Multidisciplinary Cancer Investigation Original Article

April 2022, Volume 6, Issue 2

Prediction Axillary Lymph Node Involvement Status on Breast Cancer Data

Solmaz Sohrabei[1,*], Raheleh Salari[2], Seyed Mohammad Ayyoubzadeh[3], Alireza Atashi[4,5]

[1] Department of Development, Management and Resources; Office of Statistic and Information Technology Management, Zanjan University of Medical Sciences, Zanjan, Iran

[2] Poostchi Ophthalmology Research Center, Department of Ophthalmology, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

[3] Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Science, Tehran, Iran

[4] Department of E-Health, Virtual School, Tehran University of Medical Sciences, Tehran, Iran

[5] Medical Informatics Research Group, Clinical Research Department, Breast Cancer Research Center, Motamed Cancer Institute (ACECR), Tehran, Iran

**Corresponding author: Solmaz Sohrabei, Department of Development, Management and Resources; Office of Statistic and Information Technology Management, Zanjan University of Medical Sciences, Zanjan, Iran. Tel: +9824333018421; Fax: +982433019430. E-mail: solmazsohrabee1@gmail.com*

DOI: 10.30699/mci.6.2.180-1

Submitted: 26 September 2021

Revised: 5 December 2021

Accepted: 22 December 2021

e-Published: 12 April 2022

Keywords:
Lymph Nodes
Axilla
Decision Support Systems, Clinical
Machine Learning
Breast Neoplasms

Introduction: one of the foremost usual methods for evaluating breast cancer is the removal of axillary lymph nodes (ALN) which include complications such as edema, limited hand movements, and lymph accumulation. Although studies have shown that the sentinel gland condition represents the axillary nodules context in the mammary gland, the efficacy, and safety of the guard node biopsy need to be evaluated. Subsequently, predicting axillary lymph node status before sentinel lymph node biopsy needs regular clinical data collection and would be supportive for oncologists and could keep the clinicians away from this strategy. Predictive modeling for lymph node statues may be one way to diminish the axillary lymph node dissection (ALND) and consequences.

Methods: The database used in this study was provided by Clinical Research Department, Breast Cancer Research Center, Motamed Cancer Institute (ACECR), Tehran, Iran. It contains clinical and demographic risk factors records of 5142 breast cancer patients from which a total of 38 features were selected. We performed modeling; based on six data mining algorithms (Decision Tree, Nave Bayesian, Random Forest, Support Vector Machine, Fast Large Margin, and Gradient Boosted Tree (GBT)). For evaluating the model, we used 10-fold cross-validation in Rapid Miner v9.7.001.

Results: The results showed that the GBT model has a higher ability to predict lymph node metastasis than other models with an receiver operating characteristic (ROC) of 97%, a sensitivity of 96.59%, an accuracy of 90%, and specificity of 81%

Conclusions: Obviously, we have to diagnose cancer with a needle biopsy before surgery. Used data mining predictions and use of them to create a clinical decision support system for predicting cancer and lymph node statuses can help physicians and pathologists make the best decision for a patient's ALN surgery.

INTRODUCTION

Breast cancer is the most common cause of cancer-related mortality in developing countries and also the fifth leading cause of death worldwide [1]. Five years after diagnosis, a survival rate of 88% was estimated; while it decreased to approximately 80% after 10 years from diagnosis [2]. In recent years, breast cancer surgery has evolved from an aggressive approach to a conservative method. sentinel lymph node biopsy (SLNB); as a standard method and an alternative to axillary lymph node dissection (ALND), was used for determining the breast cancer dissemination in patients with clinically node-negative axilla [3]. One of the diagnostic problems is the positive involvement of the ALNs despite the false-negative laboratory results which is of special importance because when the positive lymph node involvement is not diagnosed, appropriate treatment may not be applied to the patient [4]. The risk of missing metastases; using SLNB can range from 0% to 22% with a false-negative rate of 10% [5]. SLNB specifically examines the axillary sentinel nodes, so differences in primary tumor sites lead to significant changes in false-negative rates [6, 7]. On the other hand, it is not wise to perform an SLNB biopsy for patients with negative lymph nodes involvement. In studies with internal mammary node biopsies, approximately 7.8% of patients with negative axillary nodes had positive internal mammary biopsies [5]. Thus, to reduce breast cancer axillary surgery as a therapeutic strategy, other techniques may be applied to diminish surgical cases. Therefore, the use of machine learning techniques on diagnostic data has helped physicians to find the details covered in the laboratory and screening data [6-12]. Furthermore, breast cancer surgery-related lymphedema (BCRL) is one of the most common and persecutor complications among patients, whether they have ALND or SLNB. However, patients who have undergone ALND surgery are at higher risk for this complication. BCRL can cause edema, pain, immobility, and a feeling of heaviness in the upper limbs which in turn reduces the quality of life of patients, Therefore, predicting the status of breast cancer lymph nodes can prevent unnecessary surgeries [13]. Using clinical and pathological factors for predicting ALN metastasis before surgery and lymph node resection will not only help oncologists to make early detection and early management but will also help patients; especially those at higher risk [3, 14]. One of the solutions used in predicting and improving medical decisions is to identify and exploit patterns and relationships and discover the required knowledge among a large number of variables by machine learning algorithms. In this way, false-positive and false-negative decisions could be reduced. Therefore, unnecessary biopsies could be reduced and save time and money [2]. The purpose of this study was to evaluate the performance of some common and useful DM algorithms for early prediction of lymph node status in patients with primary breast cancer and estimate the likelihood of ALN involvement when applied to a local and valuable dataset.

