International Journal of Medical Informatics 141 (2020) 104170

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Identifying predictive factors for neuropathic pain after breast cancer surgery using machine learning

Lamin Juwara[a,d], Navpreet Arora[b], Mervyn Gornitsky[a,b], Paramita Saha-Chaudhuri[c,d,*], Ana M. Velly[a,b]

[a] Department of Dentistry, Jewish General Hospital Montreal, Quebec, Canada
[b] Faculty of Dentistry, McGill University Montreal, Quebec, Canada
[c] Department Epidemiology, Biostatistics & Occupational Health, McGill University (Montreal, Purvis Hall, 1020 Pine Avenue West, Montreal, H3A 1A2 Quebec, Canada
[d] Quantitative Life Sciences, McGill University Montreal, Quebec, Canada

ARTICLE INFO

Keywords:
Prediction
Neuropathic pain
DN4-interview
Machine learning

ABSTRACT

Introduction: Neuropathic pain (NP) remains a major debilitating condition affecting more than 26% of breast cancer survivors worldwide. NP is diagnosed using a validated 10-items Douleur Neuropathique - 4 screening questionnaire which is administered 3 months after surgery and requires patient-doctor interaction. To develop an effective prognosis model admissible soon after surgery, without the need for patient-doctor interaction, we sought to [1] identify specific pain characteristics that can help determine which patients may be susceptible to NP after BC surgery, and 2) assess the utility of machine learning models developed in objective [1] as a knowledge discovery tool for downstream analysis.

Methods: The dataset is from a prospective cohort study of female patients scheduled to undergo breast cancer surgery for the first time at the Jewish General Hospital, Montreal, Canada between November 2014 and March 2019. NP was assessed at 3 months after surgery using Douleur Neuropathique – 4 interview scores (in short, DN4-interview; range: 0–7). For the primary analysis, we constructed six ML algorithms (least square, ridge, elastic net, random forest, gradient boosting, and neural net) to identify the most relevant predictors for DN4-interview score; and compared model performance based on root mean square error (RMSE). For the secondary analysis, we built a logistic classification model for neuropathic pain (DN4-interview score $\geq$ 3 versus DN4-interview score $<$ 3) using the relevant-consensus-predictors from the primary analysis.

Results: Anxiety, type of surgery, preoperative baseline pain and acute pain on movement were identified as the most relevant predictors for DN4 - interview score. The least square regression model (RMSE = 1.43) is comparable in performance with random forest (RMSE = 1.39) and neural network model (RMSE = 1.50). The Gradient boosting model (RMSE = 1.16) outperformed the models compared including the penalized regression models (ridge regressions, RMSE = 1.28; and elastic net, RMSE = 1.31). In the secondary analysis, the preferred logistic regression classier for NP had an area under the curve (AUC) of 0.68 (95% CI = 0.57 to 0.79). Anxiety was significantly associated with the likelihood of NP (odds ratio = 2.18; 95% CI = 1.05–4.49). In comparison to their counterparts, the odds of NP were higher in participants with acute pain on movement or with present preoperative baseline pain or participants who performed total mastectomy surgery, but the differences were not statistically significant.

Conclusions: Modern machine learning models show improvements over traditional least square regression in predicting of DN4-interview score. Penalized regression methods and the Gradient boosting model out-perform other models. As a predictor discovery tool, machine learning algorithms identify relevant predictors for DN4-interview score that remain statistically significant indicators of neuropathic pain in the classification model. Anxiety, type of surgery and acute pain on movement remain the most useful predictors for neuropathic pain.

* Corresponding author at: Department of Epidemiology, Biostatistics & Occupational Health McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, H3A 1A2 Quebec, Canada.
E-mail address: paramita.sahachaudhuri.work@gmail.com (P. Saha-Chaudhuri).

https://doi.org/10.1016/j.ijmedinf.2020.104170
Received 18 October 2019; Received in revised form 8 April 2020; Accepted 3 May 2020

1. Introduction

Neuropathic pain (NP) is a severe and debilitating disorder of the peripheral and central nervous systems. It is defined as "pain arising from a direct consequence of a lesion or disease affecting the somatosensory system [1,2]. The affected nerve lesions trigger molecular changes in nociceptive neurons that become hypersensitive, resulting in spontaneous pain, shooting pain sensations and sympathetically maintained pain [3–5]. In cancer patients, for example, NP could be caused by the invasiveness of the tumor, radiotherapy and/or side effect of chemotherapy or surgery [6].

NP is highly prevalent and estimated to affect approximately 7–10% of the general population worldwide [7,8]. The condition is most common among women above the age of 45 and implicates both physical and psychological co-morbidities. Despite increased interest in neuropathic pain, treatment remains highly unsatisfactory [9]. Instead of comprehensive pain relief, current treatment efforts are symptom-specific, geared towards pain relief from specific conditions and intended for pain management [10].

Recent advances in medicine have resulted in exceptional improvements in Breast Cancer (BC) survival rates. However, NP remains a major debilitating symptom in BC patients and up to 26% of cancer survivors experience the condition [11]. Patients are typically under strong medication and increased hospital visits after BC surgery. This can result in impaired quality of sleep, loss of function, anxiety, depression, impaired cognition and/or overall life [7]. Thus, there is a significant interest in understanding the epidemiology of NP in breast cancer survivors.

Several patient-reported screening tools are available for diagnosing NP in health care settings including Leeds Assessment of Neuropathic Symptoms and Signs (LANSS) [12], the Neuropathic Pain Questionnaire (NPQ) and the Douleur Neuropathique - 4 (short form, DN4) Screening Questionnaire among others [13]. The 10-item standard DN4 screening questionnaire, which was introduced in French, is particularly popular and has been validated in many different languages [14–17]; it evaluates pain following central and/or peripheral lesions, and allows practitioners to determine if a pain condition is neuropathic. However, the utility of the standard 10-item DN4 questionnaire poses two major challenges for early detection of potential NP development: 1) A lapse period of > 3 months after BC surgery is required to administer the DN4 questionnaire, and 2) the questionnaire requires doctor-patient interactions to complete questions related to patient examination. Hence, the development of an effective and easy-to-use prognosis tool, that doesn't require patient-doctor interaction, and can be administered promptly post-surgery would greatly improve both the management and treatment of NP.

