

Machine learning for detection of lymphedema among breast cancer survivors

Mei R. Fu¹, Yao Wang², Chenge Li², Zeyuan Qiu³, Deborah Axelrod^{4,5}, Amber A. Guth^{4,5}, Joan Scagliola⁵, Yvette Conley⁶, Bradley E. Aouizerat⁷, Jeanna M. Qiu⁸, Gary Yu¹, Janet H. Van Cleave¹, Judith Haber¹, Ying Kuen Cheung⁹

¹Rory Meyers College of Nursing, ²Electrical and Computer Engineering, Tandon School of Engineering of NYU, New York University, New York, NY, USA; ³Department of Chemistry and Environmental Science, New Jersey Institute of Technology, Newark, NJ, USA; ⁴Department of Surgery, New York University School of Medicine, New York, NY, USA; ⁵NYU Perlmutter Cancer Center, New York, NY, USA; ⁶School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA; ⁷Bluestone Center for Clinical Research, Department of Oral and Maxillofacial Surgery, College of Dentistry, New York University, New York, NY, USA; ⁸Chemical and Physical Biology Concentration, Harvard University, Cambridge, MA, USA; ⁹Mailman School of Public Health, Columbia University, New York, NY, USA

Contributions: (I) Conception and design: MR Fu, Y Wang; (II) Administrative support: J Scagliola; (III) Provision of study material or patients: D Axelrod, AA Guth; (IV) Collection and assembly of data: MR Fu, J Qiu; (V) Data analysis and interpretation: MR Fu, Y Wang, C Li, Z Qiu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Mei R. Fu, PhD, RN, FAAN. Rory Meyers College of Nursing, New York University, 433 First Avenue, 4th Floor, Room 424, New York, NY 10010, USA. Email: mf67@nyu.edu.

Background: In the digital era when mHealth has emerged as an important venue for health care, the application of computer science, such as machine learning, has proven to be a powerful tool for health care in detecting or predicting various medical conditions by providing improved accuracy over conventional statistical or expert-based systems. Symptoms are often indicators for abnormal changes in body functioning due to illness or side effects from medical treatment. Real-time symptom report refers to the report of symptoms that patients are experiencing at the time of reporting. The use of machine learning integrating real-time patient-centered symptom report and real-time clinical analytics to develop real-time precision prediction may improve early detection of lymphedema and long term clinical decision support for breast cancer survivors who face lifelong risk of lymphedema. Lymphedema, which is associated with more than 20 distressing symptoms, is one of the most distressing and dreaded late adverse effects from breast cancer treatment. Currently there is no cure for lymphedema, but early detection can help patients to receive timely intervention to effectively manage lymphedema. Because lymphedema can occur immediately after cancer surgery or as late as 20 years after surgery, real-time detection of lymphedema using machine learning is paramount to achieve timely detection that can reduce the risk of lymphedema progression to chronic or severe stages. This study appraised the accuracy, sensitivity, and specificity to detect lymphedema status using machine learning algorithms based on real-time symptom report.

Methods: A web-based study was conducted to collect patients' real-time report of symptoms using a mHealth system. Data regarding demographic and clinical information, lymphedema status, and symptom features were collected. A total of 355 patients from 45 states in the US completed the study. Statistical and machine learning procedures were performed for data analysis. The performance of five renowned classification algorithms of machine learning were compared: Decision Tree of C4.5, Decision Tree of C5.0, gradient boosting model (GBM), artificial neural network (ANN), and support vector machine (SVM). Each classification algorithm has certain user-definable hyper parameters. Five-fold cross validation was used to optimize these hyper parameters and to choose the parameters that led to the highest average cross validation accuracy.

Results: Using machine learning procedures comparing different algorithms is feasible. The ANN achieved the best performance for detecting lymphedema with accuracy of 93.75%, sensitivity of 95.65%, and specificity of 91.03%.

Conclusions: A well-trained ANN classifier using real-time symptom report can provide highly accurate detection of lymphedema. Such detection accuracy is significantly higher than that achievable by current and often used clinical methods such as bio-impedance analysis. Use of a well-trained classification algorithm to detect lymphedema based on symptom features is a highly promising tool that may improve lymphedema outcomes.

