



Using machine learning to predict individual patient toxicities from cancer treatments

Katherine Marie Cole¹ · Mark Clemons^{1,2} · Sharon McGee¹ · Mashari Alzahrani¹ · Gail Larocque³ · Fiona MacDonald³ · Michelle Liu² · Gregory R. Pond⁴ · Lucy Mosquera⁵ · Lisa Vandermeer² · Brian Hutton⁶ · Ardelle Piper⁷ · Ricardo Fernandes⁸ · Khaled El Emam^{5,9}

Received: 23 November 2021 / Accepted: 16 May 2022 / Published online: 25 May 2022
© The Author(s) 2022

Abstract

Purpose Machine learning (ML) is a powerful tool for interrogating datasets and learning relationships between multiple variables. We utilized a ML model to identify those early breast cancer (EBC) patients at highest risk of developing severe vasomotor symptoms (VMS).

Methods A gradient boosted decision model utilizing cross-sectional survey data from 360 EBC patients was created. Seventeen patient- and treatment-specific variables were considered in the model. The outcome variable was based on the Hot Flush Night Sweats (HFNS) Problem Rating Score, and individual scores were dichotomized around the median to indicate individuals with high and low problem scores. Model accuracy was assessed using the area under the receiver operating curve, and conditional partial dependence plots were constructed to illustrate relationships between variables and the outcome of interest.

Results The model area under the ROC curve was 0.731 (SD 0.074). The most important variables in the model were as follows: the number of hot flashes per week, age, the prescription, or use of drug interventions to manage VMS, whether patients were asked about VMS in routine follow-up visits, and the presence or absence of changes to breast cancer treatments due to VMS. A threshold of 17 hot flashes per week was identified as being more predictive of severe VMS. Patients between the ages of 49 and 63 were more likely to report severe symptoms.

Conclusion Machine learning is a unique tool for predicting severe VMS. The use of ML to assess other treatment-related toxicities and their management requires further study.

Keywords Vasomotor symptoms · Hot flashes · Breast cancer · Machine learning · Artificial intelligence · Survivorship

Introduction

Machine learning (ML) is an “application of artificial intelligence (AI) that allows computer systems to automatically learn from experience without explicit programming” [1].

Highlights

- Machine learning (ML) is a tool to learn complex relationships between variables.
- We used ML to identify predictors of severe hot flashes in breast cancer patients.
- Hot flash frequency and age were important variables predicting severe symptoms.
- ML may aid in early identification of patients at risk of severe vasomotor symptoms.

✉ Khaled El Emam
kelemam@cheo.on.ca

Extended author information available on the last page of the article

Supervised learning models used in the oncology sphere, such as ensemble models (including gradient boosted trees) [2, 3], create predictions based on the aggregation of multiple individual models, and as such, can capture more complex relationships among variables [4]. In the setting of breast cancer, ML has been successfully applied for models examining risk of developing breast cancer [5], as well as breast cancer prognosis [6].

Vasomotor symptoms (VMS), including hot flashes and night sweats, are a common sequelae of systemic therapies for breast cancer [7, 8], and are the most common reason for discontinuation of potentially curative treatment [9–12]. As adjuvant therapies may be prescribed for up to 10 years [13, 14], appropriate management of VMS is essential to improve patient quality of life and breast cancer outcomes. Despite randomized trials showing effective interventions for the management of VMS [15], and recommendations

from guideline groups that patients be routinely asked about VMS [16, 17], it is evident that VMS remain poorly managed [18, 19]. There are many reasons for this finding, including health care provider (HCP) uncertainty regarding non-pharmacologic strategies for VMS [20], the variability in frequency and severity of VMS in EBC patients [21], heterogeneity in the efficacy of interventions across the EBC population [15], and the absence of guidelines on tailoring treatments to individual patients.

Given the heterogenous nature of the breast cancer population, the variability in VMS severity, and the multiple interventions for VMS management, new strategies for implementing patient centered care are required. The objective of the current study was to create a gradient boosted decision tree (GBDT) model to identify factors that predict patients at risk of severe VMS as defined by the HFNS problem rating score.

Materials and methods

Survey dataset

We recently conducted a survey in patients with EBC who were experiencing VMS [22] (A.1). The objective of the survey was to determine patient perspectives on the frequency and severity of VMS, and to determine the effectiveness of previously tried interventions for this problem. After collecting demographic data including menopausal status and previous systemic therapies for breast cancer, patients rated the frequency and severity of their VMS using the validated Hot Flush Night Sweats (HFNS) Problem Rating Score [23]. This is a composite score that takes the mean of the problem, distress, and disruption to daily life caused by VMS. Each of the 3 variables is classified on a 10-point scale, with “1” representing low severity, and “10” representing significant severity. The final section of the survey asked patients to report on interventions that they had received for their VMS and to rate the effectiveness of these treatments. Patients were also asked how they themselves would define effective control of their VMS, and to provide feedback on specific types of interventions that they would be willing to try in the future.