Clinical prediction models play a fundamental role in aiding decision-making and are constantly used in various clinical settings and practices ranging from risk stratification in intensive care units to medical image diagnostics [18]. The etiological complexity of NP after a major surgery like BC surgery presents an interesting challenge in clinical care [19–21]. The lack of a universally accepted and validated clinical diagnostic tool for NP makes it difficult for clinicians and policymakers to provide precise estimates of NP prevalence. Traditionally, regression models (e.g. Least square regression) have been used to identify significant predictors of binary or continuous clinical outcomes in pain studies [22,23]; however, such models do not perform well in capturing complex relationships including interactions and are not usually robust to multicollinearity [24,25], which is a fundamental feature of pain-related variables.

The objectives of the current study include 1) To examine the performance of different machine learning algorithms as predictive models for DN4-interview score (referred to DN4 score henceforth) after breast cancer surgery; 2) To use the models as a knowledge discovery and variable exploration tool in order to identify the relevant predictors for DN4 score; 3) To include the relevant predictors as covariates in the classification model for NP while leveraging residual data variability from a select component of multiple correspondence analysis of the remaining predictors.

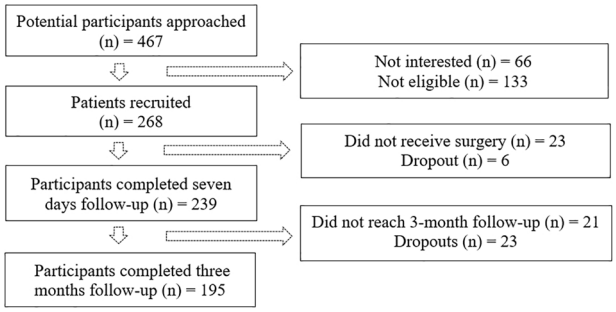

Fig. 1. Patient recruitment and follow-up.

2. Method

2.1. Data source

The data used in this study is obtained from a 3-months prospective cohort study of breast cancer patients seen at the Segal Cancer Centre, Jewish General Hospital (JGH), Montreal, Quebec, Canada from November 2014 to March 2019. The study was approved by the Research Ethics Committee of the JGH. All the potential participants signed a written consent form prior to their inclusion.

Eligible participants included in the study were adult female patients diagnosed with breast cancer and scheduled to undergo breast cancer surgery for the first time. The study excluded breast cancer patients not undergoing surgery, patients diagnosed with other cancer types or under considerable medical assistance (determined by Karnofsky Performance Scale Index < 50), and patients diagnosed with metastases or were pregnant or had no access to a telephone (Fig. 1).

Prior to the patient's scheduled surgery, a researcher at the JGH administered a set of questionnaires and validated tools to assess the relevant risk factors. More specifically, the modified-Brief Pain Inventory (m-BPI), Generalized anxiety disorder-7 (GAD-7) and Physical Health Questionnaire-8 (PHQ-8) were administered. Baseline socio-demographic information was also collected in the same session. The follow-up sessions at 7 days and 3 months after surgery were conducted through telephone interviews. In Fig. 2, we present a timeline of information collected during these sessions. All the patients in the current study completed the DN4 screening questionnaire at 3 months.

2.2. Outcomes

The primary outcome, Douleur Neuropathique – 4 interview score (DN4 score), was derived from seven post-operative questions assessed at 3 months after surgery. The questions were administered through a telephone interview and participants were asked to assess their status of three essential pain qualities 1) Burning? (yes/no), 2) Painful cold? (yes/no), and 3) Electric shock? (yes/no); and four pain associated symptoms 4) Tingling? (yes/no), 5) Pins and needles? (yes/no), 6) Numbness? (yes/no), and 7) Itching? (yes/no). DN4 score at 3 months was calculated from the sum of these questions. An example of the DN4-interview questionnaire is provided in Appendix A.

In the secondary analysis, the outcome variable considered is Neuropathic pain. It is calculated from total DN4 score. Patients were classified as neuropathic if their DN4 score was greater than or equal to 3 and non-neuropathic otherwise. The cut-off range of 3 was selected based on the recommendations presented in [13].

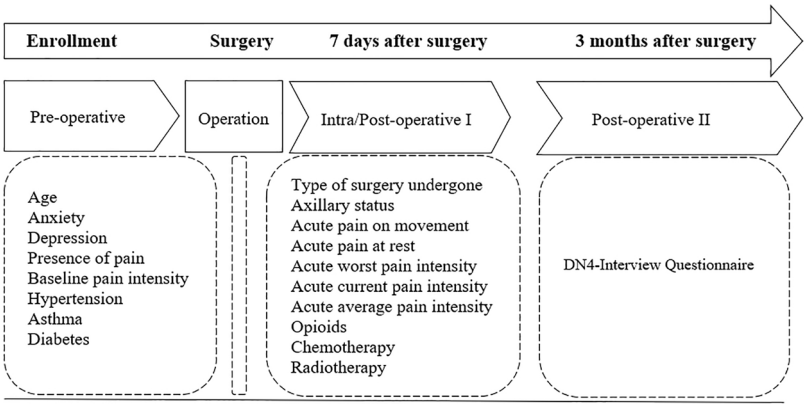

Fig. 2. Data collection timeline.

2.3. Predictors

To develop machine learning models for the DN4 score, we considered several pre-, intra- and post-operative patient characteristics as potential predictors.

The *pre-operative* predictors included age at surgery, anxiety status (no or mild versus moderate to severe), depression (no or mild versus moderate to severe), and self-reported breast pain variables (presence of any breast pain (yes/no) and rating of breast pain intensity (range, 1–10)). Predictors pertaining to the patients' medical history were also collected; these include whether the patient was diagnosed with diabetes (yes/no) or hypertension (yes/no) or asthma (yes/no).

The *Intra-operative* predictors included the type of surgery (segmental mastectomy or total mastectomy) and the patient's axillary node status (Sentinel Lymph Node Biopsy [SLNB] or Axillary Lymph Node Dissection [ALND]).