Keywords: Machine learning; real-time; symptom; lymphedema

Received: 09 February 2018; Accepted: 10 April 2018; Published: 29 May 2018.

doi: 10.21037/mhealth.2018.04.02

View this article at: <http://dx.doi.org/10.21037/mhealth.2018.04.02>

Introduction

In the digital era when mHealth has emerged as an important venue for health care, the application of machine learning has proven to be a powerful tool for health care in detecting or predicting various medical conditions. Machine learning has also been shown to provide improved accuracy over conventional statistical or expert-based systems (1,2). Symptoms are often indicators of abnormal changes occurring in body functioning or manifestations of side effects from cancer treatment (3,4). Real-time symptom report refers to the report of symptoms that patients are experiencing at the time of reporting. Lymphedema is an abnormal accumulation of lymph fluid in the interstitium of the affected limb and body areas, which usually occurs 1 to 5 years or even 20 years after cancer treatment (5-7). It is one of the most dreaded late adverse effects from breast cancer treatment because of its chronic and incurable condition as well as multiple accompanying distressing symptoms (8-11). All breast cancer survivors have the risk of developing lymphedema at any time for the rest of their lives (5-7). Thus, integrating real-time patient-centered symptom report and real-time clinical analytics to develop real-time precision prediction can improve early detection of lymphedema and long term clinical decision support.

Current methods for detecting and assessing lymphedema are cumbersome and not effective in detecting early stage lymphedema (12,13). In clinical practice, clinicians very often detect or diagnose lymphedema based on their observation of swelling (12,13). Very often, lymphedema is defined in research studies as increased limb girth of 1 to 2-cm or a 100- to 200-mL or 5% to 10% increased limb volume change (LVC) in comparing affected (or lymphedematous) and unaffected limbs (12,13). Notably, when swelling can be observed or measured in terms of limb girth or limb volume, lymphedema has typically occurred

for some time, leading to poor clinical outcomes (14,15). Different methods have been used to measure lymphedema, such as water displacement, sequential circumference limb tape measurement, or infrared perometry (12,13). Several limitations are associated with these methods, including limited reliability and no published sensitivity and specificity for water displacement and time consuming for tape measurements (12). Infrared perometry is more reliable, but it is costly. Bioelectrical impedance analysis (BIA) has shown limited sensitivity by missing 34% of lymphedema cases (13). Some research demonstrates the value of comparing pre-surgery limb volume to subsequent LVCs to detect mild lymphedema (16,17), yet pre-surgery limb volume measurement is not a clinical practice worldwide, resulting in no available pre-surgery limb volume measures for the millions of breast cancer survivors who have a lifetime risk for lymphedema. Therefore, developing an assessment system that does not require pre-surgery limb volume measures, such as using machine learning for real-time lymphedema detection based on real-time symptom report, would benefit all breast cancer survivors. Moreover, a web-based assessment tool that can accurately detect lymphedema in real-time would enable cost-effective patient-specific timely intervention (18). Patients who are informed of high risk are naturally more inclined to seek treatment and follow an intervention regime more rigorously (19-21).

More than 20 symptom features reported by breast cancer survivors have been significantly associated with lymphedema; more importantly the symptom features have discrete biological mechanisms related to inflammation and lymphatic biological mechanism (4,8,10). These symptom features include patient-reported arm swelling, heaviness, tightness, firmness, pain, aching, soreness, tenderness, numbness, stiffness, tingling, burning, limb

fatigue, limb weakness, seroma formation, breast swelling, chest wall swelling, limb hotness, blistering, as well as impaired limb mobility in the shoulder, arm, elbow, wrist, and fingers (4,8,10). The greater the number of symptoms reported, the greater the limb volume increase (4,22,23). Lymphedema symptoms may indicate a critical stage of lymphedema where lymphedema is present but changes in limb volume or limb girth cannot be detected by objective measures (6,7,22,23).

We (8) and other researchers (10) have demonstrated that it is feasible, reliable, and valid to detect lymphedema status using symptom report (8,10). Despite its clear value in detecting lymphedema, the use of real-time symptom report is still limited. This is due in part to the limitations of using conventional statistical methods to identify latent relationships between lymphedema and relevant symptoms. Machine learning is a data-driven approach to learn the association between various observable features and the class label from training data (1,2,24-28). Machine learning performs high level computing to design and program explicit algorithms which is not feasible when using a conventional statistical approach. Machine learning is able to construct algorithms that can continue improving the prediction and generate automated knowledge through data-driven predictions or decisions with incoming data. Machine learning is particularly beneficial when there are many relevant features and these features are not independent, which is the case for the lymphedema symptom features (24-28). Effective machine learning tools can discover the latent relationship(s) between lymphedema and its relevant symptom features, which is difficult to identify using conventional statistical approach. As a part of a larger research project that evaluated the feasibility, reliability, and validity of real-time lymphedema symptom report using a well-established mHealth system (18), this sub-study appraised the accuracy, sensitivity, and specificity using machine learning algorithms to detect lymphedema status based on real-time symptom report.

Methods

Ethical consideration

The approval of the study (HS#10-0251) was obtained from the institutional review board of a metropolitan university in New York City, US. Confidentiality was protected by anonymous data collection without collecting any information that might identify the participants.

