

Towards Patient-centered Decision-making in Breast Cancer Surgery

Machine Learning to Predict Individual Patient-reported Outcomes at 1-year Follow-up

André Pfob,* Babak J. Mehrara, MD,† Jonas A. Nelson, MD,†
Edwin G. Wilkins, MD,‡ Andrea L. Pusic, MD,* and Chris Sidey-Gibbons, PhD§✉

Objective: We developed, tested, and validated machine learning algorithms to predict individual patient-reported outcomes at 1-year follow-up to facilitate individualized, patient-centered decision-making for women with breast cancer.

Summary of Background Data: Satisfaction with breasts is a key outcome for women undergoing cancer-related mastectomy and reconstruction. Current decision-making relies on group-level evidence which may lead to suboptimal treatment recommendations for individuals.

Methods: We trained, tested, and validated 3 machine learning algorithms using data from 1921 women undergoing cancer-related mastectomy and reconstruction conducted at eleven study sites in North America from 2011 to 2016. Data from 1921 women undergoing cancer-related mastectomy and reconstruction were collected before surgery and at 1-year follow-up. Data from 10 of the 11 sites were randomly split into training and test samples (2:1 ratio) to develop and test 3 algorithms (logistic regression with elastic net penalty, extreme gradient boosting tree, and neural network) which were further validated using the additional site's data. AUC to predict clinically-significant changes in satisfaction with breasts at 1-year follow-up using the validated BREAST-Q were the outcome measures.

Results: The 3 algorithms performed equally well when predicting both improved or decreased satisfaction with breasts in both testing and validation datasets: For the testing dataset median accuracy = 0.81 (range 0.73–0.83), median AUC = 0.84 (range 0.78–0.85). For the validation dataset median accuracy = 0.83 (range 0.81–0.84), median AUC = 0.86 (range 0.83–0.89).

Conclusion: Individual patient-reported outcomes can be accurately predicted using machine learning algorithms, which may facilitate individualized, patient-centered decision-making for women undergoing breast cancer treatment.

Keywords: breast cancer surgery, breast reconstruction, individualized treatment, machine learning, shared decision-making

(*Ann Surg* 2023;277:e144–e152)

One in 8 women will be diagnosed with breast cancer during their lifetime.¹ Advances in treatments, screening, and awareness have led to continually decreasing breast cancer-related mortality rates in the past decades. Though women diagnosed with breast cancer are more likely than ever to survive their diagnosis, many must contend with the long-term effects of their treatment on quality of life including body image, physical, psychosocial, and sexual function.

Many women with breast cancer do not have access to sufficient information to make fully informed decisions about mastectomy and reconstruction treatment options.² Impact on body image can be an important factor in women's decision-making and reconstruction after mastectomy can help minimize the impact of mastectomy on a woman's body image.^{3,4} Patient-reported outcome (PRO) data makes it possible to accurately quantify a woman's body image and satisfaction with her breasts

From the *Patient-Reported Outcomes, Value & Experience (PROVE) Center, Harvard Medical School & Brigham and Women's Hospital, Boston, Massachusetts; †Department of Plastic & Reconstructive Surgery, Memorial Sloan Kettering Cancer Center, New York, New York; ‡Department of Surgery, University of Michigan, Ann Arbor, Michigan; and §Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, Texas. ✉cgibbons@mdanderson.org.

Preliminary results presented as part of a Poster Discussion Session at 2020 ASCO Annual Meeting, abstract number 520.

Preliminary results presented at the International Society for Quality of Life Research Annual conference 2019, oral presentation "Cutting edge research plenary" on Monday, October 21, 2019, San Diego.

Supported by the National Cancer Institute Grant No. R01 CA152192 and in part by the National Cancer Institute Support Grant No. P30 CA008748 to A.L.P. and E.M.

Appropriate institutional review board or research ethics board approval was obtained from all sites.

Clinical trial information: NCT01723423.

Andrea L. Pusic: Patents, Royalties, Other Intellectual Property: Andrea L. Pusic is a co-developer of BREAST-Q and receive royalty payments when it is used in industry-sponsored trials.

The BREAST-Q is owned by Memorial Sloan Kettering Cancer Center and the University of British Columbia. A.L.P. is a co-developer of the BREAST-Q and receives royalties when it is used in for-profit industry-sponsored clinical trials.

Authors contributions: Conception and design: Chris Sidey-Gibbons, André Pfob, Andrea L. Pusic, Edwin G. Wilkins, Collection and assembly of data: Chris Sidey-Gibbons, André Pfob, Andrea L. Pusic, Edwin G. Wilkins; Data analysis and interpretation: Chris Sidey-Gibbons, André Pfob; Manuscript writing: All authors; Final approval of manuscript: All authors; Accountable for all aspects of the work: All authors; Writing assistance: No writing assistance was provided.

Data Availability: Reasonable request should be directed to cgibbons@mdanderson.org.

To gain access, data requestors will need to sign a data access agreement.

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.annals-surgery.com.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. ISSN: 0003-4932/23/27701-e144

DOI: 10.1097/SLA.0000000000004862

both pre- and postsurgery.^{5,6} These measures have been used as end-points in observational studies to provide insight into the capacity of reconstructive surgeries to maintain or even improve women's body image after mastectomy.⁷⁻⁹ Such studies have shown that autologous reconstruction is associated with improved satisfaction with breasts compared to implant-based reconstruction at both 1- and 2-year follow-up.^{7,8}

Clinical trials are the leading paradigm for identifying superior treatments and establishing of healthcare policy but they are not without limitations.¹⁰ One limitation is that individuals may experience a better or worse outcome than those suggested from the analysis of group-level data. Machine learning refers to a set of artificial intelligence techniques that seek to predict outcomes at an individual level. Thus, machine learning has the potential to augment evidence from group-level studies and meaningfully contribute to decision-making by providing individually-tailored outcome.¹¹

In this manuscript, we develop accurate prediction algorithms for women undergoing mastectomy and reconstruction for breast cancer. We analyze the way these algorithms make decisions to provide insight into the factors which are associated with changes in satisfaction with breasts after mastectomy or reconstruction. These algorithms could facilitate individualized, patient-centered, data-driven clinical decision-making allowing care to be optimized to the characteristics of the individual patient.