Post-operative predictors were defined as patient co-morbidities measured 7 days after surgery. These self-reported predictors include acute breast pain on movement (yes/no), acute breast pain at rest (yes/no), and three other pain intensity ratings: acute worst pain intensity (range, 1–10), acute current pain intensity (range, 1–10), and acute average pain intensity (range, 1–10). Other post-operative predictors that were considered include whether the patient was prescribed any opioids (yes/no), chemotherapy treatment (yes/no) and radiotherapy (yes/no).

2.4. Statistical analysis

The primary objective was to develop prediction models for DN4 score using a combination of pre-, intra-, and post-operative predictors. The machine learning models constructed include [1] Least Square regression model [2], Ridge regression [3], Elastic Net regression [4], Random Forest [5], Gradient Boosting and [6] Neural Nets. In the next paragraph, we briefly introduced the models considered and expound on their mathematical formulation in Appendix A. The performances of the models were compared based on root mean square error (RMSE), with lower values indicating better accuracy. For each constructed model, the most relevant predictors were identified. We restricted our comparison to four predictors – the same quantity of 'statistically significant' predictors estimated using the standard least square regression model based on a 0.01 p-value threshold. The missing items were imputed once (mean matching for quantitative variables and the median for categorical variables (26)) to prevent individuals from completely dropping from the statistical analysis. Comparisons to complete case analysis are provided in Appendix A. For small proportions of missing values in predictors (e.g. 6% missing), single imputation would not introduce bias in the analysis [27,28].

To predict DN4-interview score (denoted Y) from the measured predictors (denoted $X_1, X_2, ..., X_p$), we first constructed the multivariate regression model for the output Y of the form $\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{p} X_j\hat{\beta}_j$ where $\hat{\beta}_0$ represents the model intercept and the sequence $\{\hat{\beta}_j\}_{j=1}^{P}$ are the coefficients of the predictors. To estimate the coefficients $\{\hat{\beta}_j\}_{j=0}^{P}$ using *least square regression* algorithm, we minimize the residual function $RSS(\beta) = \sum_{i=1}^{N}(y_i - x_i^T\beta)^2$; where the quadratic structure of $RSS(\beta)$ guarantees the existence of the minimum residual value [29].

Penalization techniques were next introduced to reduce the variance of the effect estimate. When patient characteristics are strongly correlated, a model based on such features results in unstable estimates and inflated variance. For example, two negatively correlated covariates may return large effect estimates with opposite signs. We considered two penalization models; *ridge regression (RR)* and *elastic net (EN)*. The least squares model is penalized for including predictors that do not sufficiently explain the outcome such that coefficients of the least contributing predictors are either set to zero (induced sparsity) or shrunk towards zero. RR shrinks the coefficients of non-useful predictors whilst EN combines shrinkage and sparsity.

Next, we considered the *random forest (RF)* model. RF is a tree-based predictive model that successively splits the dataset (starting from the predictor that is the strongest classifier) in order to produce homogeneous subpopulations. Another tree-based model considered was the *gradient tree boosting (GBM)* model that constructs a group or ensemble of shallow and weak successive trees with each tree utilizing information from previously grown trees in order to minimize prediction error.

We considered the *artificial neural net (ANN)* model next. ANN models are inspired by the neuronal structure of the mammalian cerebral cortex. A key underlying assumption of neural nets is that simple features work in tandem to output complex results. They represent a broad class of non-linear model formed from extracting linear combinations of the features (predictors) connected via weights, and then modeling the outcome as a nonlinear function of these features. The mathematical formulation and detailed description of the machine learning models considered are presented in Appendix A.

The secondary analysis sought to construct a classification model for neuropathic pain status. We formulated a *logistic regression* (LR) model to distinguish neuropathic (DN4 score $\geq$ 3) versus non-neuropathic pain (DN4 $<$ 3) using the top relevant predictors that were consistently identified by the machine learning models constructed in the primary analysis. For linear models (LS, RR, Lasso, and EN), variable importance is estimated from the absolute value of the t-statistics. Variable importance for tree-based methods (RF and GB) are estimated by first computing tree specific MSE on the out-of-bag data and permutations of the variables; the differences are then averaged and normalized by the standard error. The neural net model uses a combination of the absolute values of the weights based on the method proposed in [30].

Our choice of the LR classifier was motivated by the minimum

sample size recommendation proposed by Van der Ploeg et al. [31]. They recommend about 20 events per covariate for LR and approximately 200 events for modern learning algorithms. The LR model proposed includes the relevant predictors and the top contributing components of multiple correspondence analysis (MCA; analogous to PCA for qualitative variables) of the remaining qualitative co-morbidities. The top 4 principal components from MCA, representing > 60% of the remaining variability, were included in order to avoid overfitting the model (MCA-adjusted model). Model performance is evaluated based on its ability to discriminate between NP and non-NP (AUC score), and a calibration plot of the predicted probability of NP against the observed probabilities.

The models underwent a 2-stage validation process. The dataset was randomly partitioned into 80% train and 20% test set. With the training set, the models were internally cross-validated using 10-fold cross validation (*i.e.* we randomly split the training sample into 10 subsets, and iteratively train the model on 9 subsets while obtaining prediction errors on the left-out subset before combining the results) to compute train RMSE. We next computed bootstrapped test RMSE estimates from resampled random splits of the data. For each resampled training set, the unknown parameters are estimated using cross validation which is then used to compute the RMSE on the pseudo-independent test data (not previously used in model construction). All analyses were performed with R statistical software (version 3.4.4). In Appendix A, we specify the package(s) used for each model.

3. Results

Distributions of the covariates are presented in Table 1. The analysis included 204 eligible female participants of mean age 56.16 (SD = 14.26) years at baseline. Approximately 33% of the participants reported pre-surgical baseline pain with a mean pain intensity score of 7.46 (SD = 15.68). At baseline, 45.6% of the participant reported experiencing moderate to severe anxiety and 38.7% reported experiencing moderate to severe depression. Other comorbidities reported were diabetes (10.82%), asthma (7.22%) and allergy (33.51%). Most participants (90.77%) had undergone mastectomy segmental or excision and were prescribed at least one type of opioid (74.1%) during their postoperative pain management.