Study design

A web-based and cross-sectional study was designed to enable patients' real-time symptom report, that is, symptoms reported by patients at the time of reporting using a well-established mHealth system, The-Optimal-Lymph-Flow (TOLF) (18). TOLF is a unique patient-centered mHealth system to promote precision symptom assessment among breast cancer survivors. This mHealth system utilizes research-driven evidence-based measurement tools to assess symptoms, individual personal and clinical characteristics, quality of life, as well as self-care strategies. The TOLF system can be downloaded to laptops, electronic tablets (e.g., iPad™), or smart phones. The TOLF system is easy to use even for patients with minimal technical computer skills because the system only requires scrolling up and down and clicking on icons denoting specific questions or symptoms.

Participants

We recruited breast cancer survivors who met the following inclusion criteria: (I) older than 21 years of age; (II) had surgical treatment of lumpectomy or mastectomy as well as lymph node procedures either sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND); and (III) being diagnosed with or treated for lymphedema. Breast cancer survivors were excluded if they (I) had no surgical treatment for breast cancer (22,23); (II) had the occurrence of tumor metastasis; and (III) hereditary lymphedema.

Recruitment

Detailed description of recruitment and data collection procedures can be found in our prior publication regarding feasibility, reliability, and validity of real-time lymphedema symptom report using the TOLF system (18). Briefly, we recruited participants through StepUp-SpeakOut.org, a virtual community for breast cancer survivors. We sent a study invitation to members of StepUp-SpeakOut.org through an electronic newsletter and posted the study invitation on the organization's website. Participants were informed of voluntary and anonymous participation. Participants' submission of their complete study data represented their consent to the study. We had 417 women accessed the study, only 355 women provided complete study data. Data from these 355 patients were used for data analysis.

Data collection and instruments

Data were collected electronically using the following electronic instruments hosted in the TOLF system.

Demographic and medical information

We collected demographic information of age, ethnicity, education, marital status, employment status, weight and height. We also collected clinical information of diagnosis and location of breast cancer, diagnosis and location of lymphedema, chemotherapy, radiation, and treatment complications.

Lymphedema status

Self-reported lymphedema status was verified by the participants' responses to following questions: (I) "Did you have surgery for breast cancer?" (II) "Have you been diagnosed with or treated for lymphedema after breast cancer treatment?" (III) "If yes, when were you diagnosed with or treated for lymphedema after breast cancer treatment?" (IV) "What following self-care actions you use daily to manage your lymphedema?" Participants had to have affirmative answers to the questions to be classified in the lymphedema group. Participants, who provided negative answers to any of the above questions, would be classified in the non-lymphedema group.

Breast cancer and lymphedema symptom experience index (BCLE-SEI)

BCLE-SEI is a reliable and valid self-report instrument that measures the presence of 26 lymphedema symptom features. The symptom features include swelling in the affected body side (i.e., arm, hand, breast, and chest wall), heaviness, firmness, tightness, stiffness, pain/aching/soreness, numbness, tenderness, stiffness, redness, blistering, burning, stabbing, tingling (pins and needles), fibrosis (skin toughness or thickness), seroma formation (i.e., pocket of fluid formed), hotness or increased limb temperature, limb fatigue, limb weakness, impaired mobility in the affected body side (i.e., shoulder, arm, elbow, and wrist/fingers). We used a response frame of "now" for all participants to ensure the real-time presence of symptoms. The electronic version for the occurrence of symptom features demonstrated a high internal consistency (a Cronbach's alpha coefficient = 0.959) and high discriminant validity ($z = -6.938$, $P < 0.000$) (18).

Data management

As with any web-based study, the need of human intervention when dealing with electronic data is imperative. To ensure data quality, we used the human-in-the-loop (HITL) method to verify data accuracy and ensure minimum data errors (29). Data management strategies were reported in our prior publication (18).

Statistical data analysis

We used SPSS version 20.0 (Chicago, IL, USA) for statistical tests at the 0.05 significance level (2-sided) and 95% confidence intervals (CI). We used descriptive statistical tests for the participants' characteristics. We compared the participants with and without lymphedema in terms of demographic and clinical characteristics. We used Chi-Squared tests for contingency tables and one-way analysis of variance for continuous variables.