METHODS

Dataset and Data Pre-Processing

The dataset consists of 5142 records from 2010 to 2021 obtained by the Clinical Research Department, Breast Cancer Research Center, Motamed Cancer Institute (ACECR), Tehran, Iran. The dataset includes more than 100 features; including demographic and clinical-related features. After reviewing the articles and consulting with cancer experts, we selected 37 characteristics and variables related to the condition of the lymph nodes from this database to model them. The importance of each of them is shown in Figure 1.

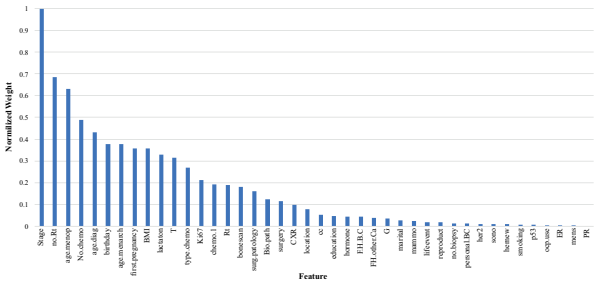

Figure 1: Normalized Weights of Different Factors; Affecting Axillary Lymph Node (ALN) by Information Gain Ratio Index

In this database, the number of patients with lymph node metastasis was 3,085 (60%) and the number of patients without lymph node metastasis was

2,056 (40%). Features with many missing values (at least 70% missing values) were removed. Missing values of other features replaced with 10-NN Rapid Miner methods (e.g., MISSING value for nominal features). In addition, features with high stability values (at least 90%) were removed. The body mass index (BMI) was calculated; using weight and height features. The values were normalized; using Z-transformation. The final dataset features are shown in Table 1.

Modeling and Evaluation

After data preprocessing, prediction models were built, using machine learning classification methods based on the number of related features on the database and regulating specific parameters to the models. Six classification methods were used in this study. For modeling and evaluation, Rapid Miner v9.7.001 was utilized. Gradient Boosted parameters were consisted of learning rate include (0.01, 0.025, 0.05, 0.1, 0.15, .2), max depth (5, 8), min sample split (0.1, 0.5, 12); and Random forest parameters were consisted of max depth (10), and min sample split (5,10). In the first step, factors normalized weights of factors affecting the ALN were calculated by Information Gain Ratio Index [15]. A multiple set of machine learning was used to generate a model for predicting ALN presence. Classification represents patterns in data that indicate whether an item belongs to a particular class of data. In Figure 2, a modeling diagram has been shown. To predict metastatic lymph node involvement in this study, the following classification methods were applied.

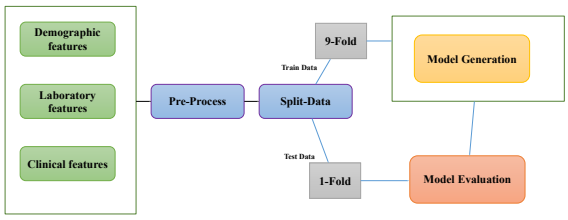

Figure 2: Modelling Diagram for Machine Learning Classification Algorithm

Decision Tree

The decision tree provides a simple way to display the impact of each event or decision, through classification [16]. The items in the data warehouse help the mining process and facilitate the creation of the forecasting model by continuously breaking the data set into smaller and more specific groups [17].