The data for acute breast pain were collected seven days after surgery. Among the participants, 112 reported acute breast pain at rest and 100 reported acute breast pain during movement respectively. The average score of current, worst and average acute pain intensities was 2.10 (SD = 2.21), 4.55 (SD = 3.16) and 2.44 (SD = 2.12) respectively. A total of 49 (24.01%) participants received chemotherapy and 103 (50.49%) patients received radiotherapy. The mean DN4 interview score for all the participants was 1.45 (SD = 1.41).

In Table 2, we show the results of estimated train RMSE and the top four relevant predictors in each machine learning method. Type of surgery, baseline pain, anxiety and acute pain on movement were consistently identified among the top-4 relevant predictors in most of the ML models assessed. Other variables that were sparingly captured include acute pain intensity and type of surgery. Overall, the penalization models (RR and EN) and GBM outperformed the remaining ML models based on estimated RMSE. In addition to the train RMSE used for model comparison in Table 2, we present (in Fig. 3) bootstrap estimates of the test RMSE from repeated resamples of the training set (Fig. 3). Distribution of top-consensus predictors stratified by NP is presented in Table 3.

The results (odds ratios (ORs) and 95% confidence intervals) of the classification model constructed in the secondary analysis are in Table 4. Effect estimates are presented for the MCA-unadjusted and MCA-unadjusted model. Including the top components from MCA improved model performance while preventing model overfitting (based on the maximum recommendation covariate inclusion criteria for logistic regression proposed in (24)). For each of the predictors, we report the OR estimates and 95%CI. In Fig. 4(A), we present the AUC (95% CI) of the MCA-unadjusted (AUC = 0.64, 95% CI = 0.50,0.79) and MCA-adjusted model (AUC = 0.68, 95% CI = 0.57,0.79). A calibration plot of the relationship between the predicted and observed probabilities for NP is presented in Fig. 4(B–C).

In the adjusted logistic model, anxiety is estimated as the most relevant predictors for NP; the odds of NP is significantly higher in patients with anxiety than patients without anxiety (OR = 2.18; 95% CI = 1.05–4.49). The odds of NP is seemingly higher for patients with acute pain on movement (OR = 2.10; 95% CI = 0.72–4.65) and patients who received total mastectomy compare to segmental mastectomy (OR = 2.10; 95% CI = 0.72–4.65). The ORs for baseline pain and the components from MCA included in the logistic model were not statistically significant. Compared to the MCA-unadjusted model (65.0% sensitivity, 54.3% specificity), the MCA adjusted model had 75.0% sensitivity and 61.5% specificity based on the optimal cutoff of 0.37 (maximizes Youden's index). Finally, the MCA adjusted LR model has positive predictive value (PPV) = 0.67 and negative predictive value (NPV) = 0.81 whilst the unadjusted model has PPV = 0.68 and NPV = 0.50 at the same cutoff.

Table 1
Patient characteristics and DN4-interview score distributions of covariates. Mean (and standard deviations, SD) are provided for continuous variables and frequencies for categorical variables. *P* Value compares mean DN4-interview score between qualitative variables categories.

	Frequency/Mean (SD)	Missing	DN4-Interview score	
			Mean (SD)	*P value*
Age	56.16 (14.26)		1.45 (1.41) †	
Baseline pain	134 / 61		1.28 (1.36) /	0.025
no / yes			1.77 (1.48)	
Baseline pain intensity	7.46 (15.68)		–	
Acute pain on movement	94 / 100	1	1.14 (1.30) /	0.004
no / yes			1.73 (1.48)	
Acute pain at rest	81 / 113	1	1.29 (1.46) /	0.186
no / yes			1.56 (1.39)	
Acute worst pain intensity	4.55 (3.16)		–	
Acute current pain intensity	2.10 (2.21)		–	
Acute average pain intensity	2.44 (2.12)		–	
Depression	119 / 76		1.39 (1.44) /	0.554
no or mild / moderate to severe			1.51 (1.38)	
Anxiety	106 / 89		1.28 (1.38) /	0.096
no or mild / moderate to severe			1.61 (1.44)	
Chemotherapy	141 / 47	7	1.43 (1.39) /	0.795
no / yes			1.49 (1.50)	
Radiotherapy	89 / 99	7	1.59 (1.45) /	0.192
no / yes			1.31 (1.37)	
Surgery type	177 / 18		1.38 (1.41) /	0.049
ms / m			2.06 (1.43)	
Axillary lymph node	55 / 140		1.23 (1.21) /	0.138
no / yes			1.55 (1.50)	
Any opioid taken	30 / 153	12	1.17 (1.29) /	0.283
no / yes			1.48 (1.47)	
Diabetes	174 / 21		1.42 (1.40) /	0.656
no / yes			1.57 (1.66)	
Hypertension	147 / 48		1.46 (1.43) /	0.615
no / yes			1.36 (1.42)	
Asthma	181 / 14		1.44 (1.43) /	0.857
no / yes			1.36 (1.45)	

Abbreviation: NP = Neuropathic pain score; Surgery type (ms = mastectomy segmental; m = total mastectomy); Axillary lymph node (slnb = sentinel lymph node biopsy, alnd = axillary lymph node dissection); †, represent all.
the participants. Tumor grade (83/195 missing) was excluded.

Table 2
Performance measure (cross-validated train RMSE) and the four most significant predictors of neuropathic DN4 score from six machine learning models.

	OLS	RR	EN	RF	GBM	NN
Train RMSE	1.43	1.28	1.31	1.39	1.16	1.50
Predictor						
Age	*			*	*	
Baseline pain	*	*	*		*	*
Baseline pain intensity				*		
Acute pain on movement	*	*	*	*	*	
Acute pain at rest						
Acute worst pain intensity					*	*
Acute current pain intensity				*		
Acute pain on average intensity					*	
Depression						
Anxiety	*	*	*	*		*
Chemotherapy						
Radiotherapy		*	*			*
Surgery type	*	*	*			*
Axillary lymph node						
Any opioid taken						
Diabetes						
Hypertension						
Asthma						

Abbreviations: OLS Least square regression, RR Ridge regression, EN Elastic net, RF Random forest, GBM Gradient boosting machines, NN Neural network, RMSE Root mean square error.