Machine learning

Since significant associations were found between symptom features and lymphedema status in this and prior research (8,10,18), we tried to develop machine learning algorithms that can classify a patient into either lymphedema or non-lymphedema class based on the 26 symptom features. We used the self-reported lymphedema status as the ground truth for the patient class. We defined "accuracy" as the percentage of patients who are correctly classified to have true lymphedema cases or non-lymphedema cases among all patients in the validation dataset. We defined "sensitivity" as the rate of the true positive lymphedema cases which measures the proportion of patients who were correctly identified as having lymphedema among those who do have it. "Specificity" refers to the true non-lymphedema cases which measures the proportion of patients who were correctly identified to have non-lymphedema among those who do not have it. The sample size of 355 participants was adequate not only for the statistical procedure but also for exploring machine learning to avoid overfitting in training a classifier with 26 symptom features, based on the recommended 5 to 10 samples-per-feature ratios for machine learning (30,31).

Overview of machine learning procedures

We compared five well-known classification algorithms:

Decision Tree C4.5 (32,33), Decision Tree C5.0 (34), gradient boosting model (GBM) (35,36), artificial neural network (ANN) (using feed forward multi-layer network) (30,37), and support vector machine (SVM) [using radial basis function (RBF) kernel function] (31). We conducted five-fold cross-validation to determine the best parameter setting for each algorithm. Specifically, the entire dataset was randomly divided into 5 data subsets. Each data subset is termed as “fold” in which the sample ratio of lymphedema and non-lymphedema reflects the sample ratio in the entire dataset. For each classifier and a candidate set of hyper parameters, we used 4 data folds to train the classifier, then evaluated its performance on the remaining 5th data fold. Such a process was repeated 5 times using a different fold for evaluation each time. We used the cross-validation accuracy, defined as the average of the classification accuracies on the 5 testing folds as the performance evaluation metric. The set of hyper parameters that led to the highest accuracy was chosen for this classifier. We further computed the standard deviation of the accuracies to characterize the variance of the classifier. By leaving out one data fold for testing, each training data subset contains approximately 284 samples. This was appropriate for the recommended 5:1 to 10:1 sample-to-feature ratio for training classifiers (33,34). A detailed description of each machine learning method is provided below.

Classification algorithms

Many powerful classification algorithms have been developed in the machine learning community including decision tree, Naïve Bayes, ANN, and SVM (25). Selection of the optimal classifier through thorough comparison of multiple classifiers for a given application has been shown to be important for clinical applications (30). We compared the performance of five renowned classification algorithms: Decision Tree of C4.5 (32,33), Decision Tree of C5.0 (34), GBM (2,38), ANN using feed forward multi-layer network (36,37), and SVM using RBF kernel function (37). Each classification algorithm has a certain user-definable parameter (called hyper parameters). We optimized these parameters using five-fold cross validation and chose the parameters that led to the highest average cross validation accuracy.

Two versions of Decision Tree implementations [i.e., C4.5 (33), C5.0 (34)] were performed and each version produced a single tree model after training. C4.5 uses the normalized information gain as a criterion for the tree

splitting decision, and uses the features that have the highest normalized information gain to split the tree at each branch. C4.5 has one hyper parameter, namely the tree depth. We found the depth of 3 achieved the best cross-validation accuracy for C4.5. Decision Tree C5.0 improves upon C4.5 and is regarded as the most advanced single-tree classifier. C5.0 first grows a large tree to fit the data closely and then prunes the initial tree by removing branches that have a relatively high error rate. One parameter for C5.0 is the Pruning Confidence Factor, which controls the severity of pruning. Smaller values that are less than the default (25%) would leave fewer nodes after pruning while larger values would lead to less pruning. The minimum case number is the other parameter, which is defined as the minimal number of samples remaining in a tree node that can be considered for further branching. Through cross validation, we found the best performance was achieved using 35% for the Pruning Confidence Factor and 8 for the minimum case number.

The other tree-based algorithm is GBM, implemented using the GBM package in R (2,38). The GBM is more robust for classification problems such as classifying lymphedema and non-lymphedema class. We trained a GBM with the binomial deviance loss function. The GBM boosting approach first resulted in a family of weak trees that were used to create a strong classifier. We used cross validation to determine the depth of each weak tree and the shrinkage factor, within [0, 1]. The shrinkage factor, also known as the learning rate, controls the contribution of each tree when added to the current model. The final model was built using depth-3 tree for each weak tree, and a shrinkage factor of 0.05.

The MATLAB Neural Network Toolbox (35,36) was used to evaluate the performance of ANN algorithm. A single hidden layer is used. We incorporated the weight decay regularization term to avoid over fitting. The best cross validation accuracy was achieved with 9 hidden nodes and a regularization coefficient of 0.3 for the weight regularization (known as `net.performParam.regularization`). When training the neural net for a given training set and a validation set, the final solution depends on the starting condition, which is randomly initialized. We ran the training program 20 times and record the one giving the highest accuracy on the validation set, which should be close to the global optimal solution for this validation set.