Naïve Bayes

Classification based on Naive Bayes uses a possible framework to solve early prediction or diagnosis problems and the assumption is that each attribute has an independent effect on the class [18].

Random Forest

Random forest; as a supervised model, helps to draw data to outputs during the model creating step. During training, data are given to the model during the training, data is given to the model to provide a suitable model for predicting [19]. The model offers the correlation between the data and the values that the user wants to predict. This algorithm builds a forest randomly. The constructed "forest" is a group of "Decision Trees". The constructed "forest" is a group of "Decision Trees". The construction of the forest; using trees is often done by the procedure of "bagging". The main idea of the bagging method is that a combination of learning models enhances the overall results of the model. In other words, a random forest merges several decision trees to make more precise and stable predictions [20].

Support Vector Machine (SVM)

The use of linear support vectors in classification problems is a new approach that in the training phase tries to select the decision boundary in such a way as to maximize the minimum distance with each of the desired classes [21].

Fast Large Margin

The Fast-Large Margin operator applies a rapid margin learner; based on the linear support vector learning. However, the outcome is similar to those delivered by classical SVM or logistic regression execution, this linear classifier is capable to work on data set with millions of instances and characteristics [22].

Gradient Boosted Trees (GBT)

GBT method is one of the best learning algorithms and handles many datasets with great accuracy GBT uses decision trees as the base learner and sums the predictions of a series of trees. At each step, a new decision tree is trained to fit the residual between

Table 1: Final Dataset Features

Feature Name	Description	Type	Values
NO. RT	Course of radiotherapy	Clinical Cancer treatment	Yes=1; No=0
AGE.DIAG	Age at diagnosing	Demographic	<100
AGE.MENOP	Age of menopause	Demographic	40-70
FIRST.PREGNANCY	Age at first pregnancy	Demographic	>30=1; <30=0
AGE.MENARCH	Age of menarche	Demographic	<12=1; >12=0
BMI	Body mass	Demographic	Underweight (< 18.5)=0; Normal (18.5-24.9)=1; Overweight (25.0-29.9)=2; Obese (≥30.0)=3
LACTATION	Breastfeeding status	Demographic	Yes=1; No=0
MARITAL	Marital status	Demographic	Single=1; Other=1
F.H.BC	Family history of breast cancer	Demographic	No=0; First degree (1 person)=1; Paternal second degree=2; Paternal third degree=3
OCP.USE	Duration of OCP use	Demographic	Duration, mo=1: No=0
EDUCATION	Academic education	Demographic	Yes=1; No=0
LIFE-EVENT	life event statues	Demographic	Yes=1; No=0
SMOKING	Smoking status	Demographic	Yes=1; No=0
HORMONE	Hormone therapy	Demographic	Yes=1; No=0
RT	Radiotherapy	Clinical cancer treatment	Yes=1; No=0
CHEMO	Number of chemotherapies	Clinical cancer treatment	Yes=1; No=0
TYPE.CHEMO	Type of chemotherapy	Laboratory	neo adjuvant=0; adjuvant=1
T	Tumor size	Laboratory	<2 cm=1; 2-5 cm=2; >5 cm=3
STAGE	Staging	Laboratory	Stage I, Stage II, Stage III, Stage IIV, Stage IVB, Stage IVC
CXR	Metastasis work up	Laboratory	Yes=1; No=0
BONE SCAN	Metastasis work up	Laboratory	Yes=1; No=0
KI67	Ki67 mutation	Laboratory	Negative=0; Positive=1
CHEMO.1	Chemotherapy regimen	Laboratory	Yes=1; No=0
SURGERY	Type of surgery	Laboratory	MRM right=0; MRM left=1; BCS RT=2; BCSLT=3
LOCATION	Tumor location	Laboratory	Uoq=1; Uiq=2; Loq=3; Liq=4; Central (nipple areole)=5; Axilla=6; Medial half=7; Multi focal=8; Multi centric=9
CC	Chief complaint	Laboratory	Mass or thickening=1; Skin symptom=2; Others=0
G	Grade	Laboratory	G1=1; G2=2; G3=3
NO. BIOPSY	Pathology of biopsy	Laboratory	Lobular carcinoma insitu=0; Ductal carcinoma insitu=1; Ductal carcinoma insitu=2; Invasive lobular carcinoma=3; Medullary=4; Micro invasion=5
PERSONAL.BC	Personal breast cancer	Laboratory	Yes=1; No=0
HER2	Her2	Laboratory	1+=1; 2+=2; 3+=3
ER	Estrogen receptor	Laboratory	Negative=0; Positive=1
PR	Progesterone receptor	Laboratory	Negative=0; Positive=1
P53	P53 mutation	Laboratory	Negative=0; Positive=1
HRT	Duration of HRT use	Clinical Cancer treatment	Duration, mo=1; No=0
HYSTERECTOMY	History of hysterectomy or oophorectomy bilateral with or without hysterectomy	Laboratory	Yes=1; No=0
MAMMO	Breast imaging-reporting and data system (BI-RADS)	Screening	Benign findings=0; Suspicious finding=1; Malignancy=2
SONO	Breast sonography	Screening	Benign findings=0; Suspicious finding=1; Malignancy=2
CLASS	Lymph node Involvement	Laboratory	Yes=1; No=0