METHODS

Patient Recruitment

Patients were recruited as part of the multicenter prospective Mastectomy Reconstruction Outcomes Consortium study (NCT01723423). This project involved 57 plastic surgeons at 11 sites across the United States and Canada. Nine of the 11 centers were academic institutions and 2 were private practices. Appropriate institutional review or research ethics board approval was obtained from all sites. Further details on this study population are published elsewhere.^{7,8}

Women were eligible to participate in the Mastectomy Reconstruction Outcomes Consortium study if they were age 18 years or older and undergoing first-time, immediate or delayed, bilateral or unilateral postmastectomy breast reconstruction for cancer treatment or prophylaxis. Women undergoing reconstruction after previous failed attempts were excluded because of potential confounding effects. Choice of reconstructive procedure was based on patient and surgeon preference and was not randomly assigned. Patients were excluded if they did not complete a preoperative baseline quality of life questionnaire.

Variables and Analysis Data

We included 7 clinical variables, 4 patient variables, 5 PROs, and 5 socioeconomic/ethnic variables which were collected at baseline as training variables for our predictive models (see Supplemental Table 1, <http://links.lww.com/SLA/D41>). We used feature scaling to normalize the value range of all items.¹² To address the rising issue of racial bias by clinical algorithms¹³ we developed each algorithm with and without socioeconomic/ethnic variables. Additionally, algorithm performance was assessed separately for all ethnic groups.

The original cohort study and our present analysis did not include the surgeon as independent variable for several reasons:

first, although the surgeons may have been of unequal skill, reconstructions were performed by a total of 57 surgeons which balances the influence of the single surgeon. Second, trying to categorize a different level of skill is rather subjective (caseload, years of experience, patient satisfaction) and may lead to even more bias.

The full analysis data for the training, test, and additional validation set was composed of all women with complete data for baseline variables and 1-year follow-up for breast satisfaction and who underwent immediate breast reconstruction (exclusion of delayed reconstruction due to possible confounders in the meantime).

Algorithms

We trained, tested, and validated 3 machine learning algorithms with increasing complexity: logistic regression with elastic net penalty, Extreme Gradient Boosting (XGBoost) tree, and neural network. We hypothesized that more complex algorithms (XGBoost tree, neural network) would show a higher performance compared to a less complex algorithm (logistic regression with elastic net penalty) by identifying complex relations within the data. Choice of algorithm and reporting on it was informed by guidelines on how to use machine learning in medicine¹⁴ and diagnostic tests¹⁵ and previously published research by our group.^{11,16,17} We provide a short description of the algorithms below and a detailed description of the algorithm development and evaluation of our study according to these guidelines^{14,15} in the online Supplementary Appendix (see online Supplement "Algorithm development" <http://links.lww.com/SLA/D47>).

- Logistic regression with elastic net penalty.^{12,18} A key advantage of this algorithm is their interpretability and their ability to attenuate the influence of certain predictors on the model, leading to greater generalizability to new datasets.
- Extreme Gradient Boosting (XGBoost) tree model.^{19,20} Gradient boosting refers to a machine learning technique where the final prediction model consists of an ensemble of several models.¹⁹ This allows for identification of more complex inter-variable relations while still being interpretable in terms of variable importance.
- Single neural network with 5 hidden units using a logistic activation function^{21,22} and resilient backpropagation.²³ Neural networks are inspired by the structure of the human brain and consist of connected units ("Neurons") - these state-of-the-art algorithms can detect even the most complex inter-variable relations.²⁴

To provide insights into the predictions made by the models we provide Shapley additive explanations²⁵ for the XGBoost tree and local interpretable model-agnostic explanations²⁶ for the neural network.

Data from 10 of the 11 sites were randomly split into training and test samples (2:1 ratio) to develop the algorithm which was further validated using the additional site's data (Fig. 1).

"R" programming language and computing software was used for all analyses.

Outcomes

The primary outcome was accuracy in predicting whether women would experience either greater or worse satisfaction with reconstructed breasts at 1-year follow-up.

We measured this outcome using the BREAST-Q "Satisfaction with Breasts" subscale, a gold-standard measure for

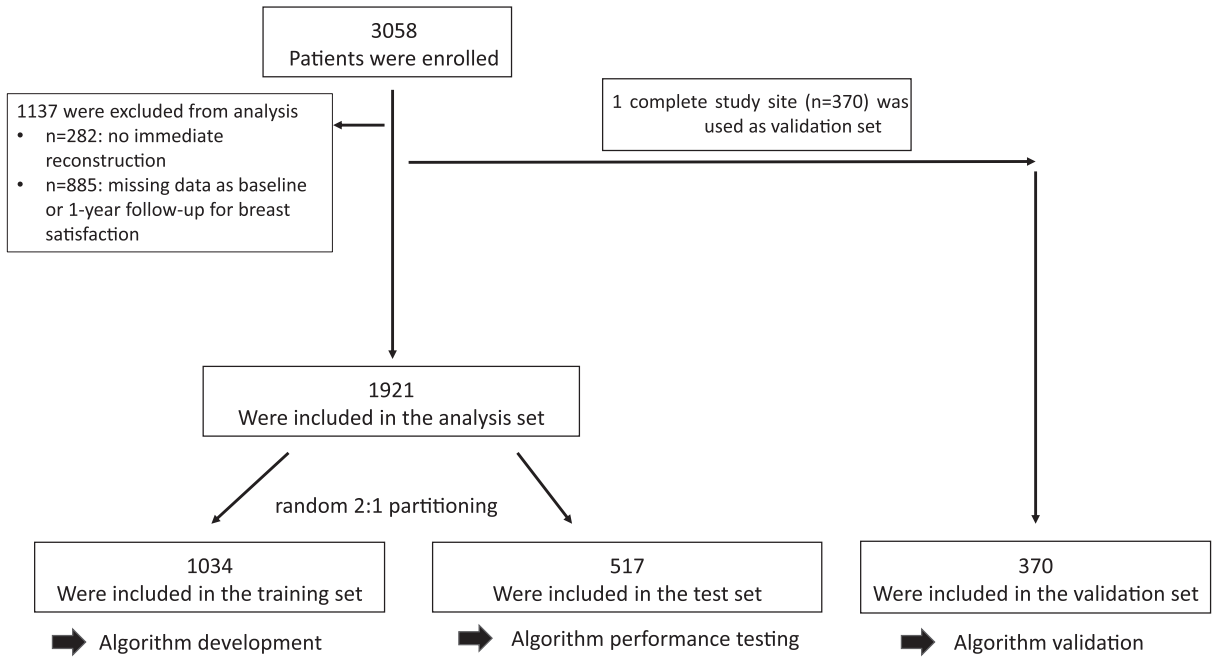


FIGURE 1. Flow of participants.