For the SVM model, we chose the Radial Basis Function as the kernel function, and used the package of `e1071` in R (37). We used cross validation to optimize two hyper

parameters, C and gamma. The parameter C controls the tradeoff between the classification accuracy for training examples and the simplicity of the decision surface. Higher C helps classify all training examples correctly by selecting more samples as support vectors. The parameter gamma controls the influence of a single training sample. Low gamma values spread the influence of support vectors to make the decision surface smoother. We found that the best cross-validation accuracy was achieved with gamma equal to 0.0019 and C equal to 100.

Sensitivity vs. specificity tradeoff

Given the progressive nature of lymphedema and the fact that early intervention enables better clinical outcome, it is extremely important to achieve a very high sensitivity even if at the expense of a reduced specificity. We considered this fact while training and optimizing individual classification algorithms and determining the final classifier by giving more weights to the sensitivity in the classification performance metric. This is particularly important, as the percentage of positive samples is modest (20%) due to the nature of the clinical problem. Our goal was to achieve a high sensitivity of $\geq 95\%$, while maintaining a sufficiently high specificity $\geq 85\%$.

Conventional statistic procedure vs. machine learning

To enable the comparison between machine learning and conventional statistical procedure for estimation of sensitivity and specificity, we used the R program (39) to estimate the best cutoff point for using the real-time report of the count of symptoms in detecting lymphedema based on Youden's method. Youden's method maximizes the sum of sensitivity and specificity (8,40,41). We used participants with lymphedema as the reference standard to calculate the sensitivity (i.e., true positive lymphedema cases) and specificity (i.e., true negative lymphedema cases). We used sensitivity and 1 minus specificity data over a range of lymphedema symptom assessment to create the ROC curves and calculated AUC with 95% CI.

Results

Participants

Table 1 presents detailed participants' information. In brief, 355 women submitted the complete study data; 208 (58.6%)

women reported to have lymphedema after breast cancer treatment. Over 60% of women with lymphedema ($n=126$) had the condition more than 1 year and 36.1% ($n=75$) had it less than a year. Lymphedema history ranged from 6 months to 10 years. The average time since the breast cancer diagnosis was 4.6 years (SD =6.054, range, 1–40 years). The majority of the participants self-reported as being white (91.3%), between the ages of 40 and 59 (53.0%), married (71.0%), with college or graduate degree (66.8%), and employed (56.1%). They represented 45 of the 50 states in the US with the highest representation from California (9.2%) and Texas (7.5%). Detailed demographic data were reported in our prior publication (18).

Conventional statistic procedure using Youden's method

The ROC curve for real-time symptom report (i.e., count of symptom features) as a continuous screening variable for discriminating between participants with lymphedema and those without the condition produced an AUC of 0.751 with 95% CI ($P<0.001$). A test with perfect sensitivity and specificity has an AUC of 1.0 while a test with poor sensitivity and specificity usually has an AUC less than 0.5 (16,40). The best cutoff point for real-time symptom report to detect lymphedema status was eight symptom features supported by an AUC of 0.742 (95% CI, 0.688–0.795; sensitivity of 0.731 (95% CI, 0.49–0.77); and specificity of 0.660 (95% CI, 0.655–0.860).

Machine learning for detecting lymphedema status

Among the five trained classifiers, the ANN achieved the best performance for lymphedema detection, with an average cross validation accuracy of 93.75%, sensitivity of 95.65%, and specificity of 91.03%. Other classifiers' performances were also significantly higher than that achievable by using bio-impedance analysis (13). Table 2 presents the average and standard deviation of cross validation accuracy, sensitivity, and specificity for each classifier under the optimal parameter setting.

Discussion

The risk of lymphedema for women who have undergone breast cancer treatment is lifelong and the time of onset varies; lymphedema can occur immediately after surgery, commonly between 1–5 years, or as long as 20 years later

Table 1 Demographic and clinical characteristics (n=355) (18)

