessed using the EuroQoL Five-Dimensional Questionnaire and total resilience score assessed using the 14-item Resilience Scale Short Version. The questions on the Patient-Reported Outcomes version of the PRO-CTCAE listed are as follows. PRO-CTCAEDepressive_1: frequency of discouragement. PRO-CTCAEDepressive_2: severity of discouragement. PRO-CTCAEDepressive_3: interference by discouragement. PRO-CTCAEjoint_3: interference by pain in joint.

breast cancer patients (38,39) is expected to continue increasing in the future.

Limitation

Our study has some limitations. First, we used cross-sectional data, so we cannot determine the direction of causality in the results. Although some clinical variables such as fatigue, resilience, quality of life and physical activity are associated with comorbid insomnia in breast cancer survivors, we could not identify the risk factors. To solve the problem, several methods have been proposed for discovering causal relationships in cross-sectional studies (13). Second, we could not distinguish the effect of cancer treatments on insomnia. A previous study has indicated that cancer treatments, such as chemotherapy and radiotherapy, cause worsening of insomnia symptoms (40). Further investigation by subgroup analysis in a larger population may provide additional clinical value.

Conclusion

In summary, our findings are consistent with the high prevalence of sleep problems in breast cancer survivors. A novel predictive

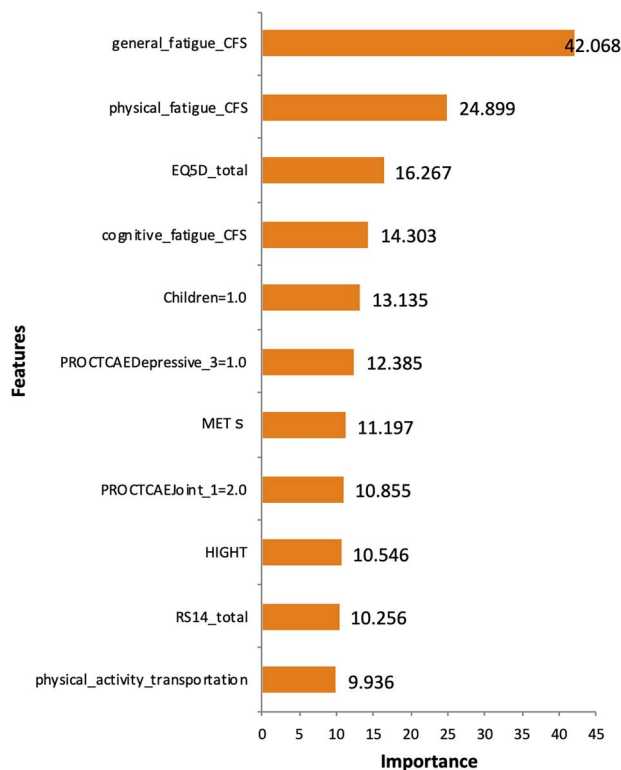


Figure 4. Ranking of variable importance for predicting the comorbidity of insomnia based on the optimal predictive model developed using the XGBoost model. Important variables are fatigue scores assessed by the CFS and total QOL score assessed using the EuroQoL Five-Dimensional Questionnaire. The questions on the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) listed are as follows. PRO-CTCAEDepressive_3: interference by discouragement. PRO-CTCAEjoint_1: frequency of pain in joint.

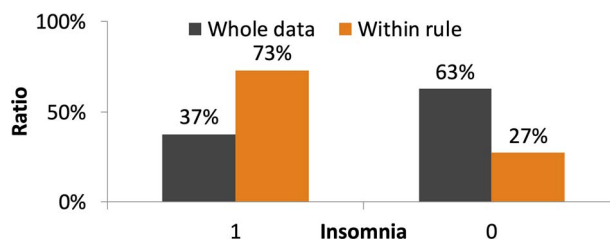


Figure 5. Results for the population segment classified as having high risk of comorbid insomnia by the RuleFit algorithm. The rules extracted by the RuleFit algorithm are as follows. PRO-CTCAE Depressive_2 (severity of discouragement in PRO-CTCAE) is >0.5. Cognitive fatigue score on the CFS is >4.5. Total score on the 14-item Resilience Scale is <78.5.

model obtained by a machine learning approach provides insight into clinical practice for early detection of comorbid insomnia and intervention in breast cancer survivors.