assessing PROs for women with breast cancer.^{5,6} The scale provides a linear score between 0 and 100 where 100 represents maximum satisfaction with breasts. Using these measures, we defined 3 types of outcome: worsened satisfaction, improved satisfaction or stable satisfaction. Both worsening and improving were defined as a clinically-meaningful negative or positive change in postoperative scores compared to baseline (preoperative).

To measure the outcome we used change greater than the minimal clinically important difference (MCID) for the questionnaire (1 = MCID, 0 = no MCID).²⁷ We provide point estimates of the outcome along with 95% confidence intervals (CIs).

We evaluated the performance of both algorithms among each individual study site and tested for possible differences in performance by the corrected Friedman test.²⁸

The development of our methods was informed by the Prediction Model Study Risk of Bias Assessment Tool (PROBAST).²⁹

RESULTS

Enrollment

A total of 3058 women meeting the study criteria were enrolled. The analysis dataset comprised 1921 women (1034 training, 517 testing, and 370 validation). Of the 1137 women who were excluded from the analysis, 282 did not receive an immediate reconstruction after mastectomy and 885 had missing data either as baseline or 1-year follow-up for breast satisfaction (Fig. 1: Flow of participants).

There were significantly more patients with flap reconstruction in the analysis dataset compared to all enrolled patients (P < 0.01). No differences due to exclusion were observed for the other variables.

Baseline and Clinical Characteristics

Baseline patient demographics and clinical characteristics are shown in Supplemental Table 1, <http://links.lww.com/SLA/D41>.

TABLE 1. Evaluation of Algorithms Trained to Predict Satisfaction With Breasts at 1-yr Follow-up

	1-yr Follow-up Satisfaction Lower Than Baseline		1-yr Follow-up Satisfaction Higher Than Baseline	
	Accuracy (95% CI)	AUC (95% CI)	Accuracy (95% CI)	AUC (95% CI)
Logistic regression with elastic net penalty				
Test set (n = 517)	0.83 (0.80–0.86)	0.85 (0.81–0.89)	0.76 (0.72–0.80)	0.84 (0.80–0.87)
Additional validation set (n = 370)	0.82 (0.78–0.86)	0.89 (0.86–0.92)	0.84 (0.79–0.87)	0.86 (0.81–0.90)
XGBoost tree				
Test set (n = 517)	0.83 (0.79–0.86)	0.85 (0.81–0.89)	0.75 (0.71–0.79)	0.81 (0.78–0.85)
Additional validation set (n = 370)	0.83 (0.79–0.87)	0.88 (0.84–0.92)	0.84 (0.79–0.87)	0.85 (0.81–0.89)
Neural network				
Test set (n = 517)	0.81 (0.77–0.84)	0.79 (0.75–0.84)	0.73 (0.69–0.77)	0.78 (0.74–0.82)
Additional validation set (n = 370)	0.82 (0.77–0.85)	0.84 (0.79–0.88)	0.81 (0.76–0.85)	0.83 (0.79–0.88)

AUC indicates area under the receiver operating characteristic curve; CI, confidence interval.