Variables	Total (n=355), n (%)	Lymphedema (n=208), n (%)	Non-lymphedema (n=147), n (%)	χ^2	P
Age				10.079	0.006
21–39	37 (10.4)	30 (14.4)	7 (4.8)		
40–59	188 (53.0)	100 (48.1)	88 (60.0)		
60–80	130 (36.6)	78 (37.5)	52 (35.4)		
Education				4.211	0.378
High school or below	37 (10.4)	18 (8.7)	19 (12.9)		
Technical school	16 (4.5)	10 (4.8)	6 (4.1)		
Partial college	65 (18.3)	44 (21.2)	21 (14.3)		
College graduate	117 (33.0)	69 (33.2)	48 (32.7)		
Graduate degree	120 (33.8)	67 (32.2)	53 (36.1)		
Marital status				8.209	0.084
Married	252 (71.0)	140 (67.3)	112 (76.2)		
Partnered	17 (4.8)	8 (3.8)	9 (6.1)		
Divorced or no partner	42 (11.8)	26 (12.5)	16 (10.9)		
Widowed	31 (8.7)	23 (11.1)	8 (5.4)		
Single or never partnered	13 (3.7)	11 (5.3)	2 (1.4)		
Employment status				2.051	0.152
No	156 (43.9)	98 (47.1)	58 (39.5)		
Yes	199 (56.1)	110 (52.9)	89 (60.5)		
Ethnicity				6.329	0.176
Asian	2 (0.6)	1 (0.5)	1 (0.7)		
African American or Black	5 (1.4)	4 (1.9)	1 (0.7)		
White	324 (91.3)	190 (91.3)	134 (91.2)		
Hispanic	5 (1.4)	1 (0.5)	4 (2.7)		
Mixed	19 (5.4)	12 (5.8)	7 (4.8)		
Location of breast cancer				0.232	0.89
Left	159 (44.8)	91 (43.8)	68 (46.3)		
Right	162 (45.6)	97 (46.6)	65 (44.2)		
Both side	34 (9.6)	20 (9.6)	14 (9.5)		
Lymph nodes procedures				32.287	<0.001
None	10 (2.8)	4 (1.9)	6 (4.1)		
SLNB*	116 (32.7)	45 (21.6)	71 (48.3)		
ALND*	95 (26.8)	65 (31.2)	30 (20.4)		
Both SLNB & ALND	134 (37.7)	94 (45.2)	40 (27.2)		

Table 1 (continued)

Table 1 (continued)

Variables	Total (n=355), n (%)	Lymphedema (n=208), n (%)	Non-lymphedema (n=147), n (%)	χ^2	P
Chemotherapy				8.577	0.035
None	122 (34.4)	59 (28.4)	63 (42.9)		
Prior to surgery	44 (12.4)	28 (13.5)	16 (10.9)		
Post-surgery	179 (50.4)	113 (54.3)	66 (44.9)		
Prior & post surgery	10 (2.8)	8 (3.8)	2 (1.4)		
Radiation				21.245	<0.001
None	191 (53.8)	92 (44.2)	99 (67.3)		
Yes	164 (46.2)	116 (55.8)	48 (32.7)		
Current BMI (mean± SD)	28.19±17.59	27.95±5.70	26.55±5.45	0.055 (t/z)	0.815
BMI prior to cancer surgery (mean ± SD)	26.70±5.80	27.29±5.99	26.06±5.28	0.928 (t/z)	0.336

SLNB, sentinel lymph nodes biopsy; ALND, axillary lymph nodes dissection; BMI, body mass index; SD, standard deviation.

Table 2 Accuracy, sensitivity, specificity by different algorithms over 5 cross-validation folds

Machine learning procedures	Decision Tree: C 4.5	Decision Tree: C 5.0	Gradient boosting model	Artificial neural network	Support vector machine
Accuracy (SD)	76.31% (0.04)	77.11% (0.04)	80.50% (0.05)	93.75% (0.03)	81.65% (0.04)
Sensitivity (SD)	89.89% (0.02)	86.05% (0.08)	81.68% (0.06)	95.65% (0.03)	85.52% (0.08)
Specificity (SD)	57.00% (0.12)	64.34% (0.14)	71.97% (0.11)	91.03% (0.04)	76.14% (0.06)

SD, standard deviation.

(5-7,9). Research has shown that one of the major daily fears for breast cancer survivors is to have lymphedema besides the fear of cancer recurrence (14,15). Using mHealth technology for detection of lymphedema status is promising in that it is pragmatic and time-efficient. Our study provides evidence that the use of a mHealth system with machine learning for real-time detection of lymphedema status displayed improved accuracy, sensitivity and specificity. In comparison with conventional statistical procedures, our study further shows that a well-trained ANN classifier can offer accurate evaluation on the patient's lymphedema status using real-time symptom report by providing 93.75% of a cross-validation accuracy, 95.65% of sensitivity, and 91.03% of specificity. These results provide initial evidence that use of a well-trained classification algorithm to detect lymphedema based on the real-time symptom report using a web-and-mobile-based mHealth system is superior to a standard statistical approach. In comparison with the use of bio-impedance analysis (13), significantly higher

classification accuracies using machine learning were achieved. It is promising to use a well-trained classification algorithm to detect lymphedema status based entirely on symptom features.

Limitations and strengths of the study

One major limitation of the study is that at the time of the study we could only use the self-reported lymphedema status as the reference standard. Although we designed multiple questions to verify our participants' lymphedema status, we were not able to verify our participants' lymphedema status through medical record review or using objective measures of limb volume at the time of our study since our participants were from 45 states of the US. Nevertheless, our study provides supporting evidence for using machine learning for detecting lymphedema following breast cancer treatment. Future research requires comparing lymphedema status with objective measures of

lymphedema such as the lymph volume change measured by the infrared perometry to further validate, as well as demographic and biomarker (e.g., genetic) data to improve the algorithm.