ground truth and current prediction. This work causes the model to focus more on complex cases and less on issues that are easy to predict. Therefore, this method has better results than many methods such as the linear regression method and bagging method [23].

RESULTS

Factors' weights affecting ALN that were calculated by Information Gain Ratio Index are shown in Figure 2. The staging, course of radiotherapy, age of menopause, and the number of chemotherapies were the four features with the most weight on ALN presence. To evaluate models and select the best model, we considered the accuracy, sensitivity, specificity, and Receiver Operating Characteristic (ROC) indicators. Sensitivity ascertains the amount of positive and accurate predictions of a classification algorithm among examples that are positive [24].

Equation 1: $SEN = \frac{TP}{(TP+FN)}$

Where TP is the number of positive data labels correctly classified, TN is the number of negative data labels correctly classified, FP is the number of negative data labels mistakenly classified, FN is the number of positive data labels mistakenly classified.

Receiver Operating Characteristic (ROC) Analysis

The ROC curve is a graph that measures the accuracy of the model prediction on a Cartesian coordinate. The y-axis represents the correct positives (sensitivity) and the x-axis represents the correct positives (1- Specificity). The larger the surface number below the ROC curve diagram of a larger category, the better the final performance of that category [25]. The performance evaluation of models is shown in Table 2. In general, the GBT showed higher accuracy and AUC in comparison with the other models. The ROC curve of this model is shown in Figure 3.

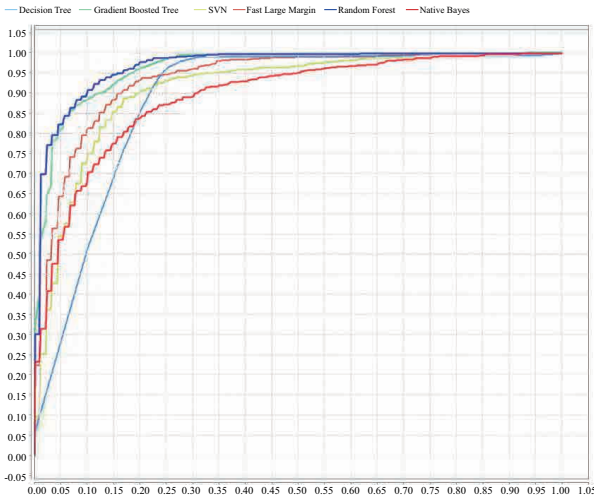

Figure 3: Receiver Operating Characteristic of Six Classification Models

DISCUSSION

In this study, some of the common machine learning algorithms were used to predict axillary lymph node involvement status in 5142 records of patients with invasive breast cancer; incorporating their clinical characteristics. Decision Tree, Naïve Bayes, Random Forest, SVM, Fast Large Margin, and Gradient Boosted were used. ROC analysis was used to evaluate the predictive models' performance. Most of the 6 models acceded high AUCs and all machine learning-based models (with AUCs; ranging from 0.885 to 0.963) had acceptable performance. In our study, most of the machine learning-based models that we developed could not indicate the association of predictive variables and the outcomes. By feature selection, the contribution of each feature to the models could be evolved. We found that the most weighted features for ALN metastasis include the stage of the tumor, course of radiotherapy, age of menopause, and the number of chemotherapy (Figure 1). The tumor stage has the highest priority