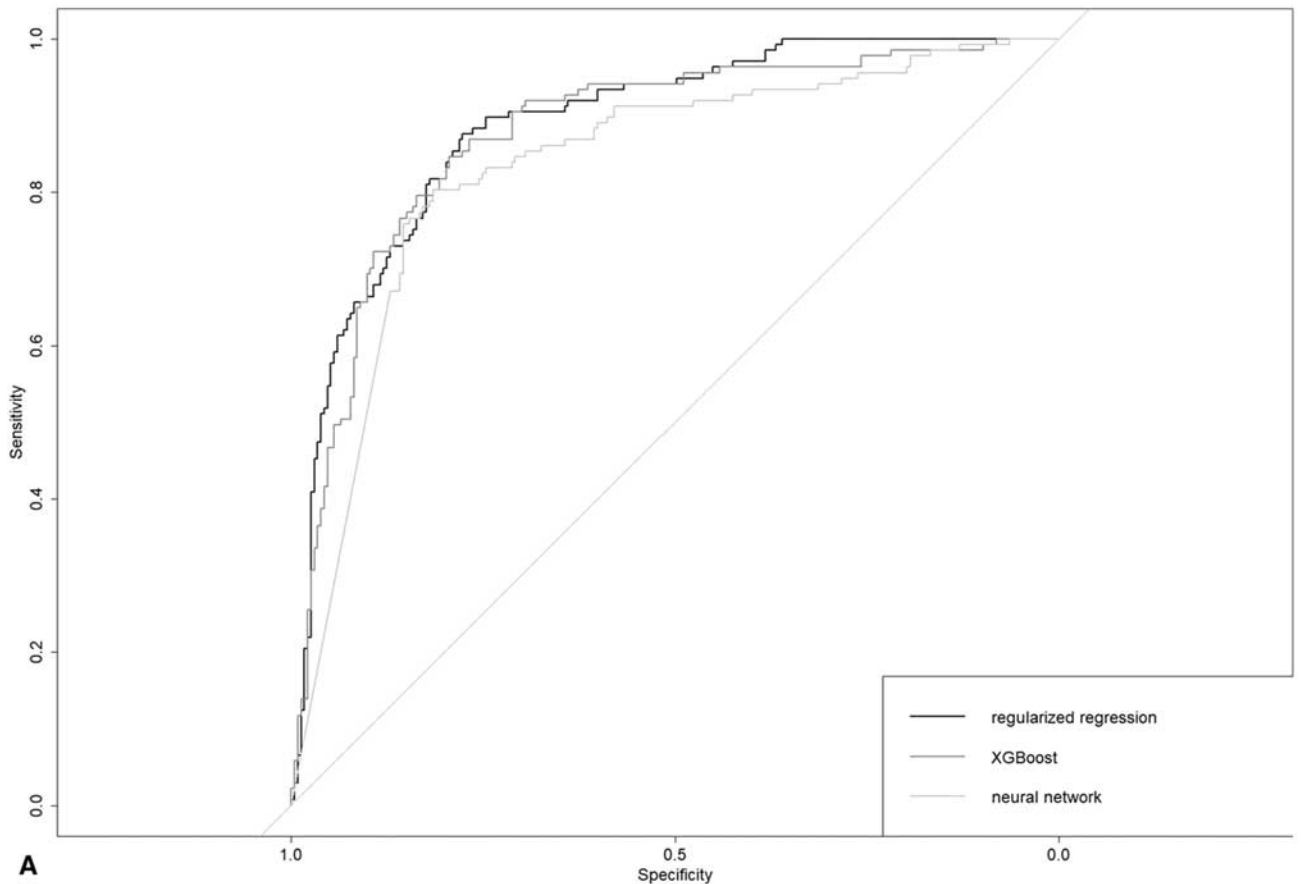


FIGURE 2. Receiver-operating characteristics curve of the elastic net regression and the neural network to predict worsened (A) and improved (B) satisfaction with breasts at 1-yr follow-up.

The mean age at baseline was 49 years (SD 10). Most mastectomies were performed due to already diagnosed breast cancer ($n = 1693$, 88.1%) with the remainder conducted prophylactically because of genetic risk.

Of the 1921 women, 571 (29.7%) experienced a clinically meaningful reduction in breast satisfaction at 1-year follow-up whilst 696 (36.2%) experienced a clinically meaningful increase in breast satisfaction.

Training, Testing, and Validation of Algorithms

Table 1 shows a summary of the logistic regression with elastic net penalty, the XGBoost tree, and the neural network algorithm we trained to predict satisfaction with breasts at 1-year follow up without socioeconomic and ethnic data. All 3 algorithms showed equally high performance to predict whether women would experience improved or worsened satisfaction with breasts compared to baseline. ROC curves are shown in Figure 2.

In the validation set, the logistic regression with elastic net penalty showed an accuracy of 0.82 (304 of 370; 95% CI, 0.78–0.86) to predict worsened satisfaction and 0.84 (309 of 370; 95% CI 0.79–0.87) to predict improved satisfaction; AUC was 0.89 (95% CI 0.86–0.92) for worsened satisfaction and 0.86 (95% CI 0.81–0.90) for improved satisfaction.

In the validation set, the XGBoost tree showed an accuracy of 0.83 (307 of 370; 95% CI, 0.79–0.87) to predict worsened satisfaction and 0.84 (309 of 370; 95% CI 0.79–0.87) to predict improved

satisfaction; AUC was 0.88 (95% CI 0.84–0.92) for worsened satisfaction and 0.85 (95% CI 0.81–0.89) for improved satisfaction.

In the validation set the neural network showed an accuracy of 0.82 (302 of 370; 95% CI, 0.77–0.85) to predict worsened satisfaction and 0.81 (299 of 370; 95% CI 0.76–0.85) to predict improved satisfaction; AUC was 0.84 (95% CI 0.79–0.88) for worsened satisfaction and 0.83 (95% CI 0.79–0.88) for improved satisfaction.

Performance between test and validation set did not differ significantly. There was no difference in algorithm performance among the individual study sites (for all algorithms $P > 0.05$). Detailed cross tabulations for all algorithms are provided in the online supplement (see Supplemental Table 2, <http://links.lww.com/SLA/D42>).

When socioeconomic and ethnic data were included as training features, the algorithm performance did not differ significantly (see Supplemental Table 3, <http://links.lww.com/SLA/D43>). In addition, the algorithms without socioeconomic and ethnic data performed equally well among all ethnic groups in the test and validation set (see Supplemental Table 4, <http://links.lww.com/SLA/D44>): median accuracy for White when predicting worsened (improved) satisfaction with breasts was 0.83, range 0.80 to 0.84, (0.81, range 0.72–0.88); for African Americans 0.87, range 0.73 to 0.90, (0.87, range 0.69–0.93); for Asians 0.87, range 0.75 to 1.00, (0.83, range 0.65–0.96). Most accuracies for American Indians could be calculated due to low numbers of patients.