Data preparation

The outcome variable of interest was hot flash severity, as per the HFNS Problem Rating Score [23]. This outcome variable was chosen as it has been previously validated as a tool for assessing severity of VMS in the breast cancer population [24]. The three items in the HFNS score had a Cronbach alpha of 0.91 on our dataset, indicating high internal consistency [25]. In previous studies of women experiencing

problematic or severe VMS, mean HFNS problem scores ranged from 5.88 to 6.3 (SD 2.2–2.6) [19, 24, 26]. Assuming that the distribution of HFNS problem scores follows a normal distribution, an integer cut-off score of 4 would capture approximately 84% of individuals with severe VMS (mean minus 1 standard deviation). A test of the normality assumption using the Kolmogorov–Smirnov (KS) statistic found $p = 0.2263$ (see Appendix C). Therefore, using this cut-off, we dichotomized patients into low and high severity scores. This coincides with the median, which results in a balanced dataset.

Variables were extracted from the patient survey questions for inclusion/exclusion in the ML model (Table B.1). Seventeen questions unrelated to the research question were excluded. These included one patient eligibility question, one patient feedback question, two questions relevant to patient treatment preferences, and four questions relevant to the perceived efficacy of interventions for VMS. To avoid redundancy, one additional question pertaining to menopausal status and two additional bothersome symptom questions were excluded. A question asking about previous treatments for VMS was removed as this was further explored in subsequent questions. Six questions relevant to hot flash severity were removed, including two questions pertaining to coping and control. While the two latter variables were part of the HFNS tool, these variables were found to be less reliable in the original validation study and were removed [23].

The remaining questions relevant to the analysis were as follows: patient age, previous systemic therapies for breast cancer, current menopausal status, hot flash/night sweats frequency, bothersome symptoms associated with VMS, changes made to breast cancer treatments due to VMS, recommendation of prescription or complementary and alternative medicine (CAM) interventions for VMS, and referral to a dedicated menopause clinic (Table 1). Responses to each question were assigned as binary or continuous variables. The most bothersome symptoms associated with VMS had 16 response options that were converted into binary variables. Variables that received less than 10% of valid responses were removed, which included 10 symptom variables (Table B.2). These variables had little variation, and therefore, were poor predictors of the outcome.

Analysis

The software used for analysis was R version 3.6, with main model construction using the *lightgbm* package (version 3.0.0). A GBDT was trained to predict the hot flash problem outcome [27, 28]. The basic process is to fit the data to a large number of trees, with each tree incorporating the prediction errors from the previous tree in a sequence as input. This approach has proven to have high prediction accuracy relative to linear models and deep learning models (see Appendix

Table 1 Variables included in the creation of machine learning model

Variable	Definitions
1. Age**	Age in years
2. Menopausal status	Current self-reported menopausal status
3. Assessment of VMS	Whether the patient is asked about/assessed for symptoms of hot flashes by HCP during clinic visits
4. Hot flashes per week	The number of hot flashes in a week that occurred in the past week
Bothersome symptoms associated with VMS	Ranking of most bothersome symptom associated with VMS
5. Feeling extremely hot/sweaty	
6. Redness of my face/chest	
7. Feeling chills/clammy after hot flashes have passed	
8. Waking up at night/difficulty sleeping	
9. Irritability	
10. Memory problems	
11. Endocrine therapy	Endocrine therapy treatment for breast cancer (e.g., tamoxifen, letrozole/Femara, anastrozole/Arimidex)
12. Ovarian function suppression (OFS)	Ovarian function suppression treatment for breast cancer (leuprolide/Lupron, goserelin/Zoladex, oophorectomy)
13. Chemotherapy	Chemotherapy treatment for breast cancer
14. Change to BC treatment	Changes made to breast cancer treatment due to hot flashes
15. Drug treatments for VMS	Drugs that were prescribed or patient tried any prescription drug
16. CAM therapies for VMS	Complementary treatments prescribed, recommended, or tried by the patient
17. Referral to menopause clinic	Patient referred or seen by a gynecologist or dedicated menopause clinic to assist in managing hot flashes
Removed due to high correlation with “hot flashes per week” variable	
Hot flashes per day	The number of hot flashes per day that occurred in the past week
Nocturnal sweats per night	The number of times per night that nocturnal hot flashes (night sweats) woke you up in the last week
Nocturnal sweats per week	The number of times per week that nocturnal hot flashes (night sweats) woke you up in the last week

VMS = vasomotor symptoms, HCP = healthcare provider, BC = Breast cancer, CAM = complementary and alternative medicine

** All variables converted to binary form with the exception of “age” and “number of VMS per day/week and night sweats per day/week”

C), is robust in datasets that have missing values, and can model interactions and non-linear relationships without having them specified a priori. We used a nested cross-validation approach for model development and accuracy estimation (see Appendix C for further details). Bayesian optimization was used for hyperparameter selection [29]. Each combination of hyperparameters was evaluated using fivefold cross-validation on the training dataset during tuning (inner loop). Generalization accuracy scores were calculated using tenfold cross-validation whereby a trained model was used to predict the probability for unseen cases (outer loop).