Table 2: Performance Comparison of Different Models for Predicting Axillary Lymph Node (ALN) Presence

Model	Accuracy, %	AUC	Sensitivity, %	Specificity, %
Decision Tree	90.75±0.43	0.885±0.006	99.43±0.56	1.00±0.82
Naïve Bayes	82.27±3.20	0.892±0.025	82.93±4.23	81.22±5.46
Random Forest	90.96±2.99	0.942±0.013	98.58±1.06	79.00±7.36
SVM	86.66±2.22	0.910±0.029	92.39±1.80	77.66±6.99
Fast Large Margin	87.61±3.57	0.934±0.024	90.47±4.17	83.14±6.42
Gradient Boosted Trees	90.98±2.54	0.963±0.022	96.68±2.03	81.31±6.60

in the predictor's features; while in one study there was no significant correlation between tumor stage and lymph node metastasis [16]. In the study of Samiei et al., random forest models were used to evaluate axillary under MRI analysis for prognosing ALN metastases and showed that AUC values were between 0.41–0.74 [26]. Early age at menarche has a significant effect on breast cancer and lymph node involvement because patients have more exposure to the estrogen hormone. Also, our findings showed that age at menopause is associated with lymph node involvement and survival which is consistent with the Korzeniowski study [17]. Meanwhile, some studies did not show any relation between these two factors [18-21]. Some studies mentioned that radiotherapy [22, 23] and chemotherapy [24, 25] are known as well-defined risk factors for lymphedema development. However, some other studies did not support this finding [27, 28]. However, in our study, these features were considered to be among the most effective risk factors. According to a recent meta-analysis, moderate evidence supports an association between adjuvant therapy (radiation and chemotherapy) and lymph node metastasis [28]; confirming our results.

In another study, it was stated that adjuvant therapy was not a significant risk factor for BCRL [29]. It seems, achieving more confidence in the role of adjuvant therapy in BCRL can provide a basis for further studies. Takada et al., proposed the AD Tree model and selected 15 of 24 clinic pathological variables. Their database contained 465 records of patients being referred for biopsy and their analysis revealed that the AUC was 0.77 for the prediction of ALN metastasis [14]. The rate of lymph involvement in women with breast cancer gradually increases from the age group of young premenopausal women to postmenopausal women. In the study of Li Q et al., 2,395 breast ultrasound images of 479 breast cancer patients were evaluated with deep learning the results of which showed AUCs of 0.940/0.886 when age features were selected [4]. Various studies have shown that increasing age increases the risk of lymph node involvement. This risk is much higher in women aged over 65 to 70 years old. However, younger women have a lower risk of lymph node involvement [30, 31]. Older age was identified as a predictor for ALN metastasis [3]. In our findings, old age ranks as the sixth important risk factor. In recent years, the first pregnancy occurs at an older age. Thus, the duration of estrogen exposure for women has increased. This clearly shows the link between high gestational age and an increased risk of breast cancer and lymph node involvement [17]. The size of the tumor was one of the most common factors in ascertaining ALN involvement status in patients with breast cancer in a previous study [16]. Similar to our study, they showed a correlation between tumor size and lymph node axillary metastasis. For this reason, it is best for people with breast cancer to be screened regularly so that complications such as lymph node involvement could be prevented, at least in the early stages and when the tumor size is small [16]. A linear correlation linearly was shown in Sopik's study; showing that very large tumors have little correlation with lymph node metastasis [32]. In a study by Liu Z p53 and ki67 were evaluated; while ki67 was statistically related to lymph vascular invasion [33].

Some studies mentioned that BMI directly affects lymphedema [23, 25, 34, 35]. In our study, this feature is one of the top ten features in predicting metastasis as well. A noteworthy point in previous studies is that the presence of high BMI in perimenopause causes the ovulation cycle and thus reduces the negative effects on breast tissue and lymph nodes. In contrast, with a high BMI after menopause, the likelihood of estrogen secretion by adipose tissue increases may result in increased adverse effects on breast tissue and lymph nodes [17]. Similar to the results of the current investigation, Edwards TL identified identified that tumor location was associated with lymph node metastasis [35]. Progesterone receptor status is correlated with ALN involvement [36, 37], as confirmed by artificial neural network-based models [36]. However, we achieved reverse results by using ML models in this study. In other studies, estrogen, progesterone, or HER-2 receptor status were not consistently related to lymph node status [26, 38, 39]. Nowadays the ML algorithms have been noted for their appropriate performance. ML algorithms can achieve better accuracy and specificity even in comparison to the opinions of experts and specialists panel [36]. In the study of Zhou CM et al., lymph node metastasis for patients with a history of intramucosal gastric cancer was predicted; using machine learning algorithms the results of which revealed an AUC value of 0.80 with