The strengths of our study included using a reliable and valid instrument for symptom report, adequate sample size for machine learning, and optimization of the hyper-parameters of machine learning algorithms using a cross-validation methodology to develop a comparatively accurate detecting algorithm. Such an algorithm developed through machine learning can be expected to work well with future data. In addition, the use of real-time symptom report that allows the use of web-and-mobile-based mHealth system in detecting lymphedema status is another strength of the study. Women treated for breast cancer have a life-time risk of lymphedema and lymphedema can occur months or years and even decades after breast cancer treatment (5), which represents a significant challenge because lymphedema typically occurs after breast cancer patients have completed treatment. Real-time lymphedema risk assessment using a web-and-mobile-based mHealth system provides sustainable access to patients even years after completion of cancer treatment. Conducting such real-time lymphedema assessment using the web-and-mobile-based interface enables ubiquitous assessment and encourages the patient to monitor their lymphedema status without the need for and cost of a clinical visit. The embedded machine learning algorithm can infer the patient's lymphedema risk based on the self-reported symptoms and encourage patients at risk to visit their health care professionals for a formal clinical assessment and diagnosis. This will reduce the cost and increase the likelihood of early detection and intervention. The developed real-time lymphedema risk assessment using an mHealth system can also be used at the doctor's office and clinics, as a decision support tool for both the patient and doctor.

Conclusions

A mHealth system designed for real-time lymphedema risk detection is feasible, reliable, and valid. A well-trained ANN classifier using real-time symptom report provided highly accurate detection of lymphedema. Using mHealth to improve health care is paramount in the era of technology and precision health care. Our study provides a novel, pragmatic and cost-effective real-time lymphedema risk detection employing a machine learning algorithm that could be used by patients or clinicians anywhere

and anytime to engage patients to monitor their ongoing lymphedema risk and encouraging patients to seek early diagnosis and intervention. Ultimately, with ongoing data collection and future biomarker data to improve the algorithm from automated machine learning refinement, accurate and real-time detection of lymphedema will enable patients and healthcare providers to accurately monitor their lymphedema risk and seek timely intervention, and has the potential to reduce the anxiety of breast cancer survivors who have minimal or no risk of lymphedema.

Acknowledgements

Our heartfelt thanks to Stepup-speakout.org for helping administer the study and review the questionnaires! Our special thanks go to the Directors of the organization: Ms. Jane Dweck, Ms. Bonnie Pike, and Dr. Judith Nudelman for their great support and insights as patient experts.

Funding: This study was supported by National Institutes of Health (NIMHD P60 MD000538-03 and NCI 1R01CA214085-01), Judges and Lawyers for Breast Cancer Alert, Pfizer Independent Grants for Learning & Change (IGL&C) (13371953 The-Optimal Lymph-Flow™ and the Pless Center for Nursing Research of NYU Rory Meyers College of Nursing. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The approval of the study (HS#10-0251) was obtained from the institutional review board of a metropolitan university in New York City, US.

References

1. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005;34:113-27.
2. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist* 2001;29:1189-232.
3. Fu MR, LeMone P, McDaniel RW. An integrated approach to an analysis of symptom management in