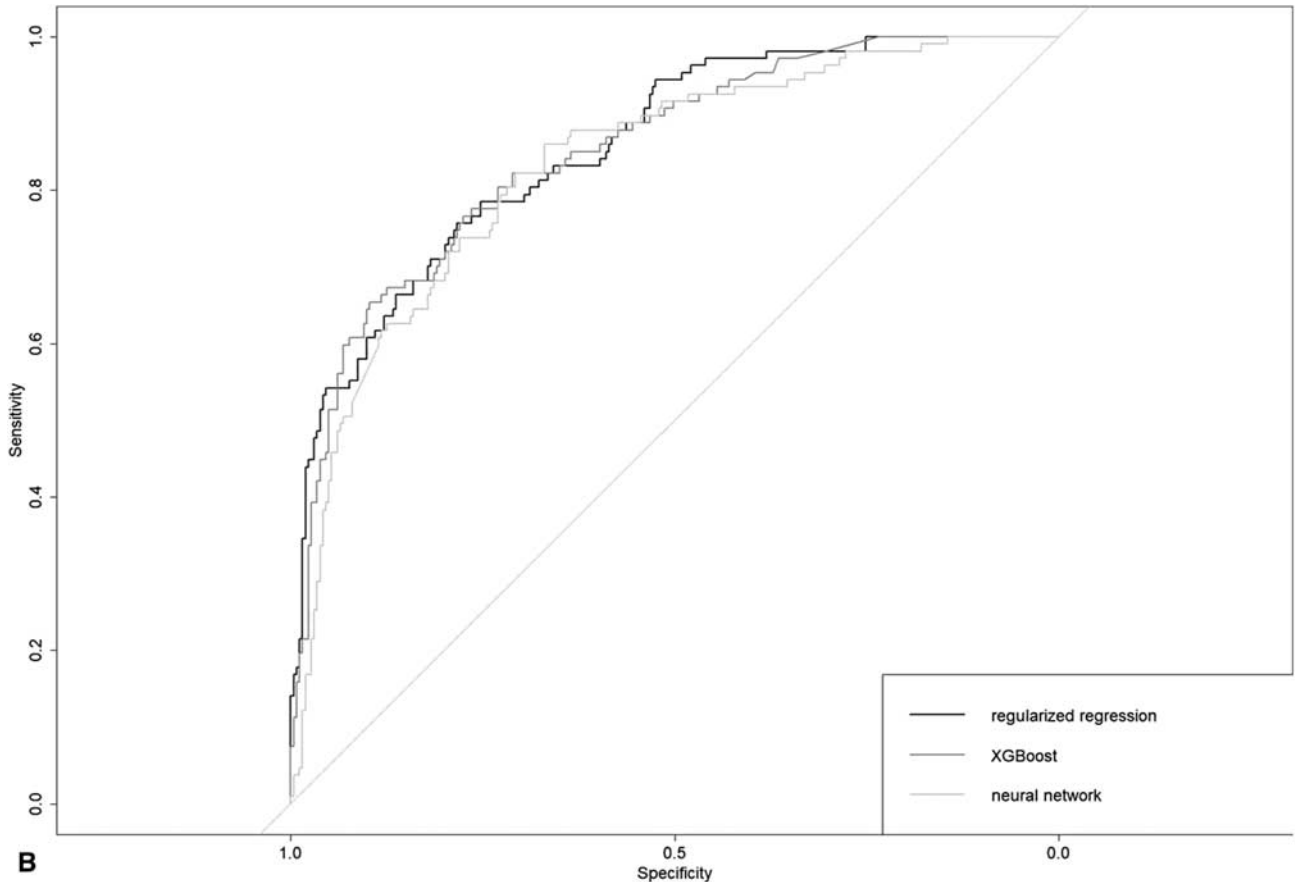


Figure 2 (Continued).

Predictive Coefficients and Insights into Variable Importance

The coefficients for the logistic regression with elastic net penalty (see Supplemental Table 5, <http://links.lww.com/SLA/D45>) clearly demonstrate the importance of baseline satisfaction with breasts in the determination of postoperative outcome, showing that a high baseline satisfaction was associated with a poorer outcome at 1-year follow up (regularized $\beta = 8.74$) and that a low baseline satisfaction was associated with a better outcome (regularized $\beta = -7.50$).

Variables related to different clinical treatments were also important in determining the predictions, though to a lesser extent:

Radiation after reconstruction increased the risk of a poorer outcome (regularized $\beta = 0.71$), whereas no radiation lowered the risk of a poorer outcome (regularized $\beta = -0.05$).

Implant-based reconstructions with tissue expanders or direct-to-implant increased the risk of a poorer outcome (regularized $\beta = 0.08, 0.34$, respectively), whereas most autologous reconstruction techniques lowered the risk of a poorer outcome: transverse rectus abdominis flap (TRAM) (regularized $\beta = -0.92$), deep inferior epigastric perforator flap (DIEP) (regularized $\beta = -0.66$), latissimus dorsi flap (regularized $\beta = -0.11$), superficial inferior epigastric artery flap (SIEA) (regularized $\beta = -1.50$). Gluteal artery perforator flap flaps, however, increased the risk of a poorer outcome (regularized $\beta = 0.59$), and mixed implant and autologous reconstruction procedure (regularized $\beta = 0.04$).

Figure 3 provides insights into the variable importance of the extreme gradient boosting tree when predicting worsened satisfaction with breasts: a higher baseline satisfaction with breasts was the most important variable for predicting worsened satisfaction with breasts. Radiation after reconstruction was associated with worsened satisfaction, as were implant based reconstruction with tissue expanders or direct-to-implant – autologous reconstruction with DIEP, TRAM, or SIEA, no radiation, simple mastectomy, and bilateral reconstruction lowered the risk of a poorer outcome.

Figure 4 provides insights into the variable importance of the neural network: a higher baseline satisfaction with breasts was the most important variable for predicting worsened satisfaction with breasts. Among the other most important variables, autologous reconstruction with SIEA, no radiation, simple mastectomy, and bilateral reconstruction lowered the risk of a poorer outcome.

When socioeconomic and ethnic data were included as training features, the coefficients for the logistic regression with elastic net penalty illustrate the drivers related to this data for determining satisfaction with breasts (see Supplemental Table 6, <http://links.lww.com/SLA/D46>): Being separated or divorced increased the risk of a poorer outcome (regularized $\beta = 0.34, 0.10$, respectively), and being unable to work (regularized $\beta = 0.37$). Full-time employment and a White ethnical background lowered the risk of a poorer outcome (regularized $\beta = -0.32, -0.45$, respectively) whereas an annual household income of $<25,000$ \$ lowered the chance of an improved outcome (regularized $\beta = -0.41, -0.32$, respectively).