the Random Forrest algorithm [37] which is less than the present value. In another study conducted by Arefan D et al., the Random Forest algorithm was shown to have the highest accuracy, AUC and accuracy were 0.81 [38]. The value of the AUC of Random Forrest in another study performed by Arefan (0.74) is less than the results obtained in the current study [38]. Other algorithms; including SVM and Naïve Bayes had fewer values in comparison with our results as well. Although they extracted three features for preoperative breast DCE-MRI radionics to distinguish ALN status. Wang J et al., used ResNet50 to diagnose lymph node metastases that had higher ACC (99.98) and AUC (100). However, digital histopathological microscopic scans were used in this study [15].

Our study has some limitations. Because of retrospective data, some biases such as selection and measurement biases are unavoidable. Further studies; using different models and algorithms on different data sets are needed.

We used six models for lymph node feature analysis that were shown to have sufficient ability to predict breast cancer lymph node metastasis. We believe that the approach used here could be embedded in CDSS and provide useful information to aid clinical decision-making before starting treatment.

ACKNOWLEDGMENTS

We would like to thank Dr Alireza Atashi, a researcher at Breast Cancer Research Center, Motamed Cancer Institute, ACECR in Tehran, for his cooperation in providing the data needed for the study.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICS APPROVAL

This study was approved by Clinical Research Department, Breast Cancer Research Center, Motamed Cancer Institute (ACECR), Tehran, Iran, with Approval ID: IR.ACECR.IBCRC.REC.1394.68

REFERENCE

1. Zahoor S, Haji A, Battoo A, Qurieshi M, Mir W, Shah M. Sentinel Lymph Node Biopsy in Breast Cancer: A Clinical Review and Update. J Breast Cancer. 2017;20(3):217-27. DOI: 10.4048/jbc.2017.20.3.217 PMID: 28970846.
2. Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR. Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform. 2013;4(124):3.
3. de Oliveira Filho HR, Doria MT, Piato JR, Soares Junior JM, Filassi JR, Baracat EC, et al. Criteria for prediction of metastatic axillary lymph nodes in early-stage breast cancer. Rev Bras Ginecol Obstet. 2015;37(7):308-13. DOI: 10.1590/S0100-720320150005343 PMID: 26247250.
4. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. Sci Rep. 2017;7(1):14331. DOI: 10.1038/s41598-017-14753-7 PMID: 29085044.
5. Qu J, Shen C, Qin J, Wang Z, Liu Z, Guo J, et al. The MR radiomic signature can predict preoperative lymph node metastasis in patients with esophageal cancer. Eur Radiol. 2019;29(2):906-14. DOI: 10.1007/s00330-018-5583-z PMID: 30039220.
6. Durur-Subasi I, Durur-Karakaya A, Karaman A, Seker M, Demirci E, Alper F. Is the necrosis/wall ADC ratio useful for the differentiation of benign and malignant breast lesions? Br J Radiol. 2017;90(1073):20160803. DOI: 10.1259/bjr.20160803 PMID: 28339285.
7. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. J Am Coll Radiol. 2018;15(3 Pt B):504-8. DOI: 10.1016/j.jacr.2017.12.026 PMID: 29402533.
8. Ashturkar AV, Pathak GS, Deshmukh SD, Pandave HT. Factors predicting the axillary lymph node metastasis in breast cancer: is axillary node clearance indicated in every breast cancer patient?: factors predicting the axillary lymphnode metastases in breast cancer. Indian J Surg. 2011;73(5):331-5. DOI: 10.1007/s12262-011-0315-5 PMID: 23024536.
9. Lopez C, Bosch-Princep R, Orero G, Fontoura Balaguero L, Korzynska A, Garcia-Rojo M, et al. Peritumoral immune infiltrates in primary tumours are not associated with the presence of axillary lymph node metastasis in breast cancer: a retrospective cohort study. PeerJ. 2020;8:e9779. DOI: 10.7717/peerj.9779 PMID: 32953267.
10. Papaioannou E, Sakellakis M, Melachrinou M, Tzoracoleftherakis E, Kalofonos H, Kourea E. A Standardized Evaluation Method for FOXP3+ Tregs and CD8+ T-cells in Breast Carcinoma: Association With Breast Carcinoma Subtypes, Stage and Prognosis. Anticancer Res. 2019;39(3):1217-32. DOI: 10.21873/anticanres.13232 PMID: 30842152.
11. Houvenaeghel G, Goncalves A, Classe JM, Garbay JR, Giard S, Charytensky H, et al. Characteristics and clinical outcome of T1 breast cancer: a multicenter retrospective cohort study. Ann Oncol. 2014;25(3):623-8. DOI: 10.1093/annonc/mdt532 PMID: 24399079.
12. Yu FH, Wang JX, Ye XH, Deng J, Hang J, Yang B. Ultrasound-based radiomics nomogram: A potential biomarker to predict axillary lymph node metastasis in early-stage in-