- patients with cancer. *Oncol Nurs Forum* 2004;31:65-70.
4. Fu MR, Conley YP, Axelrod D, et al. Precision assessment of heterogeneity of lymphedema phenotype, genotypes and risk prediction. *Breast* 2016;29:231-40.
 5. Petrek JA, Senie RT, Peters M, et al. Lymphedema in a cohort of breast carcinoma survivors 20 years after diagnosis. *Cancer* 2001;92:1368-77.
 6. Armer JM, Stewart B. Post-breast cancer lymphedema: incidence increases from 12 to 30 to 60 months. *Lymphology* 2010;43:118-27.
 7. McLaughlin SA, Wright MJ, Morris KT, et al. Prevalence of lymphedema in women with breast cancer 5 years after sentinel lymph node biopsy or axillary dissection: objective measurements. *J Clin Oncol* 2008;26:5213-9.
 8. Fu MR, Axelrod D, Cleland CM, et al. Symptom report in detecting breast cancer-related lymphedema. *Breast Cancer* 2015;7:345-52.
 9. Paskett ED, Naughton MJ, McCoy TP, et al. The epidemiology of arm and hand swelling in premenopausal breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* 2007;16:775-82.
 10. Armer JM, Radina ME, Porock D, et al. Predicting breast cancer-related lymphedema using self-reported symptoms. *Nurs Res* 2003;52:370-9.
 11. Fu MR, Deng J, Armer JM. Putting evidence into practice: cancer-related lymphedema. *Clin J Oncol Nurs* 2014;18:68-79.
 12. Armer JM, Stewart BR. A comparison of four diagnostic criteria for lymphedema in a post-breast cancer population. *Lymphat Res Biol* 2005;3:208-17.
 13. Fu MR, Cleland CM, Guth AA, et al. L-dex ratio in detecting breast cancer-related lymphedema: reliability, sensitivity, and specificity. *Lymphology* 2013;46:85-96.
 14. McLaughlin, SA, Bagaria, S, Gibson, T, et al. Trends in risk reduction practices for the prevention of lymphedema in the first 12 months after breast cancer surgery. *J Am Coll Surg* 2013;216:380-9.
 15. Fu MR, Ridner SH, Hu SH, et al. Psychosocial impact of lymphedema: a systematic review of literature from 2004 to 2011. *Psychooncology* 2013;22:1466-84.
 16. Stout Gergich NL, Pfalzer LA, McGarvey C, et al. Preoperative assessment enables the early diagnosis and successful treatment of lymphedema. *Cancer* 2008;112:2809-19.
 17. Springer BA, Levy E, McGarvey C, et al. Pre-operative assessment enables early diagnosis and recovery of shoulder function in patients with breast cancer. *Breast Cancer Res Treat* 2010;120:135-47.
 18. Fu MR, Axelrod D, Guth AA, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. *mHealth* 2016;2:28.
 19. Fu MR, Axelrod D, Guth AA, et al. Proactive approach to lymphedema risk reduction: a prospective study. *Ann Surg Oncol* 2014;21:3481-9.
 20. Fu MR, Chen CM, Haber J, et al. The effect of providing information about lymphedema on the cognitive and symptom outcomes of breast cancer survivors. *Ann Surg Oncol* 2010;17:1847-53.
 21. Fu MR, Axelrod D, Haber J. Breast-cancer-related lymphedema: information, symptoms, and risk reduction behaviors. *J Nurs Scholarsh* 2008;40:341-8.
 22. Cormier JN, Xing Y, Zaniletti I, et al. Minimal limb volume change has a significant impact on breast cancer survivors. *Lymphology* 2009;42:161-75.
 23. Czerniec SA, Ward LC, Refshauge KM, et al. Assessment of breast cancer-related arm lymphedema--comparison of physical measurement methods and self-report. *Cancer Invest* 2010;28:54-62.
 24. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226-38.
 25. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd edition. New York: Springer, 2009.
 26. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005;34:113-27.
 27. Quinlan JR. C4.5: programs for machine learning. San Francisco: Morgan Kaufmann Publishers, 1993.
 28. Haykin S. Neural networks: a comprehensive foundation. New York: Macmillan, 1994.
 29. Fu MR, Ryan JC, Cleland CM. Lymphedema knowledge and practice patterns among oncology nurse navigators. *J Oncol Navig Surviv* 2012;3:8-15.
 30. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007;2:59-77.
 31. Somorjai RL, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics* 2003;19:1484-91.
 32. Wang Q, Ou Y, Julius AA, et al. Tracking tetrahymena pyriformis cells using decision trees. 21st Int Conf Pattern Recognit. Available online: <https://arxiv.org/abs/1207.3127>
 33. Quinlan JR. C4. 5: programs for machine learning. San

- Mateo, CA: Elsevier, 2014.
34. Quinlan JR. See5: an informal tutorial. 2017. Available online: <https://www.rulequest.com/see5-win.html>
 35. MATLAB Neural Network Toolbox: The MathWorks, Inc. 2018. Available online: <http://www.mathworks.com/products/neural-network/>
 36. Haykin SS. Neural networks: a comprehensive foundation. 2nd edition. Princeton, NJ: Prentice Hall, 1999.
 37. Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. J Stat Softw. Available online: <https://www.jstatsoft.org/v15/i09/paper>
 38. Ridgeway G. Package 'gbm' Version 2.1.3. 2017. Available online: <https://cran.r-project.org/web/packages/gbm/gbm.pdf>
 39. R Development Core Team. R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria, 2013. Available online: <http://www.R-project.org/>
 40. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.
 41. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670-5.

doi: 10.21037/mhealth.2018.04.02

Cite this article as: Fu MR, Wang Y, Li C, Qiu Z, Axelrod D, Guth AA, Scagliola J, Conley Y, Aouizerat BE, Qiu JM, Yu G, Van Cleave JH, Haber J, Cheung YK. Machine learning for detection of lymphedema among breast cancer survivors. *mHealth* 2018;4:17.