Authors' contributions

All authors were responsible for the acquisition, analysis or interpretation of data and the critical revision of the manuscript for important intellectual content. T.U., D.I. and Y.J.M. took care of the drafting of the manuscript. T.U. and D.I. were in charge of the statistical analysis. Y.S. and T.N. were in charge of administrative,

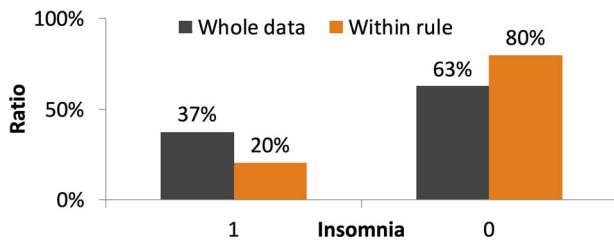


Figure 6. Results for the population segment classified as having low risk of comorbid insomnia by the RuleFit algorithm. The rules extracted by the RuleFit algorithm are as follows. Total score on the 14-item Resilience Scale is >62.5. Total score on the EuroQoL Five-Dimensional Questionnaire is >0.869. Amount of physical activity is not decreased after diagnosis of breast cancer.

technical or material support. N.S. was responsible for the patient and public involvement. Y.J.M. was in charge of the supervision.

Acknowledgements

The authors wish to thank Prof Uchitomi (National Cancer Center Japan), Dr Shimazu (National Cancer Center Japan) and Mr Motohashi (SUSMED) for their generous support and helpful advice. The authors also wish to thank Ms Akutsu for her efforts in data management. This article is based on results obtained from a project, P20006, commissioned by the New Energy and Industrial Technology Development Organization.

Funding

This study was supported by the National Cancer Center Research and Development Fund (30-A-17).

Conflict of interest statement

Ochi has received research support from Nippon Suisan Kaisha. Ueno and Ichikawa are presidents and shareholders of SUSMED, Inc. Matsuoka has received speaker fees from Suntory Wellness, Pfizer, Mochida, Eli Lilly and Morinaga Milk, and Cimic and is conducting collaborative research with SUSMED. All other authors declare that they have no competing interests regarding this work.

Role of the funder/sponsor

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

- Hori M, Matsuda T, Shibata A, et al. Cancer incidence and incidence rates in Japan in 2009: a study of 32 population-based cancer registries for the Monitoring of Cancer Incidence in Japan (MCIJ) project. *Jpn J Clin Oncol* 2015;45:884–91.
- Savard J, Morin CM. Insomnia in the context of cancer: a review of a neglected problem. *J Clin Oncol* 2001;19:895–908.
- Kim K, Uchiyama M, Okawa M, Liu X, Ogihara R. An epidemiological study of insomnia among the Japanese general population. *Sleep* 2000;23:41–7.
- Itani O, Kaneita Y, Munezawa T, et al. Nationwide epidemiological study of insomnia in Japan. *Sleep Med* 2016;25:130–8.