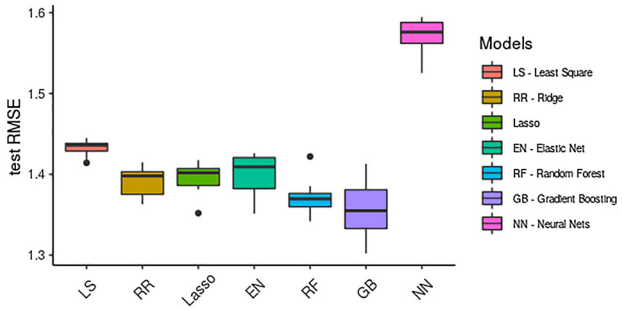

Fig. 3. Bootstrapped estimate of test RMSE for each machine learning model. Each resampled test RMSE is generated from estimating the model parameters (through cross validation) on a randomly selected training set (80% data) and computing the RMSE on the test set. The median test RMSE estimates were LS = 1.4362, RR = 1.3980, Lasso = 1.4019, EN = 1.4093, RF = 1.3697, GB = 1.3550, and NN = 1.5760.

4. Discussion

The study assessed the performance of multiple machine learning models to identify predictive factors for neuropathic pain after breast cancer surgery (*i.e.* six models for DN4-interview score and two models for neuropathic pain classification). The models predict neuropathic pain score using self-assessed patient questionnaires and relevant covariates collected at most seven days after surgery without the traditional need for doctor-patient interaction at 3 months after surgery.

In predicting DN4-interview scores, gradient boosting machines and penalization methods out-perform traditional statistical methods like least squares regression based on their estimated RMSE. We found the neural net model to be the worst-performing model for predicting neuropathic pain. Even though neural network models typically perform well in modeling complex datasets, their performance in difficult-to-recruit study settings like ours is potentially limited by the available sample size. Our analysis consistently identified anxiety, type of surgery, pre-operative baseline pain and acute pain on movement as the major relevant predictors of DN4-interview score in the various models. However, anxiety and the type of surgery are most indicative of potential neuropathic pain. This supports findings from previous studies [11] that have shown psychological factors to be significantly associated with neuropathic pain of similar pathophysiology. After including additional data variability from the top-4 components from MCA, we continued to see a significant association between anxiety and neuropathic pain. A borderline statistically significant association is also observed between acute pain on movement and neuropathic pain. The MCA adjusted model had a good predictive score (AUC) and calibration (intercept and slope).

Of note, we make a distinction between traditional statistical modeling (e.g., linear and logistic regression for continuous and binary outcomes respectively) and recently popularized statistical learning methods (e.g. tree-based methods and neural nets). While traditional statistical modeling techniques such as linear and logistic regression fall under the "machine learning" (ML) umbrella, recent use of ML is akin to the algorithm-based modeling techniques [32,33]. Therefore, to distinguish between the modeling techniques built from a statistical perspective and the algorithm-driven techniques built from a computer science and informatics perspective, we chose to use statistical modeling for the former and ML for the latter.

Extensive and thorough clinical examination remains a gold

Table 4
Logistic regression (LR) models for Neuropathic Pain (n = 45 of 195 patients). Model LR-1 includes only the relevant predictors. Model LR-2 is adjusted with the top 4 individual components from MCA. The c-statistics of the logistic classifier indicates model discrimination index.

Model Predictor	LR-1 (MCA-unadjusted)		LR-2 (MCA-adjusted)	
	OR	95% CI	OR	95% CI
Baseline pain (no/yes)	1.17	(0.58, 2.38)	1.27	(0.62, 2.64)
Anxiety (no or mild /moderate to severe) /yes)	1.85	(0.95, 3.63)	2.18	(1.05, 4.49)
Acute pain on movement (no/yes)	1.57	(0.78, 3.18)	1.68	(0.72, 3.88)
Surgery (ms/m)	1.88	(0.69, 5.10)	1.78	(0.64 4.93)
MCA1	N/A		1.17	(0.48, 2.83)
MCA2	N/A		0.96	(0.35, 2.66)
MCA3	N/A		1.20	(0.47, 3.04)
MCA4	N/A		0.67	(0.26, 1.79)
C-index	0.64		0.68	

Abbreviations: ORodds ratio; 95% CI95% confidence interval; MCA(j)jth; component of Multiple Correspondence Analysis; LRlogistic regression; OLSordinary least square regression; Reference category is underlined.

Table 3
Distribution of predictive variables of patients stratified by Neuropathic Pain status (DN4-interview score ≥3 versus DN4-interview score < 3). Histogram of DN4-interview score is provided in Appendix A.

Predictor	Neuropathic pain (n = 45).	Non-neuropathic pain(n = 150).
Baseline pain (no/yes)	29(64.4%) / 16(35.6%)	105 (70.0%) / 45 (30.0%)
Anxiety (no or mild /moderate to severe)	18 (40.0%) / 27 (60.0%)	88 (58.7%) / 62 (41.3%)
Acute pain on movement (no/yes)	17 (37.8%) / 28 (62.2%)	77 (51.7) / 72 (48.3%)
Surgery type (ms/m)	38 (84.4%) / 7 (15.6%)	139 (92.7%) / 11 (7.3%)
Age [D]	56.33 (SD = 13.13)	57.70 (SD = 14.42)

Distributions are presented as n (%) and mean (SD = standard deviation). [D] demographic information is not a consensus predictive factor.