vasive breast cancer. Eur J Radiol. 2019;119:108658. DOI: 10.1016/j.ejrad.2019.108658 PMID: 31521878.

13. Li F, Lu Q, Jin S, Zhao Q, Qin X, Jin S, et al. A scoring system for predicting the risk of breast cancer-related lymphedema. Int J Nurs Sci. 2020;7(1):21-8. DOI: 10.1016/j.ijnss.2019.12.007 PMID: 32099855.
14. Takada M, Sugimoto M, Naito Y, Moon HG, Han W, Noh DY, et al. Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model. BMC Med Inform Decis Mak. 2012;12:54. DOI: 10.1186/1472-6947-12-54 PMID: 22695278.
15. Wang J, Liu Q, Xie H, Yang Z, Zhou H. Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks. Cancers (Basel). 2021;13(4). DOI: 10.3390/cancers13040661 PMID: 33562232.
16. Mohammed AA. Predictive factors affecting axillary lymph node involvement in patients with breast cancer in Duhok: Cross-sectional study. Ann Med Surg (Lond). 2019;44:87-90. DOI: 10.1016/j.amsu.2019.07.011 PMID: 31341618.
17. Unlu O, Kiyak D, Caka C, Yagmur M, Yavas HG, Erdogan F, et al. Risk factors and histopathological features of breast cancer among women with different menopausal status and age at diagnosis. J BUON. 2017;22(1):184-91. PMID: 28365953.
18. Barnett GC, Shah M, Redman K, Easton DF, Ponder BA, Pharoah PD. Risk factors for the incidence of breast cancer: do they affect survival from the disease? J Clin Oncol. 2008;26(20):3310-6. DOI: 10.1200/JCO.2006.10.3168 PMID: 18612147.
19. Kyogoku S, Hirohata T, Takeshita S, Nomura Y, Shigematsu T, Horie A. Survival of breast-cancer patients and body size indicators. Int J Cancer. 1990;46(5):824-31. DOI: 10.1002/ijc.2910460513 PMID: 2228311.
20. Nair MK, Sankaranarayanan R, Nair KS, Amma NS, Varghese C, Padmakumari G, et al. Overall survival from breast cancer in Kerala, India, in relation to menstrual, reproductive, and clinical factors. Cancer. 1993;71(5):1791-6. DOI: 10.1002/1097-0142(19930301)71:5<1791::aid-cncr2820710513>3.0.co;2-w PMID: 8448742.
21. Mohle-Boetani JC, Grosser S, Whittemore AS, Malec M, Kampert JB, Paffenbarger RS, Jr. Body size, reproductive factors, and breast cancer survival. Prev Med. 1988;17(5):634-42. DOI: 10.1016/0091-7435(88)90056-4 PMID: 3237661.
22. Zou L, Liu FH, Shen PP, Hu Y, Liu XQ, Xu YY, et al. The incidence and risk factors of related lymphedema for breast cancer survivors post-operation: a 2-year follow-up prospective cohort study. Breast Cancer. 2018;25(3):309-14. DOI: 10.1007/s12282-018-0830-3 PMID: 29397555.
23. Shahpar H, Atieh A, Maryam A, Fatemeh HS, Massoome N, Mandana E, et al. Risk factors of lymph edema in breast cancer patients. Int J Breast Cancer. 2013;2013:641818. DOI: 10.1155/2013/641818 PMID: 23862068.
24. Paskett ED, Naughton MJ, McCoy TP, Case LD, Abbott JM. The epidemiology of arm and hand swelling in premenopausal breast cancer survivors. Cancer Epidemiol Biomarkers Prev. 2007;16(4):775-82. DOI: 10.1158/1055-9965.EPI-06-0168 PMID: 17416770.
25. Deo SV, Ray S, Rath GK, Shukla NK, Kar M, Asthana S, et al. Prevalence and risk factors for development of lymphedema following breast cancer treatment. Indian J Cancer. 2004;41(1):8-12. PMID: 15105573.