The generalization accuracy of the predictions from the model was assessed using two metrics. The first was the area under the receiver operating characteristic curve (AUROC) [30], which is commonly used to evaluate the performance of binary classifiers in machine learning. It is a measure of the area under a plot of the false positive rate against recall. For binary classification tasks, an AUROC of 0.5 is the expected performance of a random classifier, while

an AUROC of 1 is the expected performance of a perfect classifier. The second complementary metric which focuses on predictions of the positive class is the area under the precision-recall curve (AUPRC) [31]. The AUPRC value of a random classifier is the rate of the positive class [32] which is 0.5, with 1 being the perfect classifier.

Model interpretation

We used two approaches to interpret the overall GBDT model: “permute and re-learn” to determine variable importance, and conditional partial dependence plots to visualize the functional form of the relationships between the predictors and the outcome.

Permutation of variables is a technique that involves the shuffling of variables to evaluate their impact on the accuracy of prediction models [33], and is commonly used to evaluate variable importance in machine learning models. There is evidence that permuting a variable

is biased towards predictors that are correlated with other predictors that have many categories [34, 35]. Therefore, we instead permuted and reconstructed the model from the training data within each cross-validation iteration (permute and retrain the model), and then computed the difference in prediction accuracy between the original and permuted models [36, 37]. This difference gives the gain in accuracy by including a particular variable in the model. This will give us more reliable variable importance measures.

To illustrate the functional relationships between the most important predictor variables and the outcome of interest, conditional partial dependence plots were constructed [38]. Regular partial dependence plots are commonly used but have been subject to criticism as not all observations in the plot may plausibly be observed, leading to poor predictions due to extrapolation [37]. Conditional partial dependence plots aim to minimize extrapolation by calculating partial dependence within conditional subgroups, and then pools the results across subgroups. They also isolate the effect of a variable so we can view its impact, within the model, on the outcome.

Results

Description of the data

The original survey dataset comprised 383 patients. Ten patients were excluded, as these individuals were not experiencing VMS at the time of survey completion, totalling 373 patients who fulfilled eligibility (Table 2). An additional 13 patients were excluded from the gradient boosted trees analysis, as incomplete data was available for the HFNS problem rating score, totalling 360 patients included in the final analysis. These patients were recruited from the Ottawa Hospital Cancer Centre and the London Regional Cancer Centre, Ontario, Canada. As responses to all questions were optional, some variables had fewer than 360 responses.

The mean age of the participants was 56.3 (SD 10.5) (Table 2). The majority of women were post-menopausal at the time of survey completion, and treatment received included endocrine therapy ($n = 319/360$, 88.6%) and chemotherapy ($n = 205/360$, 56.9%). A minority of patients reported receiving drug interventions ($n = 112/248$, 31.1%) or CAM interventions ($n = 62/298$, 17.2%) to manage their VMS, and 18% of patients reported changes in their breast cancer therapy secondary to VMS.

Table 2 Baseline characteristics and summary statistics for patient sample, excluding symptom variables

	Number of respondents	Mean \pm SD
Age		
Mean age	360*	56.3 \pm 10.5
18–24		<i>N</i> (%)
25–39		1 (0.3%)
40–59		23 (6.4%)
60–74		204 (56.7%)
75+		117 (32.5%)
15 (4.2%)		
Hot flashes per week	Number of respondents	Median (IQR)
	295**	15 (IQR 5–35)
Binary variables included in the model	Yes or 1 count (% of total sample)	No or 0 count (% of total sample)
Menopausal at time of survey completion**	198 (55.0%)	130 (36.1%)
Routinely asked about VMS in clinic**	210 (58.3%)	129 (35.8%)
Endocrine therapy	319 (88.6%)	41 (11.4%)
Chemotherapy	205 (56.9%)	155 (43.1%)
Ovarian suppression	70 (19.4%)	290 (80.6%)
Change to BC treatment secondary to VMS	66 (18.3%)	294 (81.7%)
Drug treatments for VMS [‡]	112 (31.1%)	248 (68.9%)
CAM treatments for VMS [‡]	62 (17.2%)	298 (82.8%)
Referral to menopause clinic** [‡]	24 (6.7%)	335 (93.1%)

*Sample size fewer than 373 patients, as “I don’t know” and missing values were not counted

**Responses not provided by all survey participants

[‡]Included patients responding “yes” to any question containing this variable as an option

The four variables pertaining to frequency of hot flashes and night sweats (hot flashes per day, hot flashes per week, night sweats per day, night sweats per week) were highly