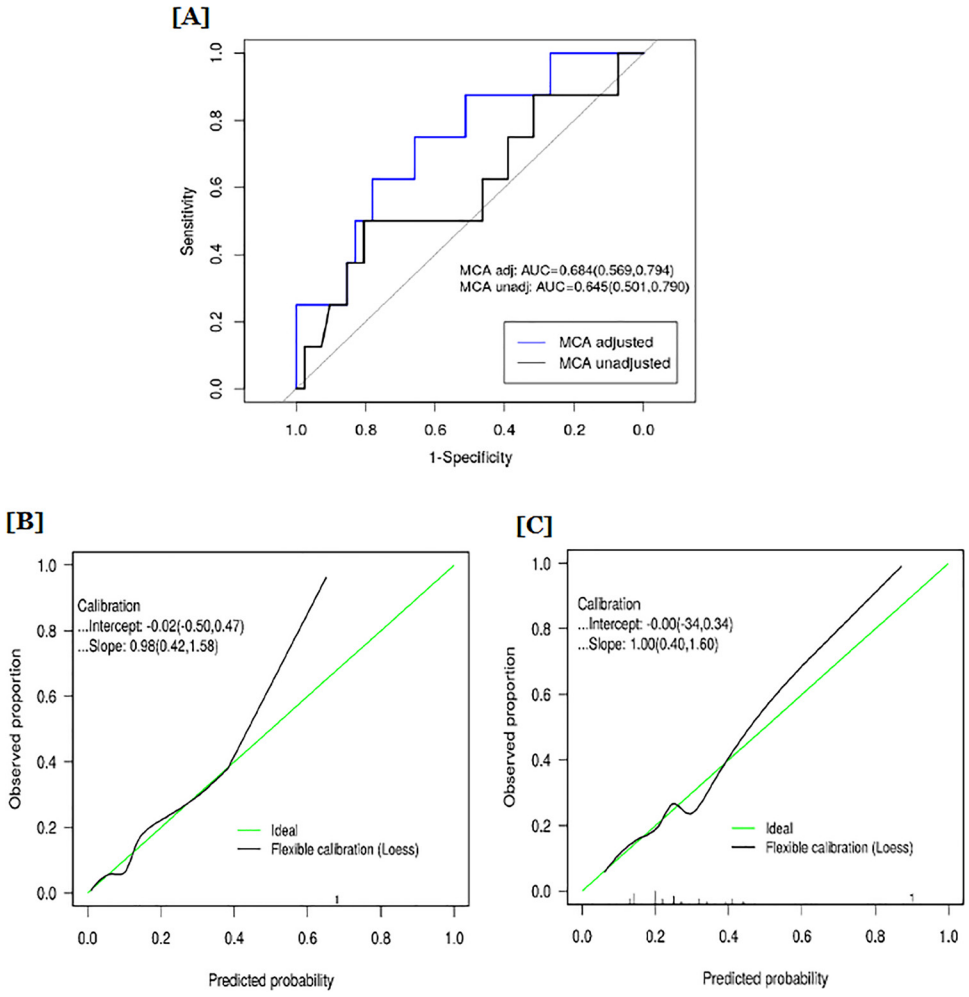

Fig. 4. Validation plots of the logistic regression classifiers. (A) – AUCs for MCA adjusted and unadjusted logistic regression (LR) classifiers. 95% Confidence intervals are presented. [B-C] Calibration plot to compare predicted and observed probabilities (B) – MCA unadjusted logistic classifier; (C) – MCA adjusted logistic classifier. Perfect calibration is indicated by the 45-degree line in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

standard for diagnosing NP. The Assessment Committee of the Neuropathic Pain Special Interest Group (NeuPSIG) of the International Association for the Study of Pain (IASP) (Haanpää et al., 2011) stated: "Clinical examination, including accurate sensory examination, is the basis of neuropathic pain diagnosis. For more accurate sensory profiling, quantitative sensory testing is recommended for selected cases in clinic, including the diagnosis of small fiber neuropathies and for research purposes." At the same time, the guideline mentioned: "It is important to emphasize that the clinical examination can never prove any pain to be of neuropathic origin, it can only provide supporting evidence for altered function of the nervous system." Therefore, it is important to note that the current standard of clinical examination is an imperfect, but an extremely important and useful "gold standard" for diagnosing NP. In the absence of clinical examination and under a resource limitation impacting patient-provider interaction, we believe the use of DN4 to capture NP was appropriate in our setting.

The novelty of our study lies in using easily obtainable information to predict high/low DN4 score and therefore use the high DN4 score as a surrogate for NPs. We hope that our model could ultimately be used for identifying NP without requiring patient-provider interaction as is currently required to obtain the standard DN4 score. Hence, in the current manuscript, our objective was to develop a predictive model for high/low DN4 score, rather than NP directly, and as such, follow-up studies would be required to assess the accuracy of our model to accurately predict NP. In the current form, the proposed models could potentially be utilized as a pre-screening tool to aid care providers, support clinical research, and facilitate informed decisions to improve patient outcomes. This is particularly important in health systems such as ours (Quebec, Canada) where patients often wait for long periods for an appointment with the healthcare provider.

Previous studies in pain research that implemented machine learning algorithms either focused on specific pain context e.g., migraine [34] or were implemented on a select group of classification algorithms accessible to direct interpretation [35]. To the best of our knowledge, the current study is the first to compare the performance of machine learning models to identify risk factors for neuropathic pain after breast cancer surgery without the need for doctor-patient interaction after 7 days post-surgery. The learning algorithms implemented in the current study have various strengths over conventional traditional methods; they have proven to be effective in capturing non-linear relations due to their hierarchical structure [29,36,37]. Another considerable strength over traditional models is they are more robust to the influence of outlier data [38].

Pain-related co-morbidities at baseline and seven days post-surgery are highly dependent [39,40]. As such, the presence of multicollinearity must be addressed during statistical modeling. Traditional regression models are negatively impacted by multicollinearity leading to unstable estimates of association effect sizes and inflated variance estimates. Such a negative impact is less consequential when the research objective is prediction instead of the interpretation of effect estimates.

However, capturing the complex association correctly is important for clinical management [41,42]. Our manuscript fills this gap and demonstrates that ML methods can be successfully deployed to characterize such association and obtain an acceptable prediction of DN4 score for BC patients after surgery. Unlike the standard DN4- questionnaire that requires a 3-month waiting period post-surgery, these models can be used promptly after surgery and can inform proper pain management and course of action for both the patients and health professional.