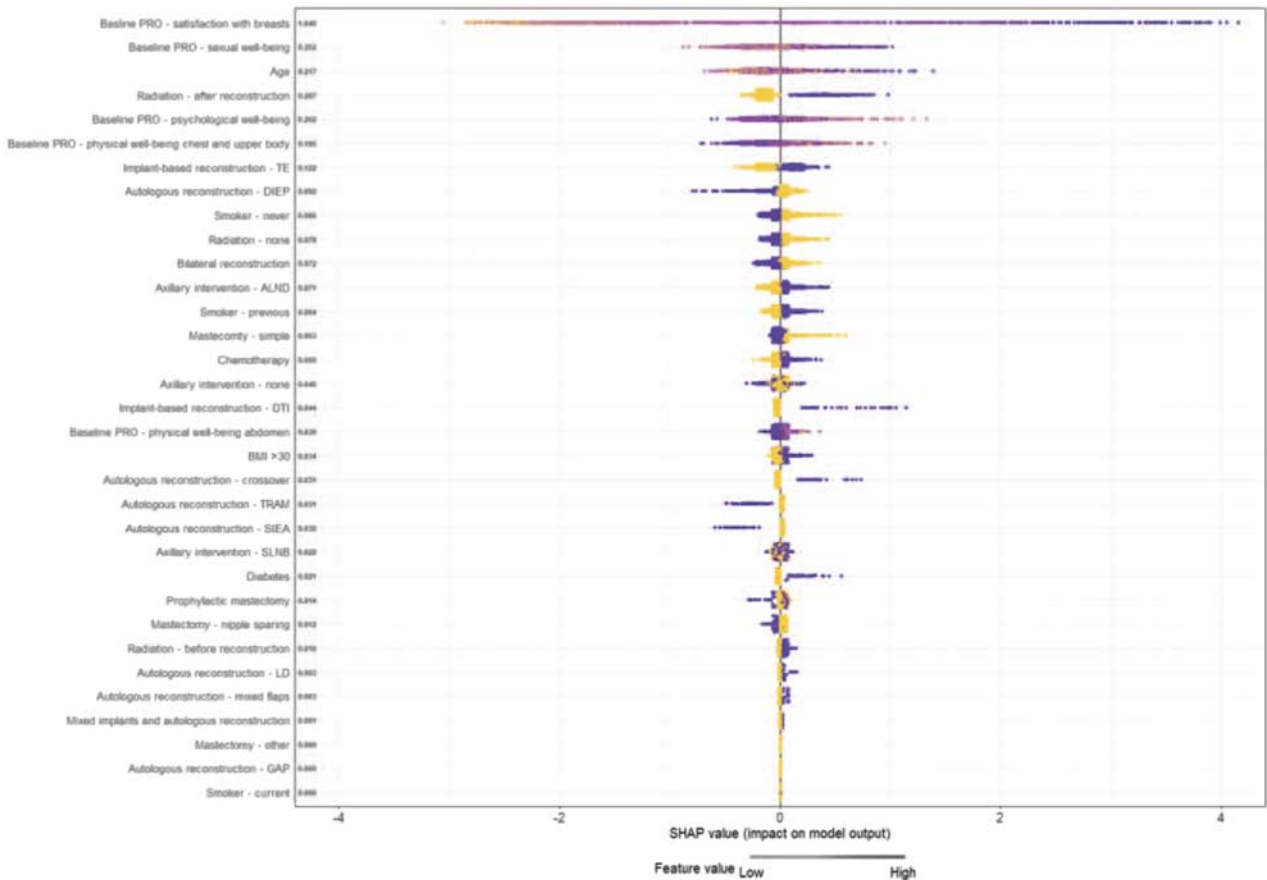


FIGURE 3. Insights into variable importance for the extreme gradient boosting tree model to predict worsened satisfaction using local interpretation methods. Shapley additive explanations (SHAP) summary plot of the extreme gradient boosting (XGBoost) tree model. Positive SHAP values on the x-axis indicate that the variable was important for predicting worsened satisfaction with breasts; negative values indicates that the variable was important for predicting no worsened satisfaction with breasts. Purple indicates a high variable value (eg, radiation after reconstruction: yes; baseline PRO - satisfaction with breasts: higher values); yellow indicates a low variable value (eg, radiation after reconstruction: no; baseline PRO - satisfaction with breasts: lower values). The values on the y-axis represent the overall global variable importance. PRO indicates patient-reported outcome.

The effects observed for patient and clinical variables and PROs at baseline were comparable for the models with and without socioeconomic and ethnic data.

DISCUSSION

We demonstrate that individual PROs at 1-year follow up of women undergoing breast cancer treatment can be accurately predicted using machine learning algorithms. About 30% of patients in our sample experienced a clinically meaningful reduction in breast satisfaction at 1-year follow-up - a majority of these women could have been identified using our predictive algorithms and alternative treatments for these individual patients could have been recommended. Thus, our results challenge the notion that different treatment procedures are associated with certain outcomes for all patients and suggest that a “one-size-fits-all” approach for clinical decision-making fails to achieve optimal outcomes for individual patients. Our algorithms facilitate individualized predictions for women considering mastectomy and reconstruction.

The current gold standard in clinical decision-making is to counsel patients based on physician preference and

group-level evidence, which may not offer optimal choice of treatment for individuals and often cannot objectively incorporate patients’ individual characteristics and needs. The key take-away for clinicians may be that instead of relying on generalizing, group-based evidence, clinicians may use algorithms like ours to confidently tailor treatment recommendations to the individual patient - this may also apply to other fields than breast cancer. The strength of these algorithms is that an outcome prediction for each individual patient is made based on the patient’s individual situation (in our example characterized by the 7 clinical, 4 patient’ and 5 PROs). Thus, our research provides a basis for the development of truly individualized, patient-centered, and data-driven tools to support clinical decision-making. There is evidence that patients benefit from preoperative decision aids³⁰ and we anticipate that the utilization of our predictive models will allow patients to more confidently make decisions (eg, about mastectomy and reconstruction) and allowing physicians to assuredly recommend the course of action that will maximize the benefit of the individual patient in their care.

Our models suggest that findings from previous studies may be more nuanced than they first appeared. Postoperative