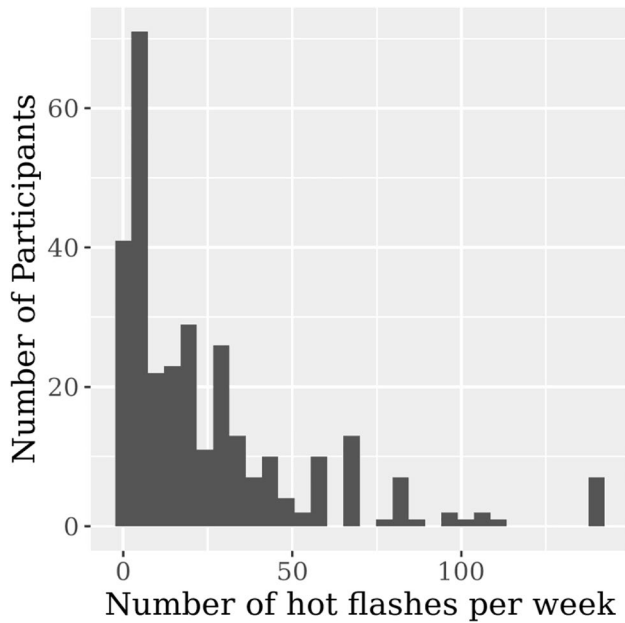


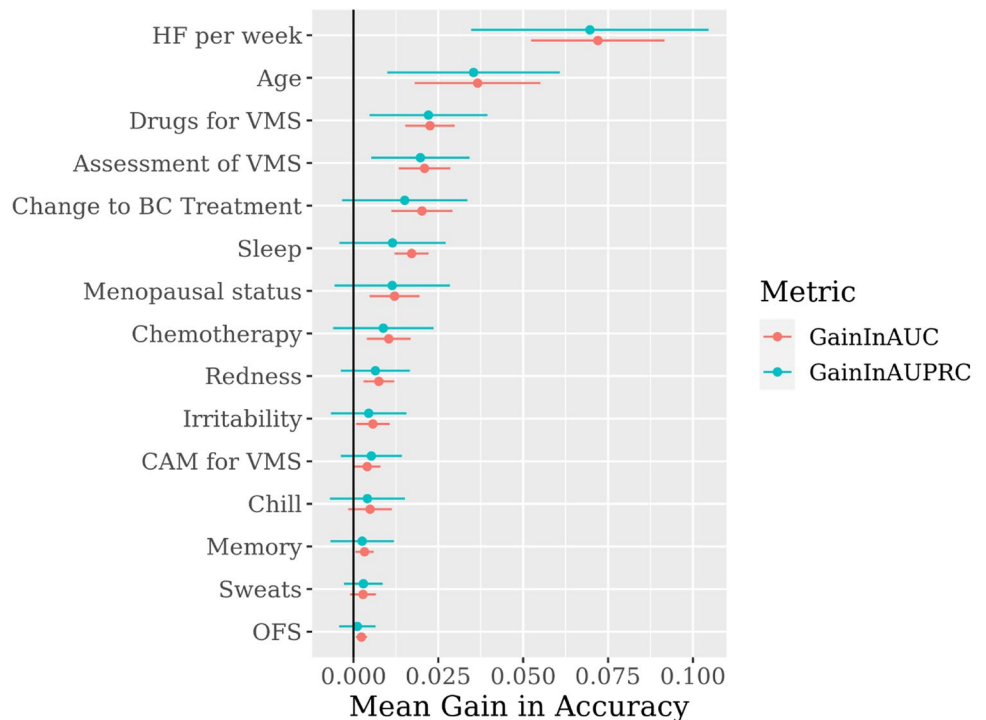
Fig. 1 Distribution of the number of vasomotor symptoms per week. For this variable, counts at 140 hot flashes per week were topcoded, as there were very few observations above that threshold, and thus represented outliers in the population. This was done for the purposes of data presentation only, and affected a total of five patients (5/295, 1.7%)

correlated ($r > 0.7$) (Table B.3). To avoid redundancy in the features that were used in modelling, we utilized the “hot flashes per week” variable only. The number of hot flashes per week is a commonly utilized endpoint in clinical trials of VMS in breast cancer patients, leading to its selection as the variable of interest [39, 40]. The median number of hot flashes per week was 15 (IQR 5–35) (Table 2, Fig. 1).

Model accuracy and variable importance

The model accuracy for predicting severity of VMS based on the AUROC value was 0.731 (SD ± 0.074) and the accuracy based on the AUPRC value was 0.687 (SD ± 0.079). The full confusion matrix and additional generalization performance metrics are included in Appendix C (Fig. C.3). The most important variable impacting the problem scale was the number of hot flashes per week (gain in AUROC 0.072 ± 0.019) (Fig. 2) followed by “age” (0.037 ± 0.018) and prescription and/or use of drugs to mitigate VMS (0.023 ± 0.007). Other highly ranked variables included whether the patient was asked about VMS in routine follow-up (0.021 ± 0.007) and whether changes were made to breast cancer treatment due to VMS (0.020 ± 0.009). The specific symptom variables had weaker impacts on the model, with waking up at night/difficulty sleeping (0.017 ± 0.005) having the greatest importance. Variables of lower importance included whether a patient had received or is still receiving endocrine therapy (0.0009 ± 0.0009), or ovarian function suppression treatments (0.0002 ± 0.001) for their breast

Fig. 2 Variable importance using the permutation and retrain method with change in accuracy ± one standard deviation. The variables are ranked from the most important to the least important. “HF” is hot flashes, “VMS” is vasomotor symptoms, “BC” is breast cancer, “CAM” is complementary and alternative medicine, and “OFS” is ovarian function suppression



cancer, which were the sixteenth and the fifteenth most important variables, respectively. The use of complementary medicines had lower relative importance in the model for the management of VMS (0.004 ± 0.004). The model with the six most important variables had similar generalization performance as the full model (see Appendix C, Table C.1). This suggests that a simpler model with fewer variables would provide comparable performance to the more complex model with 17 variables. We also removed the two variables with unclear causality (drug treatment for VMS and alteration of BC treatments) which gave poorer prediction performance than either of the other two models (Appendix C).