- Doi Y, Minowa M, Okawa M, Uchiyama M. Prevalence of sleep disturbance and hypnotic medication use in relation to sociodemographic factors in the general Japanese adult population. *J Epidemiol* 2000;10:79–86.
- Leysen L, Lahousse A, Nijs J, et al. Prevalence and risk factors of sleep disturbances in breast cancer survivors: systematic review and meta-analyses. *Support Care Cancer* 2019;27:4401–33.
- Desai K, Mao JJ, Su I, et al. Prevalence and risk factors for insomnia among breast cancer patients on aromatase inhibitors. *Support Care Cancer* 2013;21:43–51.
- Bower JE. Behavioral symptoms in patients with breast cancer and survivors. *J Clin Oncol* 2008;26:768–77.
- Palesh O, Aldridge-Gerry A, Zeitzer JM, et al. Actigraphy-measured sleep disruption as a predictor of survival among women with advanced breast cancer. *Sleep* 2014;37:837–42.
- Trudel-Fitzgerald C, Zhou ES, Poole EM, et al. Sleep and survival among women with breast cancer: 30 years of follow-up within the Nurses' Health Study. *Br J Cancer* 2017;116:1239–46.
- Collins KP, Geller DA, Antoni M, et al. Sleep duration is associated with survival in advanced cancer patients. *Sleep Med* 2017;32:208–12.
- Inoue T, Ichikawa D, Ueno T, et al. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Rep* 2020;1:8–16.
- Shimizu Y, Tsuji K, Ochi E, et al. Study protocol for a nationwide questionnaire survey of physical activity among breast cancer survivors in Japan. *BMJ Open* 2020;10:e032871.
- Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. *J Psychosom Res* 2000;48:555–60.
- Okajima I, Nakajima S, Kobayashi M, Inoue Y. Development and validation of the Japanese version of the Athens Insomnia Scale. *Psychiatry Clin Neurosci* 2013;67:420–5.
- Soldatos CR, Dikeos DG, Paparrigopoulos TJ. The diagnostic validity of the Athens Insomnia Scale. *J Psychosom Res* 2003;55:263–7.
- van der Heijden GJMG, Donders ART, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006;59:1102–9.
- Cessie SL, Van Houwelingen JC. Ridge estimators in logistic regression. *J R Stat Soc Ser C Appl Stat* 1992;41:191.
- Liu Z, Shen Y, Ott J. Multilocus association mapping using generalized ridge logistic regression. *BMC Bioinformatics* 2011;12:384.
- Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38:367–78.
- Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neuro-robot* 2013;7:21.
- Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat* 2008;2:916–54.
- Breiman L, Last M, Rice J. Random forests: finding quasars. *Statistical Challenges in Astronomy*. Springer-Verlag, 2006; 243–54.
- Tibshirani R. Regression shrinkage and selection via the lasso. *J R I State Dent Soc* 1996;58:267–88.
- Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. *Gigascience* 2017;6:1–9.
- Chaibub Neto E, Pratap A, Perumal TM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ Digit Med* 2019;2:99.
- Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
- Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003;29:1043–51.
- Okuyama T, Akechi T, Kugaya A, et al. Development and validation of the cancer fatigue scale: a brief, three-dimensional, self-rating scale for assessment of fatigue in cancer patients. *J Pain Symptom Manage* 2000;19:5–14.

30. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
31. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. *J Nurs Meas* 1993;1:165–78.
32. Nishi D, Uehara R, Kondo M, Matsuoka Y. Reliability and validity of the Japanese version of the Resilience Scale and its short version. *BMC Res Notes* 2010;3:310.
33. Bean HR, Diggins J, Ftanou M, Weihs KL, Stanton AL, Wiley JF. Insomnia and fatigue symptom trajectories in breast cancer: a longitudinal cohort study. *Behav Sleep Med* Published online January 20 2021;19:814–27.
34. Palagini L, Moretto U, Novi M, et al. Lack of resilience is related to stress-related sleep reactivity, hyperarousal, and emotion dysregulation in insomnia disorder. *J Clin Sleep Med* 2018;14:759–66.
35. Schmitz KH, Campbell AM, Stuver MM, et al. Exercise is medicine in oncology: engaging clinicians to help patients move through cancer. *CA Cancer J Clin* 2019;69:468–84.
36. Rock CL, Doyle C, Demark-Wahnefried W, et al. Nutrition and physical activity guidelines for cancer survivors. *CA Cancer J Clin* 2012;62:243–74.
37. Avancini A, Trestini I, Tregnago D, et al. Physical activity for oncological patients in COVID-19 era: no time to relax. *JNCI Cancer Spectr* 2020;4:kaa071.
38. Tsuji K, Ochi E, Okubo R, et al. Effect of home-based high-intensity interval training and behavioural modification using information and communication technology on cardiorespiratory fitness and exercise habits among sedentary breast cancer survivors: habit-B study protocol for a randomised controlled trial. *BMJ Open* 2019;9:e030911.
39. Hirano T, Motohashi T, Okumura K, et al. Data validation and verification using blockchain in a clinical trial for breast cancer: regulatory sandbox. *J Med Internet Res* 2020;22:e18938.
40. Savard J, Ivers H, Savard M-H, Morin CM. Cancer treatments and their side effects are associated with aggravation of insomnia: results of a longitudinal study. *Cancer* 2015;121:1703–11.