Our study has some limitations. A major limitation of the current study is the small size of the study. Machine learning models are prone to overfitting under insufficient amount of data which may result in misleading relationships. Neural networks for example, even under regression modeling, are highly vulnerable to small sample sizes; we therefore only considered a single layer network to address the problem of overfitting. Classification experiments are even more challenging in a difficult-to-recruit setting like ours. For example, Van der Ploeg, et al. (2014) (31) recommends about 20 events per covariate for logistic regression and approximately 200 events per covariate for modern learning algorithms. Our implementation of the logistic model was based on the generous minimum recommended event of 10 per variable (number of events: 45, number of variables: 4 for the logistic regression model (MCA unadjusted model)) suggested in [43,44]. With possible linkage to provincial electronic health records, we hope to mitigate these issues in the future. Another shortcoming of the machine learning models is they are not adequately robust to influential training features. Model performance is largely dependent on the quality of the dataset used in constructing the model. Even though recent increased interest in pain research has made more datasets available, researchers continue to grapple with sample size problems due to small sizes of subgroups in complex pain-related datasets [37].

5. Conclusions

In conclusion, the current study has identified relevant predictors for DN4-interview score based on six different machine learning methods. In order of relevance, anxiety, acute pain on movement, type of surgery and preoperative baseline pain were the most relevant predictors. Our work has demonstrated a general agreement between the constructed machine learning models, with the penalized regression models and tree-based models seemingly out-performing other prediction models. In the logistic classification model for neuropathic pain, modelled using the relevant predictors identified in the primary analysis as covariates, we showed that anxiety remains significantly associated with neuropathic pain whilst the type of surgery of the patient and acute pain on movement are seemingly indicative of the development of NP. Our findings suggest a potential prognosis tool for detecting neuropathic pain as early as 7 days after surgery without the involvement of a health professional, which could potentially improve the treatment and management of NP in patients after BC surgery.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

Financial disclosure of all authors

LJ received funding from Mitacs. PS-C's research is supported by Natural Sciences and Engineering Research Council of Canada (RGPIN-2017 − 06100), and Fonds de la recherche en santé du Québec (Salary Award). NA, MG, and AMV have no financial disclosures to report.

Summary points

What was already known?

- Neuropathic pain is diagnosed at 3 months after breast cancer (BC) surgery (with the aid of a primary care physician) using a validated DN4 questionnaire.
- Previous studies have shown neuropathic pain after BC surgery to be associated with psychological factors (anxiety and depression) and baseline pain symptoms following surgery.
- Current lack of easy-to-use prognosis tool/model that is admission soon after surgery to aid pain management and treatment.

What this study added to our knowledge?

- Anxiety, type of surgery, acute pain on movement and preoperative baseline pain are major prognosis markers for the development of neuropathic pain after BC surgery.
- Unconventional modeling techniques, e.g. gradient boosting trees, out-perform traditional regression models in capturing existing complex relationships between pain characteristics.

CRediT authorship contribution statement

Lamin Juwara: Formal analysis, Writing - original draft, Writing - review & editing. **Navpreet Arora:** Data curation, Writing - original draft, Writing - review & editing. **Mervyn Gornitsky:** Writing - original draft, Writing - review & editing. **Paramita Saha-Chaudhuri:** Supervising analysis, Writing - original draft, Writing - review & editing. **Ana M. Velly:** Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

Authors declare no conflict of interest.

Acknowledgments

We thank Harsimrat Kaur, Gurveen Gill and Neha Aggarwal (Department of Dentistry, Jewish General Hospital) for their assistance with patient recruitment and data collection; Jennifer Baoteng for her thoughtful comments; and the anonymous reviewers for their insightful recommendations that has considerably improved our manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijmedinf.2020.104170.

References

[1] D.E. Moulin, A.J. Clark, I. Gilron, M.A. Ware, C.P.N. Watson, B.J. Sessle, et al., Pharmacological management of chronic neuropathic pain - consensus statement and guidelines from the Canadian Pain Society, Pain Res. Manag. 12 (1) (2007) 13–21.
[2] N.B. Finnerup, N. Attal, S. Haroutounian, E. McNicol, R. Baron, R.H. Dworkin, et al., Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis, Lancet Neurol. 14 (February 2) (2015) 162–173.
[3] A. Binder, R. Baron, Sodium channels in neuropathic pain–friend or foe? Nat. Clin. Pract. Neurol. 3 (April 4) (2007) 179.
[4] S. Ratté, S.A. Prescott, Afferent hyperexcitability in neuropathic pain and the inconvenient truth about its degeneracy, Curr. Opin. Neurobiol. 36 (February) (2016) 31–37.
[5] M.N. Baliki, A.V. Apkarian, Nociception, pain, negative moods, and behavior selection, Neuron 87 (August 3) (2015) 474–491.
[6] C. Reyes-Gibby, P.K. Morrow, M.I. Bennett, M.P. Jensen, S. Shete, Neuropathic pain in breast cancer survivors: using the ID pain as a screening tool, J. Pain Symp. Manage. 39 (May 5) (2010) 882–889.