Functional form of relationships

Conditional partial dependence plots were generated to show the functional form of the relationships between the most important variables and hot flash severity as per the HFNS Problem Rating Score. Each graph represents the impact of the variable factoring out the effects of the other variables. Given that the cut-off point for the outcome variable represented 50% of the observations, a predicted probability of a HFNS problem score of greater than 0.5 implies the presence of severe VMS.

For the most important variable, which was the number of hot flashes per week, women who experience ≥ 17 hot flashes per week are more likely to consider their hot flashes as problematic (Fig. 3a). The peak probability score occurs at 103 hot flashes per week. Patients aged between 49 and 63 were more likely to report problematic hot flashes, with the maximum peak in women aged 56. Patients older than 63 were less likely to report problematic hot flashes (Fig. 3b). Among the 9 participants younger than or equal to age 35, 7 patients reported severe VMS (data not shown).

The probability of severe VMS was higher if a patient was offered a prescription or over-the-counter drug to manage

their VMS (Fig. B.1a), and patients who reported that they were routinely asked about VMS were less likely to rate severe symptoms (Fig. B.1b). Finally, the probability of severe VMS was higher in patients who had a change in their breast cancer therapy due to VMS (Fig. B.2). The impact of the top two symptom variables that ranked of highest importance in the model was as follows: sleep disturbances and “redness of the face”/flushing (Fig. B.5). For both symptoms, individuals who reported experiencing sleep problems or “redness of the face/flushing” were estimated to have a higher probability of having problematic hot flashes.

Discussion

Vasomotor symptoms are common in EBC patients [7, 8]; however, effectively managing this problem remains a challenge. The reasons for the complexity of this problem are many: unlike the general population, for example, estrogen replacement is contraindicated in breast cancer patients due to its role in tumorigenesis [41]. Moreover, a multitude of other challenges exist in managing VMS in this population, including systemic cancer treatments that worsen or induce VMS [7, 8], a lack of randomized controlled trials directly comparing active interventions for VMS [15], variation in patient preferences [19, 22], and health care provider uncertainty in the selection of optimal interventions for VMS [20].

While linear regression models serve as the backbone for traditional statistical modelling, they are limited in their capacity to capture complex interactions between variables, and as such, non-linear methods are often employed post hoc [42]. ML models can learn such complex relationships among multiple variables, making these techniques useful modalities for creating prediction models. Identifying patients at greatest risk of problematic and distressing VMS is critical to effective survivorship care in this patient population. In the current study using ML models, we have

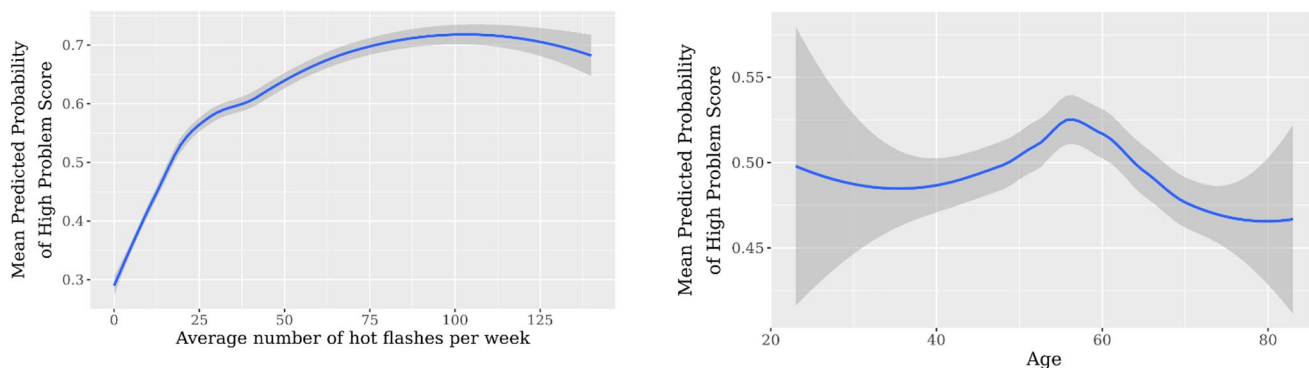


Fig. 3 The probability of reporting severe vasomotor symptoms based on **a** the number of hot flashes per week and **b** patient age. The shaded region represents one standard deviation across the fivefold cross-validation

identified important factors that are predictive of patients at risk of severe VMS.