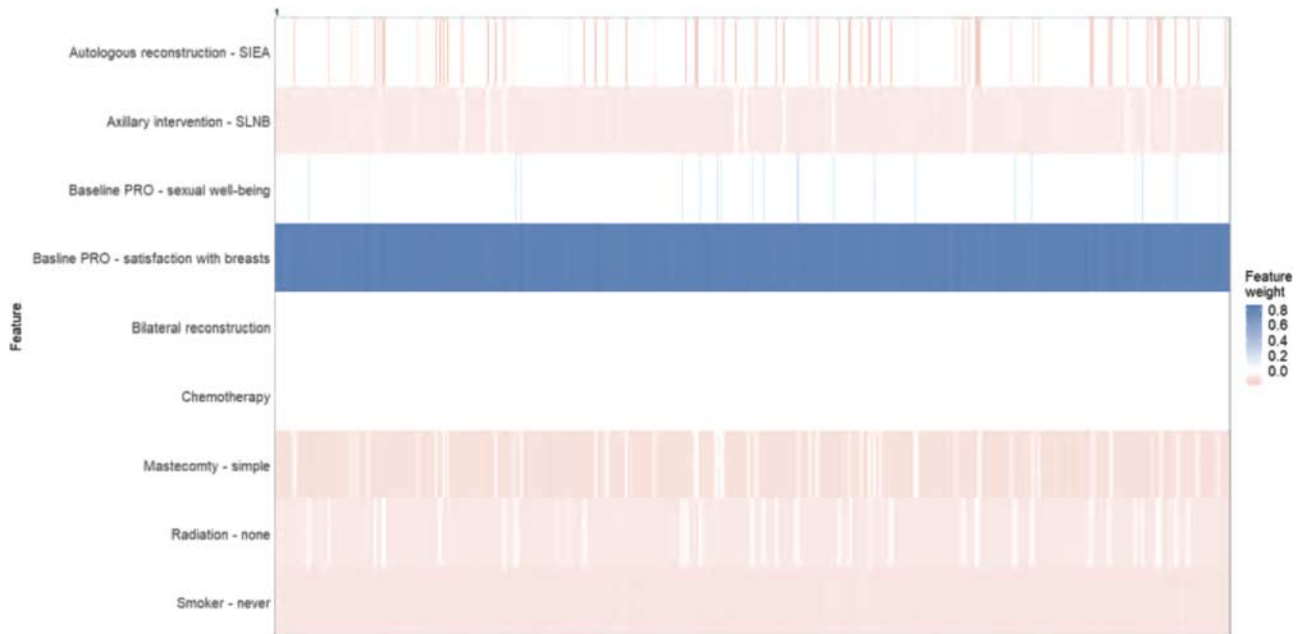


FIGURE 4. Insights into variable importance of the neural network model to predict worsened satisfaction using local interpretation methods. Local interpretable model-agnostic explanations (LIME) summary plot for the neural network and its predictions on the test set. Blue indicates that the variable was important for predicting worsened satisfaction with breasts; red indicates that the variable was important for predicting no worsened satisfaction with breasts.

outcomes are influenced more heavily by baseline PROs than by type of reconstructive surgery. This finding warrants further investigation and may have an important impact on decision-making and usefulness of observation studies with appreciable baseline differences in primary outcomes. In cases such as these, where complicated relationships exist between surgeries and related outcomes, machine learning can be used to facilitate precise estimates of individual risk, rather than relying solely on group-based statistics which may not produce ideal recommendations for individuals.³¹ Although previous group-level evidence suggests that autologous reconstruction is superior to implants,^{7,8} individual patients might benefit from alternative treatments (implant-based instead of autologous reconstruction if the algorithm predicts better PROs for the alternative treatment procedure). Also, our results suggest that not all autologous procedures are generally associated with better outcomes: although TRAM, DIEP, latissimus dorsi flap, and SIEA flaps generally lowered the risk of a poorer outcome, gluteal artery perforator flap flaps and mixed implant-autologous procedures increased the risk of a poorer outcome (see Supplemental Table 5, <http://links.lww.com/SLA/D45>).

Regarding the effect of radiation on the outcome of breast reconstruction, results from previous research are inconsistent.³² There is a general notion that radiation impairs satisfaction with reconstructed breasts⁹ but the timing of radiation therapy (before or after reconstruction for either implant or autologous reconstruction) is under intense debate and scrutiny.³² Our results suggest that on a group-level, radiation after breast reconstruction increases the risk of a poor outcome, whereas radiation before reconstruction has no effect on the outcome. However, the inconsistent results from previous studies clearly show the need for individualized outcome predictions instead of general group-based evidence.

Regarding the effect of mastectomy type on satisfaction with breasts, a recent single institution analysis of ~1900 women

concluded that nipple-sparing mastectomy offers no benefits over simple mastectomy.³³ Interestingly, after the insights into the predictions made by the XGBoost tree model and the neural network (Figs. 3 and 4) our results show that simple mastectomy actually lowers the risk of a poorer outcome compared to nipple-sparing mastectomy; although we acknowledge that we have no more detailed information (ie, incision type, preoperative breast ptosis). The finding regarding nipple-sparing mastectomy may have an important impact on policy for breast reconstruction as nipple-sparing mastectomy is currently seen as state-of-the-art surgical technique in breast cancer reconstruction.

Recently, awareness has been raised to the issue of racial bias in clinical algorithms commonly used to guide health care decisions.¹³ Incorporating socioeconomic and ethnic data as training variables in a clinical prediction model may lead to algorithms learning to make predictions based on implicit biases inherent in dataset. Our results indicate that for a model which incorporates baseline PROs the decision whether or not to include socioeconomic and ethnic data as training variables had no effect on the model performance (see Table 1 and Supplemental Table 3, <http://links.lww.com/SLA/D43>); also the main patient and clinical drivers for the predictions made remained the same (see Supplemental Table 5, <http://links.lww.com/SLA/D45> and Supplemental Table 6, <http://links.lww.com/SLA/D46>). In addition, the algorithms without socioeconomic and ethnic data performed equally well among all ethnic groups in the test and validation set (see Supplemental Table 4, <http://links.lww.com/SLA/D44>). The incorporation of baseline PROs may have enabled the algorithms to account for known individual differences among ethnic groups in terms of body image³⁴ without specifically knowing the ethnic background and thus to maximize the individual PRO while avoiding racial bias by the algorithm.

Our group has recently developed other algorithms to identify women at risk of experiencing financial toxicity related

to their breast cancer surgery.¹⁷ We believe that by combining those algorithms with those described here, care could be optimized for individuals around both PROs and financial burden.

Our study was limited by a number of factors. Although we used data from a large prospective study, the small number of operations for some procedure types limits the generalizability of our findings. In addition, although the dataset we used is the largest of its kind, in the scale of big data research, this is a small sample with a limited range of both features and cases. We hope that data-sharing initiatives will allow future research to combine more cases with a greater number of variables to improve the accuracy of such tools further. The fact that type of reconstruction was not randomized but chosen upon patient and surgeon preference may have limited the predictive accuracy, too. Another potential risk of bias is that our full analysis set consists of only 62% of the enrolled patients, however, this was mainly caused by missing information in the 1-year follow up¹ which is a known problem with PROs at long term follow-ups² and did not significantly alter the distribution of the study cohort (except for more patients with autologous than implant-based reconstruction).