26. Samiei S, Granzier RWY, Ibrahim A, Primakov S, Lobbes MBI, Beets-Tan RGH, et al. Dedicated Axillary MRI-Based Radiomics Analysis for the Prediction of Axillary Lymph Node Metastasis in Breast Cancer. Cancers (Basel). 2021;13(4). DOI: 10.3390/cancers13040757 PMID: 33673071.
27. Park JH, Lee WH, Chung HS. Incidence and risk factors of breast cancer lymphoedema. J Clin Nurs. 2008;17(11):1450-9. DOI: 10.1111/j.1365-2702.2007.02187.x PMID: 18482142.
28. Petrek JA, Senie RT, Peters M, Rosen PP. Lymphedema in a cohort of breast carcinoma survivors 20 years after diagnosis. Cancer. 2001;92(6):1368-77. DOI: 10.1002/1097-0142(20010915)92:6<1368::aid-cncr1459>3.0.co;2-9 PMID: 11745212.
29. Kim M, Park IH, Lee KS, Ro J, Jung SY, Lee S, et al. Breast Cancer-Related Lymphedema after Neoadjuvant Chemotherapy. Cancer Res Treat. 2015;47(3):416-23. DOI: 10.4143/crt.2014.079 PMID: 25544575.
30. Wildiers H, Van Calster B, van de Poll-Franse LV, Hendrickx W, Roislien J, Smeets A, et al. Relationship between age and axillary lymph node involvement in women with breast cancer. J Clin Oncol. 2009;27(18):2931-7. DOI: 10.1200/JCO.2008.16.7619 PMID: 19451448.
31. Botteri E, Bagnardi V, Goldhirsch A, Viale G, Rotmensz N. Axillary lymph node involvement in women with breast cancer: does it depend on age? Clin Breast Cancer. 2010;10(4):318-21. DOI: 10.3816/CBC.2010.n.042 PMID: 20705566.
32. Sopik V, Narod SA. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer. Breast Cancer Res Treat. 2018;170(3):647-56. DOI: 10.1007/s10549-018-4796-9 PMID: 29693227.
33. Liu Z, Li R, Liang K, Chen J, Chen X, Li X, et al. Value of digital mammography in predicting lymphovascular invasion of breast cancer. BMC Cancer. 2020;20(1):274. DOI: 10.1186/s12885-020-6712-z PMID: 32245448.
34. Soran A, D'Angelo G, Begovic M, Ardic F, Harlak A, Samuel Wieand H, et al. Breast cancer-related lymphedema--what are the significant predictors and how they affect the severity of lymphedema? Breast J. 2006;12(6):536-43. DOI: 10.1111/j.1524-4741.2006.00342.x PMID: 17238983.
35. Edwards TL. Prevalence and aetiology of lymphoedema after breast cancer treatment in southern Tasmania. Aust N Z J Surg. 2000;70(6):412-8. DOI: 10.1046/j.1440-1622.2000.01839.x PMID: 10843395.
36. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017;318(22):2199-210. DOI: 10.1001/

jama.2017.14585 PMID: 29234806.

37. Zhou CM, Wang Y, Ye HT, Yan S, Ji M, Liu P, et al. Machine learning predicts lymph node metastasis of poorly differentiated-type intramucosal gastric cancer. Sci Rep. 2021;11(1):1300. DOI: 10.1038/s41598-020-80582-w PMID: 33446730.
38. Arefan D, Chai R, Sun M, Zuley ML, Wu S. Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features. Med Phys. 2020;47(12):6334-42. DOI: 10.1002/mp.14538 PMID: 33058224.
39. Yang J, Wang T, Yang L, Wang Y, Li H, Zhou X, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer Using Mammography-Based Radiomics Method. Sci Rep. 2019;9(1):4429. DOI: 10.1038/s41598-019-40831-z PMID: 30872652.