This is the first time, to our knowledge, that this approach has been applied to predicting patients at risk of severe VMS. Moreover, our results demonstrate that the frequency of hot flashes in a given week was the single greatest predictor of severe VMS as per the HFNS problem score. Our model has helped to identify that patients experiencing 17 or more hot flashes per week (or more than 2 hot flashes per day) are more likely to experience severe symptoms. This is an important finding, as it could be used in clinical practice to identify patients that can be offered early VMS interventions to help mitigate their symptoms. The results also demonstrate that women near the age of “natural menopause” (median age of 51 in Canada) [43] are more likely to report severe symptoms, which is consistent with previous studies in the non-breast cancer population demonstrating that severe VMS are more frequent in women transitioning to menopause or in early post-menopause, than in late menopause [44]. The importance of regular assessment of VMS in routine follow-up is emphasized by the finding that women who report being regularly asked about their symptoms are less likely to rate them as severe. This is an important finding, as only 58% of patients from our original patient survey reported being regularly asked about these symptoms in routine follow-up visits [22], which is substantiated elsewhere in the literature [19]. Similarly, while specific symptoms ranked lower in the model than the variables mentioned above, screening for sleep disturbances, which was the 6th most important variable, is an additional simple method to screen for patients at risk of distressing symptoms. Finally, while the recommendation for drugs to mitigate VMS, or changes to breast cancer treatments in response to poorly controlled VMS were associated with increased severity of symptoms, these variables, while interesting and important predictors (see Appendix C), do not assist in early detection of patients experiencing severe symptoms. This is an example of how a cross-sectional survey cannot disentangle cause and effect. Interestingly, exposure to endocrine therapy and ovarian function suppression rated as low importance in predicting distress from VMS, and will be evaluated further in a future prospective trial.

There are several limitations of this study. A study oversight involved the inclusion of an 11-point scale starting at “zero” included in the online version of the survey, and a 10-point scale starting at “one” in the paper version of the survey. While this oversight may introduce challenges comparing mean cumulative HFNS problem scores *between* patients, the dichotomization of high/low severity scores about the median (i.e., a score of 4) ensures that very low score individuals would be grouped together in the “low” problem score group, regardless of whether they utilized a scale beginning at “zero” or “one.” Moreover, the majority

of patients conducted the survey electronically (259 electronic vs 114 paper), and thus utilized the 11-point scale. Secondly, while the survey from which the dataset was derived was conducted at a single-time point, the cross-sectional nature of the study does not permit evaluation of change in symptoms over time. As indicated above, the presence or absence of a drug treatment for VMS ranked highly in the variables of importance. This data indicates that patients with severe symptoms likely required escalated therapy; however, the ability of ML to predict efficacy of these medications remains unknown and requires a future prospective study to further assess this relationship. Recall bias also likely influences the results of this study, as patients were required to provide averages for the number of VMS per week, which can be subject to inaccuracies. Our survey did not ask patients about the severity of hot flashes prior to the breast cancer diagnosis, nor the age at diagnosis of menopause. These variables are likely to influence the problematic nature of VMS, and will be examined more closely in subsequent studies. The survey moreover did not integrate other risk factors for menopause, such as smoking, education, and socioeconomic status. Finally, the number of patients younger than 35, and those who utilized complementary therapies are small, and as such, their importance in the model will require further study.

Future prospective studies are needed to identify and treat patients experiencing bothersome VMS. We plan to use the above techniques in future trials to create prediction models that will guide patients and clinicians in the selection of individualized treatments for VMS, as well as the assessment of toxicities from other therapies.

Conclusion

Machine learning offers a novel way to assess treatment toxicity in early breast cancer patients. In the current study, we demonstrated that the number of hot flashes per week and age were the two most important predictors for bothersome VMS in breast cancer patients. Patients experiencing more than 17 hot flashes per week are more likely to experience bothersome symptoms, and represent a higher risk group that may benefit from therapeutic interventions. Women between the ages of 49 and 63 are also more likely to report bothersome symptoms. Future studies specifically looking at these high-risk groups are needed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00520-022-07156-6>.

Author contribution KC, SMG, MC, MA, LV, FM, AP, and GL designed the survey and prepared the protocol. ML collected the data and coordinated the study. KC, MC, SMG, and KEE were involved in data preparation. KEE and LM created the machine learning algorithm

and analyzed the data. KC, MC, SMG, and KEE wrote the manuscript. All authors had full access to the data and take responsibility for the integrity of the data and accuracy of the data analysis. All authors were involved in the critical review of the manuscript and approved the final version.

Funding This work was supported by the Rethinking Clinical Trials (REaCT) program platform at the Ottawa Hospital, which is supported by The Ottawa Hospital Foundation and its donors.

Data availability The dataset generated during and/or analyzed during the current study are available from the corresponding author on request with approval of the Ontario Cancer Research Ethics Board.

Declarations

Statement of human rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Ontario Cancer Research Ethics Board (OCREB).

Statement on the welfare of animals This article does not contain any studies with animals performed by any of the authors.