Our population also had somewhat limited racial and ethnic diversity and the majority of the study sites were high-volume academic centers all located in North America. This may be reflected in the greater-than-average number of bilateral mastectomies (>50%) compared to trends in the United States and around the world. Future validation is recommended to assess the effectiveness of these tools in diverse populations.

Finally³ although we examined outcomes which are known to be important to women with breast cancer, we did not assess all relevant outcomes. It is also known that different types of reconstruction may impact physical well-being. Other, non-PROs, such as risk of complications are also important, though incorporating these was beyond the remit of the current study. Thus, the choice of outcome poses a possible risk of bias as assessed by the PROBAST tool,²⁹ though satisfaction with reconstructed breasts is a recommended key outcome by the International Consortium for Health Outcomes Measurement. All other aspects of our study would not seem to be at risk of bias when assessed by PROBAST tool, though further independent evaluation would be welcomed.

CONCLUSIONS

Machine learning algorithms trained with clinical, patient, and PRO data to predict postoperative outcomes have the potential to support clinical patient-centered decision-making and personalized care. This study demonstrates the ability to predict postoperative PROs for women undergoing mastectomy and breast reconstruction at 1-year follow up with high accuracy. Our examination of these algorithms reveals new insights into the drivers of posttreatment satisfaction with breasts for women undergoing cancer treatment. Specifically, we demonstrate that differences in baseline PRO scores have far greater influence on postoperative satisfaction with breasts than treatment decisions. The algorithm may also identify individual patients that might benefit from alternative treatments than suggested by group-level evidence (eg, implant-based vs autologous reconstruction). Further research is warranted to assess whether use of the algorithm leads to improved care and outcomes in a clinical setting.

REFERENCES

1. Howlader N, Noone A, Krapcho M, et al. *SEER Cancer Statistics Review, 1975-2016*. Bethesda, MA: National Cancer Institute; 2019.

- Fagerlin A, Lakhani I, Lantz PM, et al. An informed decision? Breast cancer patients and their knowledge about treatment. *Patient Educ Couns*. 2006;64:303–312.
- Parker PA, Peterson SK, Bedrosian I, et al. Prospective study of surgical decision-making processes for contralateral prophylactic mastectomy in women with breast cancer. *Ann Surg*. 2016;263:178–183.
- Frost MH, Hoskin TL, Hartmann LC, et al. Contralateral prophylactic mastectomy: long-term consistency of satisfaction and adverse effects and the significance of informed decision-making, quality of life, and personality traits. *Ann Surg Oncol*. 2011;18:3110–3116.
- Pusic AL, Klassen AF, Scott AM, et al. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124:345–353.
- Cano SJ, Klassen AF, Scott AM, et al. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg*. 2012;129:293–302.
- Santosa KB, Qi J, Kim HM, et al. Long-term patient-reported outcomes in postmastectomy breast reconstruction. *JAMA Surg*. 2018;153:891–899.
- Pusic AL, Matros E, Fine N, et al. Patient-reported outcomes 1 year after immediate breast reconstruction: results of the mastectomy reconstruction outcomes consortium study. *J Clin Oncol*. 2017;35:2499–2506.
- Nelson JA, Allen RJ, Polanco T, et al. Long-term patient-reported outcomes following postmastectomy breast reconstruction: an 8-year examination of 3268 patients. *Ann Surg*. 2019;270:473–483.
- Frieden TR. Evidence for health decision making—beyond randomized, controlled trials. *N Engl J Med*. 2017;377:465–475.
- Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019 191. 2019;19:1–18.
- Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16:385–395.
- Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453.
- Liu Y, Chen PHC, Krause J, et al. How to read articles that use machine learning: users, guides to the medical literature. *JAMA*. 2020;322:1806–1816.
- Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6:e012799.
- Pfob A, Sidey-Gibbons C, Lee H, et al. Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy. *Eur J Cancer*. 2021;143:134–146.
- Sidey-Gibbons C, Asaad M, Pfob A, et al. Machine learning algorithms to predict financial toxicity associated with breast cancer treatment. *J Clin Oncol*. 2020;38:2047–12047.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189–1232.
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery; 2016:785–794.
- Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. New Delhi: Pearson Education; 1999.
- Hastie T, Tibshirani R, Friedman J. *Elements of Statistical Learning*. 2nd ed. New York: Springer; 2009.
- Riedmiller M, Braun H. Direct adaptive method for faster back-propagation learning: The RPROP algorithm. In: 1993 IEEE International Conference on Neural Networks. Publ by IEEE; 1993:586–591.
- Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
- Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017. Available at: <http://arxiv.org/abs/1705.07874>. Accessed August 28, 2020.
- Ribeiro MX, Singh S, Guestrin C. “Why Should I Trust You??: Explaining the Predictions of Any Classifier. Available at: <https://arxiv.org/abs/1602.04938>. 2016. Accessed August 28, 2020.
- Voineskos SH, Klassen AF, Cano SJ, et al. Giving meaning to differences in BREAST-Q scores: minimal important difference for breast reconstruction patients. *Plast Reconstr Surg*. 2020;145:11e–20e.
- Demšar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res*. 2006;7:1–30.

29. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170:51–58.
30. Berlin NL, Tandon VJ, Hawley ST, et al. Feasibility and efficacy of decision aids to improve decision making for postmastectomy breast reconstruction: a systematic review and meta-analysis. *Med Decis Mak.* 2019;39:5–20.
31. Scott IA. Machine learning and evidence-based medicine. *Ann Intern Med.* 2018;169:44–46.
32. Ho AY, Hu ZI, Mehrara BJ, et al. Radiotherapy in the setting of breast reconstruction: types, techniques, and timing. *Lancet Oncol.* 2017;18:e742–e753.
33. Romanoff A, Zabor EC, Stempel M, et al. A comparison of patient-reported outcomes after nipple-sparing mastectomy and conventional mastectomy with reconstruction. *Ann Surg Oncol.* 2018;25:2909–2916.
34. Rubin LR, Chavez J, Alderman A, et al. Use what god has given me: difference and disparity in breast reconstruction. *Psychol Heal.* 2013;28:1099–1120.