[7] L. Colloca, T. Ludman, D. Bouhassira, R. Baron, A.H. Dickenson, D. Yarnitsky, et al., Neuropathic pain, Nat. Rev. Dis. Primer. 3 (February) (2017) 17002.
[8] D. Bouhassira, E. Luporsi, I. Krakowski, Prevalence and incidence of chronic pain with or without neuropathic characteristics in patients with cancer, Pain 158 (June 6) (2017) 1118.
[9] Chronic Pain Has Arrived in the ICD-11 - IASP [Internet], (2019) [cited 2019 Aug 6]. Available from: https://www.iasp-pain.org/PublicationsNews/NewsDetail.aspx?ItemNumber=8340&navItemNumber=643.
[10] L. Colvin, Special section on pain: progress in pain assessment and management, BJA Br. J. Anaesth. 119 (October 4) (2017) 703–705.
[11] S. Pereira, F. Fontes, T. Sonin, T. Dias, M. Fragoso, J. Castro-Lopes, et al., Neuropathic pain after breast cancer treatment: characterization and risk factors, J. Pain Symptom Manage. 54 (6) (2017) 877–888.
[12] M. Bennett, The LANSS pain scale: the Leeds assessment of neuropathic symptoms and signs, Pain. (2020) 92–147.
[13] D. Bouhassira, N. Attal, H. Alchaar, F. Boureau, B. Brochet, J. Bruxelle, et al., Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4), Pain 114 (March 1–2) (2005) 29–36.
[14] P. Sykioti, P. Zis, A. Vadalouca, I. Siafaka, E. Argyra, D. Bouhassira, et al., Validation of the greek version of the DN4 diagnostic questionnaire for neuropathic pain, Pain Pract. Off. J. World Inst. Pain 15 (September 7) (2015) 627–632.
[15] J.G. Santos, J.O. Brito, D.C. de Andrade, V.M. Kaziyama, K.A. Ferreira, I. Souza, et al., Translation to Portuguese and validation of the Douleur Neuropathique 4 questionnaire, J. Pain 11 (5) (2010) 484–490.
[16] H.-J. Kim, J.-H. Park, D. Bouhassira, J.-H. Shin, B.-S. Chang, C.-K. Lee, et al., Validation of the korean version of the DN4 diagnostic questionnaire for neuropathic pain in patients with lumbar or lumbar-radicular pain, Yonsei Med. J. 57 (March 2) (2016) 449–454.
[17] S.P. Madani, H.R. Fateh, B. Forogh, S.-M. Fereshtehnejad, T. Ahadi, P. Ghaboussi, et al., Validity and reliability of the persian (Farsi) version of the DN 4 (Douleur neuropathique 4 questions) questionnaire for differential diagnosis of neuropathic from non-neuropathic pains, Pain Pract. 14 (5) (2014) 427–436.
[18] E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, et al., Assessing the performance of prediction models: a framework for some traditional and novel measures, Epidemiol. Camb. Mass 21 (January 1) (2010) 128–138.
[19] T.S. Jensen, R. Baron, Translation of symptoms and signs into mechanisms in neuropathic pain, Pain 102 (March 1–2) (2003) 1–8.
[20] M.M. Backonja, C.E. Argoff, Neuropathie pain-definition and implications for research and therapy, J. Neuropathic Pain Sympt Palliation 1 (2) (2005) 11–17.
[21] S.L. Thorp, Pathophysiology of cancer pain, in: A. Gulati, V. Puttanniah, B.M. Bruel, W.S. Rosenberg, J.C. Hung (Eds.), Essentials of Interventional Cancer Pain Management [Internet], Cham: Springer International Publishing, 2019, pp. 13–17, , https://doi.org/10.1007/978-3-319-99684-4_3 [cited 2019 Sep 16]. Available from:.
[22] J.M. Armer, M.E. Radina, D. Porock, S.D. Culbertson, Predicting breast cancer-related lymphedema using self-reported symptoms, Nurs. Res. 52 (December 6) (2003) 370–379.
[23] B. Rehberg, S. Mathivon, C. Combescure, Y. Mercier, G.L. Savoldelli, Prediction of acute postoperative pain following breast cancer surgery using the pain sensitivity questionnaire: a cohort study, Clin. J. Pain 33 (1) (2017) 57–66.
[24] F.E. Harrell, K.L. Lee, R.M. Califf, D.B. Pryor, R.A. Rosati, Regression modelling strategies for improved prognostic prediction, Stat. Med. 3 (June 2) (1984) 143–152.
[25] Y. Haitovsky, Multicollinearity in regression analysis: comment, Rev. Econ. Stat. 51 (4) (1969) 486–489.
[26] S. Van Buuren, Multiple imputation of discrete and continuous data by fully conditional specification, Stat. Methods Med. Res. 16 (3) (2007) 219–242.
[27] F.E.H. Harrell, Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, Springer, 2015, p. 598 p.
[28] Z. Zhang, Missing data imputation: focusing on single imputation, Ann. Transl. Med. 4 (January 1) (2016) 9.
[29] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, The elements of statistical learning: data mining, inference and prediction, Math. Intell. 27 (2) (2005) 83–85.
[30] M. Gevrey, I. Dimopoulos, S. Lek, Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecol. Model. 160 (February 3) (2003) 249–264.
[31] T. van der Ploeg, P.C. Austin, E.W. Steyerberg, Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints, BMC Med. Res. Methodol. 14 (December 1) (2014) 137.
[32] T.A. Gerds, T. Cai, M. Schumacher, The performance of risk prediction models, Biom. J. 50 (4) (2008) 457–479.
[33] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: concerns and ways forward. Hernandez Montoya AR, editor, PLoS One 13 (March 3) (2018) e0194889.
[34] A. Parthipan, I. Banerjee, K. Humphreys, S.M. Asch, C. Curtin, I. Carroll, et al., Predicting inadequate postoperative pain management in depressed patients: a machine learning approach, PLoS One 14 (2) (2019) e0210575.
[35] B. Emir, M. Kuhn, K. Johnson, B. Parsons, (414) Predictors of response to pregabalin for broad neuropathic pain: results from 11 machine learning methods from a 6-week German observational study, J. Pain 17 (April 4) (2016) S78.
[36] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016, p. 655 p.
[37] J. Lötsch, A. Ultsch, Machine learning in pain research, Pain 159 (4) (2018) 623.
[38] G.H. John, Robust Decision Trees: Removing Outliers From Databases, KDD, 1995, pp. 174–179.
[39] R. Sipilä, A.-M. Estlander, T. Tasmuth, M. Kataja, E. Kalso, Development of a screening instrument for risk factors of persistent pain after breast cancer surgery, Br. J. Cancer 107 (October) (2012) 1459.
[40] A.A. Juhl, P. Christiansen, T.E. Damsgaard, Persistent pain after breast cancer treatment: a questionnaire-based study on the prevalence, associated treatment variables, and pain type, J. Breast Cancer 19 (4) (2016) 447–454.
[41] R.F. Gunst, J.T. Webster, Regression analysis and problems of multicollinearity, Commun. Stat. Theory Methods 4 (3) (1975) 277–292.
[42] E.R. Mansfield, B.P. Helms, Detecting multicollinearity, Am. Stat. 36 (3a) (1982) 158–160.
[43] M. Pavlou, G. Ambler, S. Seaman, M. De Iorio, R.Z. Omar, Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events, Stat. Med. 35 (March 7) (2016) 1159–1177.
[44] M. Pavlou, G. Ambler, S.R. Seaman, O. Guttmann, P. Elliott, M. King, et al., How to develop a more accurate risk prediction model when there are few events, BMJ 351 (August) (2015) h3868.