Informed consent Participation and completion of the survey implied consent to participate in the study.

Consent for publication Participation and completion of the survey implied consent to publication of aggregate survey findings.

Conflict of interest SMG reports receipt of honorarium from Novartis for insights on management of breast cancer patients. BH and MC report consulting fees from Cornerstone Research, outside the submitted works. All other authors declare no competing interests. KEE is a co-founder of and SVP at Replica Analytics, a spin-off company from the University of Ottawa that develops synthetic data generation software. LM is an employee of Replica Analytics. GRP reports consulting fees from Merck and Profound Medical, honorarium for DSMB membership from Takeda and has a close family member who works for Roche Canada and owns stock in Roche, all outside the submitted works.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

- Rashidi HH, Tran NK, Betts EV, Howell LP, Green R (2019) Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 6:2374289519873088
- Lynch CM, Abdollahi B, Fuqua JD, de Carlo AR, Bartholomai JA, Balgemann RN et al (2017) Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform* 108:1–8
- Deist TM, Dankers FJWM, Valdes G, Wijsman R, Hsu IC, Oberije C et al (2018) Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys* 45:3449–3459
- Lu W, Fu D, Kong X, Huang Z, Hwang M, Zhu Y et al (2020) FOLFOX treatment response prediction in metastatic or recurrent colorectal cancer patients via machine learning algorithms. *Cancer Med* 9:1419–1429
- Nindrea RD, Aryandono T, Lazuardi L, Dwiprahasto I (2018) Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. *Asian Pac J Cancer Prev* 19:1747–1752
- Montazeri M, Beigzadeh A (2016) Machine learning models in breast cancer survival prediction. *Technol Health Care* 24:31–42
- Howell A, Cuzick J, Baum M, Buzdar A, Dowsett M, Forbes JF et al (2005) Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet* 365:60–62
- Pagani O, Regan MM, Walley BA, Fleming GF, Colleoni M, Lang I et al (2014) Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. *N Engl J Med* 371:107–118
- Chirgwin JH, Giobbie-Hurder A, Coates AS, Price KN, Ejlertsen B, Debled M et al (2016) Treatment adherence and its impact on disease-free survival in the Breast International Group 1–98 trial of tamoxifen and letrozole, alone and in sequence. *J Clin Oncol* 34:2452–2459
- Hershman DL, Kushi LH, Shao T, Buono D, Kershenbaum A, Tsai WY et al (2010) Early discontinuation and nonadherence to adjuvant hormonal therapy in a cohort of 8,769 early-stage breast cancer patients. *J Clin Oncol* 28:4120–4128
- McCowan C, Shearer J, Donnan PT, Dewar JA, Crilly M, Thompson AM et al (2008) Cohort study examining tamoxifen adherence and its relationship to mortality in women with breast cancer. *Br J Cancer* 99:1763–1768
- Yood MU, Owusu C, Buist DS, Geiger AM, Field TS, Thwin SS et al (2008) Mortality impact of less-than-standard therapy in older breast cancer patients. *J Am Coll Surg* 206:66–75
- Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V et al (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* 381:805–816
- Goss PE (2007) Letrozole in the extended adjuvant setting: MA.17. *Breast Cancer Res Treat*. 105(Suppl 1):45–53
- Hutton B, Hersi M, Cheng W, Pratt M, Barbeau P, Mazzarello S et al (2020) Comparing interventions for management of hot flashes in patients with breast and prostate cancer: a systematic review with meta-analyses. *Oncol Nurs Forum* 47:E86–E106
- Runowicz CD, Leach CR, Henry NL, Henry KS, Mackey HT, Cowens-Alvarado RL et al (2016) American Cancer Society/American Society of Clinical Oncology breast cancer survivorship care guideline. *CA Cancer J Clin* 66:43–73
- Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT et al (2019) Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 30:1194–1220
- Peate M, Saunders C, Cohen P, Hickey M (2021) Who is managing menopausal symptoms, sexual problems, mood and sleep disturbance after breast cancer and is it working? Findings from a

- large community-based survey of breast cancer survivors. *Breast Cancer Res Treat* 187:427–435
19. Fenlon D, Morgan A, Khambaita P, Mistry P, Dunn J, Ah-See ML et al (2017) Management of hot flushes in UK breast cancer patients: clinician and patient perspectives. *J Psychosom Obstet Gynaecol* 38:276–283
 20. Cole KM, Clemons M, Alzahrani M, Larocque G, MacDonald F, Vandermeer L et al (2021) Developing patient-centred strategies to optimize the management of vasomotor symptoms in breast cancer patients: a survey of health care providers. *Breast Cancer Res Treat* 188:343–350
 21. Chang HY, Jotwani AC, Lai YH, Jensen MP, Syrjala KL, Fann JR et al (2016) Hot flashes in breast cancer survivors: frequency, severity and impact. *Breast* 27:116–121
 22. Cole K, Clemons M, El Emam K, Larocque G, MacDonald F, Vandermeer L et al (2022) Vasomotor symptoms in early breast cancer—a “real world” exploration of the patient experience. *Support Care Cancer* 30(5):4437–4446. <https://doi.org/10.1007/s00520-022-06848-3>
 23. Hunter MS, Liao KL (1995) A psychological analysis of menopausal hot flushes. *Br J Clin Psychol* 34:589–599
 24. Mann E, Smith MJ, Hellier J, Balabanovic JA, Hamed H, Grunfeld EA et al (2012) Cognitive behavioural treatment for women who have menopausal symptoms after breast cancer treatment (MENOS 1): a randomised controlled trial. *Lancet Oncol* 13:309–318
 25. Nunnally J, Bernstein IH (1994) *Psychometric theory*. McGraw-Hill, New York
 26. Ayers B, Hunter MS (2013) Health-related quality of life of women with menopausal hot flushes and night sweats. *Climacteric* 16:235–239
 27. Bühlmann Peter, Torsten H (2007) Boosting algorithms: regularization, prediction and model fitting. *Stat Sci* 22:477–505
 28. Ke G, Meng Q, Finley T, Wang T, Chen W, Weidong M, et al. (2017) LightGBM: a highly efficient gradient boosting decision tree [abstract]. *NIPS’ 17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 3149–57.
 29. Snoek J, Larochelle H, Adams R (2012) Practical Bayesian optimization of machine learning algorithms [abstract]. *Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 2*. Curran Associates Inc., Red Hook, NY, USA, 2951–59
 30. Pepe MS (2004) *The statistical evaluation of medical tests for classification and prediction*, 1st edn. Oxford University Press, Oxford
 31. Davis J, Goadrich M (2006) The relationship between precision-recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. 148:233–40
 32. Boyd K, Costa VS, Davis J, Page CD (2012) Unachievable region in precision-recall space and its effect on empirical evaluation. *Proceedings of the 29th International Conference on International Conference on Machine Learning*. Edinburgh, Scotland: Omnipress. p. 1619–26.
 33. Janitzka S, Strobl C, Boulesteix AL (2013) An AUC-based permutation variable importance measure for random forests. *BMC Bioinf* 14:119
 34. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinf* 8:25
 35. Nicodemus KK, Malley JD, Strobl C, Ziegler A (2010) The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinf* 11:110
 36. Mentch L, Hooker G (2016) Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J Mach Learn Res* 17:841–881
 37. Hooker G, Mentch L (2019) Please stop permutating features: an explanation and alternatives. *ArXiv*, abs/1905.03151
 38. Molnar C, Konig G, Bischl B, Casalicchio G (2020) Model-agnostic feature importance and effects with dependent features- a conditional subgroup approach. <http://arxiv.org/abs/2006.04628>
 39. Biglia N, Bounous VE, Susini T, Pecchio S, Sgro LG, Tuninetti V, et al. (2018) Duloxetine and escitalopram for hot flushes: efficacy and compliance in breast cancer survivors. *Eur J Cancer Care (Engl)* 27
 40. Loprinzi CL, Kugler JW, Barton DL, Dueck AC, Tschetter LK, Nelimark RA et al (2007) Phase III trial of gabapentin alone or in conjunction with an antidepressant in the management of hot flashes in women who have inadequate control with an antidepressant alone: NCCTG N03C5. *J Clin Oncol* 25:308–312
 41. Clemons M, Goss P (2001) Estrogen and the risk of breast cancer. *N Engl J Med* 344:276–285
 42. Bertsimas D, Wiberg H (2020) Machine learning in oncology: methods, applications, and challenges. *JCO Clin Cancer Inform* 4:885–894
 43. Costanian C, McCague H, Tamim H (2018) Age at natural menopause and its associated factors in Canada: cross-sectional analyses from the Canadian Longitudinal Study on Aging. *Menopause* 25:265–272
 44. Mitchell ES, Woods NF (2015) Hot flush severity during the menopausal transition and early postmenopause: beyond hormones. *Climacteric* 18:536–544

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Katherine Marie Cole¹ · Mark Clemons^{1,2} · Sharon McGee¹  · Mashari Alzahrani¹ · Gail Larocque³ · Fiona MacDonald³ · Michelle Liu² · Gregory R. Pond⁴ · Lucy Mosquera⁵ · Lisa Vandermeer² · Brian Hutton⁶ · Ardelle Piper⁷ · Ricardo Fernandes⁸ · Khaled El Emam^{5,9}

¹ Department of Medicine, Division of Medical Oncology, The University of Ottawa, Ottawa, Canada

² Cancer Therapeutics Program, The Ottawa Hospital Research Institute, Ottawa, ON, Canada

³ The Ottawa Hospital Cancer Centre, Ottawa, Canada

⁴ Department of Oncology, McMaster University, Hamilton, ON, Canada

⁵ CHEO Research Institute, University of Ottawa, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada

⁶ Clinical Epidemiology Program, The Ottawa Hospital Research Institute, Ottawa, ON, Canada

⁷ University of Ottawa Health Services, Ottawa, ON, Canada

⁸ Division of Medical Oncology, Department of Oncology, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

⁹ School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada