# Clinical Decision Support for Axillary Lymph Node Staging in Newly Diagnosed Breast Cancer Patients Based on 18F-FDG PET/MRI and Machine Learning

Janna Morawitz[1], Benjamin Sigl[2], Christian Rubbert[1], Nils-Martin Bruckmann[1], Frederic Dietzel[1], Lena J. Häberle[3], Saskia Ting[4], Svjetlana Mohrmann[5], Eugen Ruckhäberle[5], Ann-Kathrin Bittner[6], Oliver Hoffmann[6], Pascal Baltzer[2], Panagiotis Kapetas[2], Thomas Helbich[2], Paola Clauser[2], Wolfgang P. Fendler[7], Christoph Rischpler[7], Ken Herrmann[7], Benedikt M. Schaarschmidt[8], Andreas Stang[9], Lale Umutlu[8], Gerald Antoch[1], Julian Caspers*[1], and Julian Kirchner*[1]

[1]Department of Diagnostic and Interventional Radiology, Medical Faculty, University of Duesseldorf, Duesseldorf, Germany; [2]Department of Biomedical Imaging and Image-Guided Therapy, Division of General Radiology, Medical University of Vienna, Vienna, Austria; [3]Institute of Pathology, Medical Faculty, Heinrich Heine University and University Hospital Duesseldorf, Duesseldorf, Germany; [4]Institute of Pathology, University Hospital Essen, West German Cancer Center, University of Duisburg–Essen and the German Cancer Consortium (DKTK), Essen, Germany; [5]Department of Gynecology, University of Duesseldorf, Medical Faculty, Duesseldorf, Germany; [6]Department of Gynecology and Obstetrics, University Hospital Essen, University of Duisburg–Essen, Essen, Germany; [7]Department of Nuclear Medicine, University Hospital Essen, University of Duisburg–Essen and German Cancer Consortium (DKTK), Essen, Germany; [8]Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, University of Duisburg–Essen, Essen, Germany; and [9]Institute of Medical Informatics, Biometry, and Epidemiology, Essen University Medical Center, Essen, Germany

In addition to its high prognostic value, the involvement of axillary lymph nodes in breast cancer patients also plays an important role in therapy planning. Therefore, an imaging modality that can determine nodal status with high accuracy in patients with primary breast cancer is desirable. Our purpose was to investigate whether, in newly diagnosed breast cancer patients, machine-learning prediction models based on simple assessable imaging features on MRI or PET/MRI are able to determine nodal status with performance comparable to that of experienced radiologists; whether such models can be adjusted to achieve low rates of false-negatives such that invasive procedures might potentially be omitted; and whether a clinical framework for decision support based on simple imaging features can be derived from these models. **Methods:** Between August 2017 and September 2020, 303 participants from 3 centers prospectively underwent dedicated whole-body 18F-FDG PET/MRI. Imaging datasets were evaluated for axillary lymph node metastases based on morphologic and metabolic features. Predictive models were developed for MRI and PET/MRI separately using random forest classifiers on data from 2 centers and were tested on data from the third center. **Results:** The diagnostic accuracy for MRI features was 87.5% both for radiologists and for the machine-learning algorithm. For PET/MRI, the diagnostic accuracy was 89.3% for the radiologists and 91.2% for the machine-learning algorithm, with no significant differences in diagnostic performance between radiologists and the machine-learning algorithm for MRI (P = 0.671) or PET/MRI (P = 0.683). The most important lymph node feature was tracer uptake, followed by lymph node size. With an adjusted threshold, a sensitivity of 96.2% was achieved by the random forest classifier, whereas specificity, positive predictive value, negative predictive value, and accuracy were 68.2%, 78.1%, 93.8%, and 83.3%, respectively. A decision tree based on 3 simple imaging features could be established for MRI and PET/MRI. **Conclusion:** Applying a high-sensitivity threshold to the random forest results might potentially avoid invasive procedures such as sentinel lymph node biopsy in 68.2% of the patients.

**Key Words:** breast cancer; lymph node metastases; machine learning; PET/MRI

With more than 2.3 million cases in 2020, breast cancer represents the world's most prevalent cancer (*1*). In primary breast cancer, axillary lymph node involvement is the most important predictor of overall survival and recurrence in breast cancer patients (*2*) and has a decisive influence on the therapy regime. Whereas a few years ago mastectomy and extensive axillary dissection were performed in most clinically node-positive patients, advances in imaging, among other factors, have helped to make therapeutic options for local control much less invasive (*3,4*). When imaging procedures such as sonography and mammography do not reveal affected axillary lymph nodes, sentinel lymph node biopsy is now the gold standard for clinically node-negative patients (*5*). With regard to the planned therapy, this is decisive, because depending on these findings, axillary dissection and axillary radiation are further therapy options (*6*). Nearly 60% of breast carcinoma patients do not have lymph node metastases at the time of initial diagnosis (*7*). These patients, in particular, would benefit from deescalation of invasive procedures. Although the recently introduced Node-RADS (Reporting and Data System) classification tries to standardize reporting of possible lymph node metastases (*8*), no universal consensus exists on objective criteria for evaluation of metastatic disease in the axillary lymph nodes of breast cancer patients, and N staging by imaging remains a challenge (*7,9,10*).
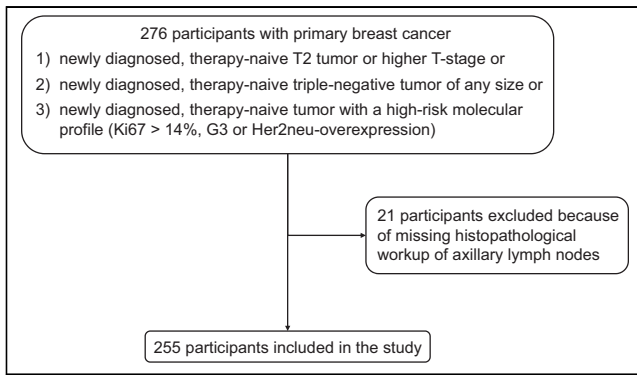
FIGURE 1. Flowchart of included and excluded participants. G3 = grade 3; Her2neu = human epidermal growth factor receptor type 2.

In recent years, artificial intelligence and machine learning have emerged strongly into the medical imaging field (11). Thus, incorporating machine-learning models into imaging-based decision-support tools has great potential to enhance diagnostic workup in breast cancer patients.

Therefore, the aim of this study was to investigate whether, in newly diagnosed breast cancer patients, machine-learning prediction models based on simple and easily assessable imaging features on MRI or PET/MRI are able to detect lymph node metastases with performance comparable to that of experienced radiologists; whether such models can be adjusted to achieve low rates of false-negatives such that invasive procedures might potentially be omitted; and whether a clinical framework for decision support based on simple imaging features can be derived from these models.

## MATERIALS AND METHODS

Because of the multiple aims of this study, the workflow was structured into 3 consecutive steps involving different methods. All calculations were based on the assessment of predefined imaging features of axillary lymph nodes by radiologists. First, machine-learning–based prediction models applying random forest classifiers were developed using the imaging features derived from the radiologist reader assessments, and their predictive performance on an independent test sample was compared with that of radiologists. Second, the random forest classifiers were adjusted to

minimize false-negative results by receiver-operating-characteristic (ROC) area-under-the-curve (AUC) optimization. Third, to facilitate a simple decision framework for everyday clinical routine, a simple decision tree classifier was trained on the imaging features independently of the optimized random forest classifiers trained beforehand.

### Participant Population, Inclusion Criteria, and Imaging Protocol

The study sample consisted of 2 samples: a training sample derived from 2 centers (University Hospital Duesseldorf and University Hospital Essen) and a testing sample from a third center (Medical University of Vienna, General Hospital).

For the training sample, 255 participants were prospectively included (Fig. 1). All had newly diagnosed, therapy-naïve breast cancer with at least one of the following criteria for a worse prognosis: a newly diagnosed, therapy-naïve T2 tumor or a higher T stage; a newly diagnosed, therapy-naïve triple-negative tumor of any size; or a newly diagnosed, therapy-naïve tumor with a high-risk molecular profile (Ki-67 > 14%, grade 3, or overexpression of human epidermal growth factor receptor type 2). All participants underwent whole-body $^{18}$F-FDG PET/MRI. Some participants have been reported before (7,12,13). This study was approved by the local ethics committees (study 6040R, 17-7396-BO + 510-2009). The test sample consisted of 48 participants. All PET/MRI examinations were performed on an integrated hybrid 3.0-T PET/MRI system (Biograph mMR; Siemens Healthcare) (14).

### Image Analysis

Imaging data from the training and test samples were analyzed by 1 reader, whereas data from the test sample were additionally rated by a second reader. MRI or PET/MRI datasets were analyzed in random order using an Osirix workstation (Pixmeo SARL). Readers were unaware of participant identity and all clinical information except for the diagnosis of breast cancer. For every participant, the presence or absence of axillary lymph node metastasis was evaluated on MRI and subsequently on PET/MRI separately. This assessment was of predefined imaging features of the most suggestive axillary lymph node for each participant. The morphologic features for the assessment of lymph node metastases were short-axis diameter in millimeters, irregular margin (yes/no), inhomogeneous cortex (yes/no), intact nodal border (yes/no), perifocal edema (yes/no), absence of fatty hilum (yes/no), and contrast medium enhancement (yes/no) (Fig. 2). On PET/MRI, tracer uptake in terms of the SUV$_{max}$ of the selected lymph node was assessed by manually drawing a region of interest around the respective lymph node. A lymph node SUV$_{max}$ ratio was calculated, with the blood pool SUV$_{max}$ of the ascending aorta as the denominator. Considering all criteria together, each reader then made a final evaluation of the lymph node status, although an absolute number of positive findings did not have to be present to evaluate the lymph node as benign or malignant.

### Reference Standard

In all participants, the histopathologic findings for the axillary lymph nodes served as the reference standard. If available, sentinel lymph node biopsy or axillary dissection was used. Otherwise, histopathologic results were derived from pretherapeutic ultrasound-guided core-needle biopsy of the suggestive lymph node. If no sufficient pretherapeutic sampling of lymph nodes was available, sentinel lymph node excision or axillary dissection after neoadjuvant systemic therapy was used as the reference standard. In these cases,
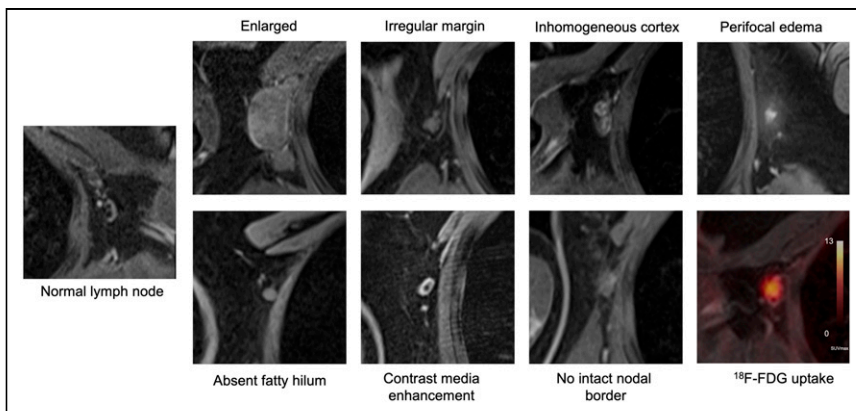


FIGURE 2. Examples of morphologic and metabolic features for assessment of axillary lymph nodes in axial T1-weighted, volume-interpolated breath-hold examination, fat-saturated, contrast-enhanced images. Enlarged lymph node has short-axis diameter of 31 mm. Lymph node with increased $^{18}$F-FDG uptake has SUV$_{max}$ of 13.1.

**TABLE 1**
Participant Demographics and Tumor Characteristics

| Parameter | Training sample | Testing sample | P |
|---|---|---|---|
| Total participants | 255 | 48 | |
| Mean age (±SD) | 51.2 ± 11.9 y | 52.2 ± 12.2 y | 0.689 |
| Lymph node status (reference standard) | | | |
| Negative | 154 (60.4%) | 26 (54.2%) | 0.420 |
| Positive | 101 (39.6%) | 22 (45.8%) | |
| Menopause status | | | |
| Premenopausal | 111 (43.5%) | 18 (37.5%) | 0.737 |
| Perimenopausal | 25 (9.8%) | 5 (10.4%) | |
| Postmenopausal | 119 (46.7%) | 25 (52.1%) | |
| Ki-67 | | | |
| Positive > 14% | 226 (88.6%) | 41 (85.4%) | 0.528 |
| Negative < 14% | 29 (11.4%) | 7 (14.6%) | |
| Progesterone status | | | |
| Positive | 169 (66.3%) | 29 (60.4%) | 0.433 |
| Negative | 86 (33.7%) | 19 (39.6%) | |
| Estrogen status | | | |
| Positive | 187 (73.3%) | 28 (58.3%) | <0.01 |
| Negative | 68 (26.7%) | 20 (41.7%) | |
| HER2neu expression | | | |
| 0 | 97 (38.0%) | 23 (47.9%) | 0.479 |
| 1+ | 73 (28.6%) | 14 (29.2%) | |
| 2+ | 34 (13.3%) | 5 (10.4%) | |
| 3+ | 51 (20.0%) | 6 (12.5%) | |
| Tumor grade | | | |
| 1 | 10 (3.9%) | 4 (8.3%) | 0.025 |
| 2 | 137 (53.7%) | 16 (33.3%) | |
| 3 | 108 (42.4%) | 28 (58.3%) | |
| Histology | | | |
| No special type | 222 (87.1%) | 42 (87.5%) | < 0.01 |
| Lobular invasive | 25 (9.8%) | 0 (0%) | |
| Other | 8 (3.1%) | 6 (12.5%) | |

HER2neu = human epidermal growth factor receptor type 2.
Data are number and percentage, except for age.

additional histopathologic preparations were evaluated, using focal fibrosis or focal necrosis as retrospective indicators of previously viable lymph node metastasis (*15,16*).

**Model Development**

Predictive models were developed for MRI and PET/MRI separately using random forest classifiers. For each modality, a random forest classifier was trained using the imaging features derived from the reader assessment as input features and the dichotomous reference standard (benign or malignant) as output.

To further optimize the classification of the models for sensitivity and minimize false-negatives (to identify a rule-out criterion), an adjusted random forest model was developed by adjusting the classification threshold of a trained random forest model on an independent validation set that was split from the training sample beforehand (80:20 stratified split) so that sensitivities of more than 0.95 were achieved on this validation set.

To additionally create more clinically interpretable classifiers, simple decision-tree classifiers with a maximum depth of 3 were additionally built using Gini impurity as the optimization criterion.

The model was developed using the scikit-learn library (version 0.24.2) in Python 3.9.

**Statistics**

For statistical analyses, SPSS Statistics (version 21; IBM) was used. Demographic participant data were reported using descriptive statistics. The Cohen κ was used to calculate interrater reliability between the 2 readers regarding prediction of lymph node status (metastatic vs. nonmetastatic) on MRI and PET/MRI. The diagnostic performance of the radiologists and machine-learning models for lymph node status on

TABLE 2
**TABLE 2**
Diagnostic Performance of MRI and PET/MRI in Assessment of Lymph Node Status of Radiologists and Random Forest Classifier Within Testing Sample

| Assessor | MRI | PET/MRI |
|---|---|---|
| Radiologists | | |
|   Sensitivity | 84.6 (65.1–95.6) | 92.3 (74.9–99.1) |
|   Specificity | 90.9 (70.8–98.9) | 86.4 (65.1–97.1) |
|   PPV | 91.7 (74.4–97.7) | 88.9 (73.5–96.8) |
|   NPV | 83.3 (66.8–92.6) | 90.5 (71.3–97.3) |
|   Accuracy | 87.5 (74.8–95.3) | 89.6 (77.3–96.5) |
| Random forest algorithm | | |
|   Sensitivity | 88.5 (69.9–97.6) | 88.5 (69.9–97.6) (reader 1), 88.5 (69.9–97.6) (reader 2) |
|   Specificity | 86.4 (65.1–97.1) | 86.4 (65.1–97.1) (reader 1), 81.8 (59.7–94.8) (reader 2) |
|   PPV | 88.5 (72.6–95.7) | 88.5 (72.6–95.7) (reader 1), 85.2 (70.1–93.4) (reader 2) |
|   NPV | 86.4 (68.3–94.9) | 86.4 (68.3–94.9) (reader 1), 85.7 (67.0–94.7) (reader 2) |
|   Accuracy | 87.5 (74.8–96.3) | 87.5 (74.8–96.3) (reader 1), 85.4 (72.2–93.9) (reader 2) |

Data are percentages, with 95% CIs in parentheses.

MRI and PET/MRI was assessed by determining sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and ROC AUC. A McNemar test was used to compare the diagnostic performance of the radiologists with that of the machine-learning models. A Pearson $\chi^2$ test was used to compare the tumor characteristics between the training and validation samples. Statistical significance was defined as a $P$ value of less than 0.05.

## RESULTS

### Participant Demographics and Reference Standard

In this study, 255 female participants (mean age, 51.2 ± 11.9 y) from 2 centers were included for the training sample (Fig. 1). According to the reference standard, 101 of the 255 (39.6%) were node-positive and 154 (60.4%) were node-negative.

For the testing sample, 48 female participants (mean age, 52.2 ± 12.2 y) from a third center were evaluated. According to the reference standard, 26 of the 48 (54.2%) were node-negative and 22 (45.8%) were node-positive. The demographics and tumor characteristics of all participants are in Table 1.

### Radiologist Performance

On the basis of MRI data, the radiologist was able to determine the correct lymph node status in 218 of 255 participants (85.5%) in the training set. This yielded a diagnostic performance indicated by sensitivity, specificity, PPV, NPV, and accuracy of 74.3%, 92.9%, 87.2%, 84.6%, and 85.5%, respectively, for the training sample (Supplemental Table 1; supplemental materials are available at http://jnm.snmjournals.org). Corresponding results for radiologist performance (identical results for both readers) based on MRI in the testing sample were 84.6%, 90.9%, 91.7%, 83.3%, and 87.5% (Table 2).

When taking PET/MRI into account, the radiologist was able to determine the correct lymph node status in 221 of 255 participants (86.7%), and sensitivity, specificity, PPV, NPV, and accuracy were 84.0%, 88.4%, 82.4%, 89.5%, and 86.7%, respectively, for the training sample (Supplemental Table 1). In the testing sample,

radiologist performance on PET/MRI data was 92.3%, 86.4%, 88.9%, 90.5%, and 89.6%, respectively (Table 2).

With regard to the individual features, there were isolated differences in the subjective evaluation of lymph nodes by the raters (irregular margin, κ = 0.919; inhomogeneous cortex, κ = 0.879; perifocal edema, κ = 0.776; absence of fatty hilum, κ = 0.865; contrast medium enhancement, κ = 0.947; absent intact nodal border, 0.957; all $P < 0.001$), but together these led to an equal evaluation of lymph node status, so that the interrater reliability with regard to lymph node status was excellent (κ = 1.0, $P < 0.001$).
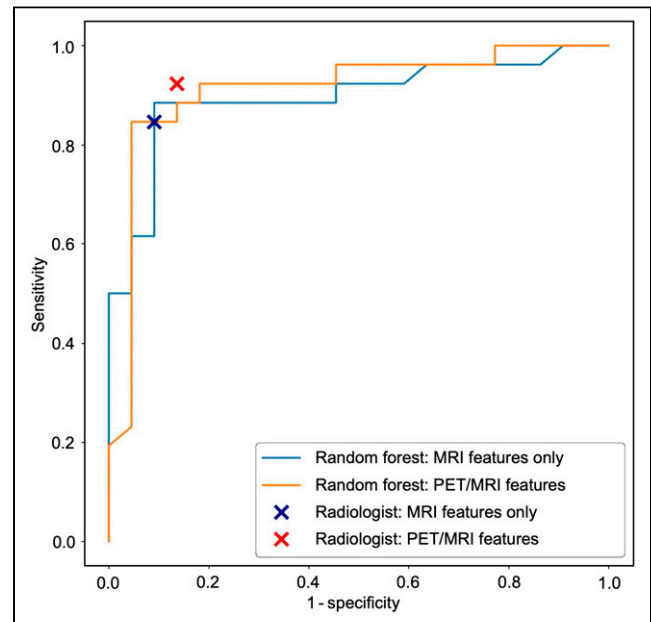


**FIGURE 3.** ROC AUC for random forest model performance on testing data and for prediction of lymph node status by radiologists on MRI and PET/MRI. LN = lymph node.
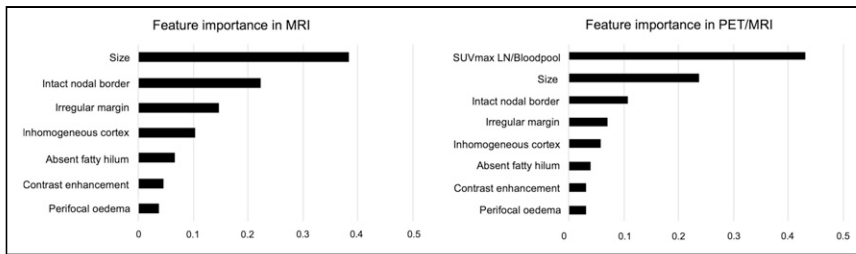
**FIGURE 4.** Importance of different morphologic and metabolic features of lymph nodes.

## Random Forest Algorithm Performance

The trained random forest classifiers yielded an accuracy of 88.3% for MRI and of 99.2% for PET/MRI on the training data, which is indicative of a very good model fit to the training data (Supplemental Table 1). When applied to the independent datasets of the testing sample, the respective random forest classifier was able to determine the correct lymph node status in 42 of 48 participants (87.5%) (23 true-positive and 19 true-negative) for MRI features, whereas 3 participants were rated false-positive and 3 participants false-negative (both readers, Table 2). The performance was unchanged when applying the PET/MRI-based random forest classifier to the testing sample, with 42 of 48 correct classifications (87.5%) (23 true-positive and 19 true-negative), whereas 3 participants were rated false-positive and 3 participants false-negative on the basis of the lymph node assessment of reader 1. On the basis of the lymph node assessment of reader 2, there were 41 of 48 correct classifications (85.4%) (23 true-positive and 18 true-negative), whereas 4 participants were rated false-positive and 3 participants false-negative. Sensitivity, specificity, PPV, and NPV for both classifiers on PET/MRI were 88.5%, 86.4%, 88.5%, and 86.4%, respectively, for reader 1 and 88.5%, 81.8%, 85.2%, 85.7%, 85.4%, respectively, for reader 2 (Table 2).
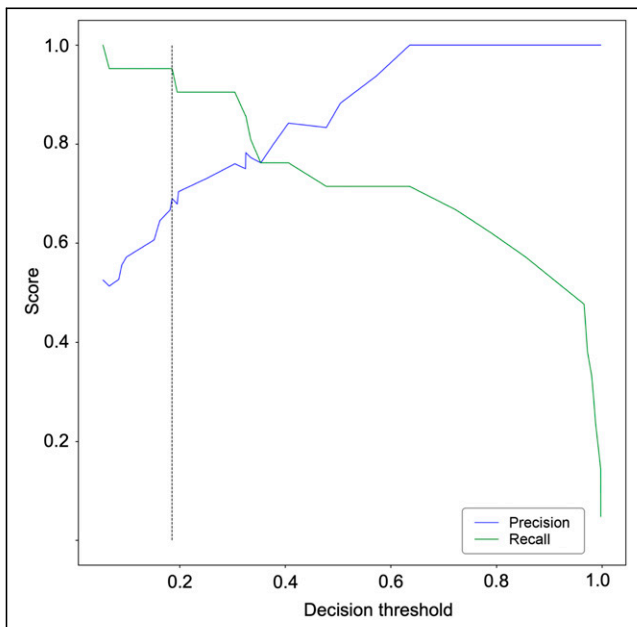


**FIGURE 5.** Precision and recall scores as function of decision threshold on internal validation sample. *x* represents threshold values, and *y* is score of precision or recall. Adjusted decision threshold for optimized sensitivity is indicated by dashed line.

## Comparison of Radiologist Performance and Random Forest Algorithm

In the testing sample, the highest ROC AUC was achieved by the random forest classifier based on PET/MRI data, with a value of 91.2% (95% CI, 82.8%–99.6%), followed by an ROC AUC of 89.5% (95% CI, 80.4%–98.7%) by the random forest classifier based on MRI data (Fig. 3).

There were no significant differences in the assessment of lymph node status between the radiologists and the random forest classifier, either for MRI features ($P = 0.67$) or for PET/MRI features ($P = 0.68$).

## Feature Importance

The most important feature in MRI was size, followed by intact nodal border and irregular margin, whereas the most important features for predicting the nodal status in PET/MRI were tracer uptake as indicated by the ratio of the $SUV_{max}$ of the lymph node to the $SUV_{max}$ of the ascending aorta, followed by size and intact nodal border (Fig. 4).

## Decision Threshold Adjustment

To minimize the classifier's false-negatives with regard to clinical need, we adjusted the decision threshold of the random forest classifier on PET/MRI data as a trade-off between precision (i.e., PPV) and recall (i.e., sensitivity). The default decision threshold in the random forest classifier was 0.5. Figure 5 shows precision and recall as a function of decision values in the internal validation sample. The optimal decision threshold for this purpose was obtained at 0.19. A sensitivity (recall) of 96.2% was achieved, with only 1 false-negative in the test sample, whereas specificity, PPV, NPV, and accuracy were 68.2%, 78.1%, 93.8%, and 83.3%, respectively, at this threshold. Applying these results to everyday routines in our cohort would mean that it would be possible to save 68.2% (15/22) of the women from an unnecessary biopsy, although 3.8% (1/26) of the affected women would be missed (Tables 3 and 4).

## Decision Tree for Clinical Decision Support

The decision tree classifier for distinguishing benign from malignant lymph nodes achieved an accuracy of 89.6% and an ROC AUC of 87.6% (95% CI, 77.6%–97.5%) for MRI in the testing sample and an accuracy of 89.6% and ROC AUC of 89.0% (95% CI, 79.7%–98.4%) for PET/MRI data in the testing sample.

These decision trees can support clinical decision making based on 3 simple imaging features each (Fig. 6A). For MRI, the root node indicative of the most important feature is size, which is consistent with the feature importance from random forests. Here, a short-axis diameter of at least 7.5 mm serves as a cutoff for highly suggestive lymph nodes. ROC AUC evaluation of this feature alone shows a sensitivity of 71.6% and specificity of 86.4% ($J = 0.580$)

**TABLE 3**
Confusion Matrix for Adjusted Threshold

| Actual | Predicted | |
|---|---|---|
| | Negative | Positive |
| Negative | 15 | 7 |
| Positive | 1 | 25 |

## TABLE 4
### Performance Metrics for Adjusted Threshold

| Metric | Data |
|---|---|
| Sensitivity | 96.2% (80.4%–99.9%) |
| Specificity | 68.2% (45.1%–86.1%) |
| PPV | 78.1% (65.9%–86.9%) |
| NPV | 93.8% (68.2%–99.1%) |
| Accuracy | 83.3% (69.8%–92.5%) |

Data in parentheses are ranges.

## TABLE 5
### Confusion Matrices for Decision Trees

| Actual | Predicted | | | |
|---|---|---|---|---|
| | MRI | | PET/MRI | |
| | Negative | Positive | Negative | Positive |
| Negative | 20 | 2 | 21 | 1 |
| Positive | 3 | 23 | 4 | 22 |

for this cutoff. A cutoff of 12.5 mm led to a specificity of 100% but a sensitivity of 34.3% (J = 0.343) (Fig. 6B). The decision tree and these cutoffs were determined from the training data. The combination of an $^{18}$F-FDG uptake more than 1.3-fold that in the aorta ascendens and a short-axis diameter of 7.5 mm is sufficient to characterize a lymph node as malignant.

The confusion matrices and performance metrics for the decision trees are shown in Tables 5 and 6. The performance of the decision trees on the training data is shown in Supplemental Table 2. Supplemental Table 3 shows the detection rates for lymph nodes on $^{18}$F-FDG PET/MRI per nodal stage (cN0–cN3c).

## DISCUSSION

Our study demonstrated that lymph node metastases in patients with newly diagnosed breast cancer can be diagnosed using simple imaging features from MRI and PET/MRI, both by radiologists and by machine-learning–based prediction models, with comparably high accuracies. However, our results indicate that a machine-learning–based prediction model can be advantageous in a clinical setting because it provides the opportunity for decision threshold adjustments. Compared with the current gold standard, in which every clinically node-negative patient would undergo sentinel lymph node biopsy, use of the random forest classifier on PET/MRI data would make it possible to prevent unnecessary biopsy in 68.2% of the women, although 3.8% of the women would be missed. This ability is important for such a model to be suitable for the clinical setting, in which diagnostic imaging might potentially omit invasive procedures such as lymph node biopsy when false-negatives can reliably be reduced. Furthermore, we derived a decision tree for clinical decision support based on simple imaging features from MRI and PET/MRI, which can assist clinicians in the diagnostic workup with regard to lymph node involvement in breast cancer. Although application of the model evaluated here does not, per se, save time in the evaluation of lymph node criteria, the clear cascade of the 3 easily assessable imaging features can be helpful for the radiologist when classifying axillary lymph nodes in daily routine.

Different machine-learning algorithms for the detection of axillary lymph node metastases have previously been shown to provide diagnostic performance comparable to or better than that of experienced physicians in other specialties (17), but only a few applications have been introduced into everyday routine.

This study further rated the relevance of various imaging features of lymph nodes. Although the size of a lymph node, as characterized by the short-axis diameter, is a generally accepted criterion for assessing metastatic status (8), diagnostic accuracy can be increased by adding factors such as contour and signal intensity. Nevertheless, the feature importance of the random forest classifier and the good performance of the simple decision tree classifier indicate that only a few features are necessary to predict lymph node malignancy with high accuracy. Our findings are in line with those of Ramírez-Galván et al. (18), who found lymph node size to be the most important morphologic feature. However, according to our investigation, a short-axis diameter of at least 7.5 mm seems to be most
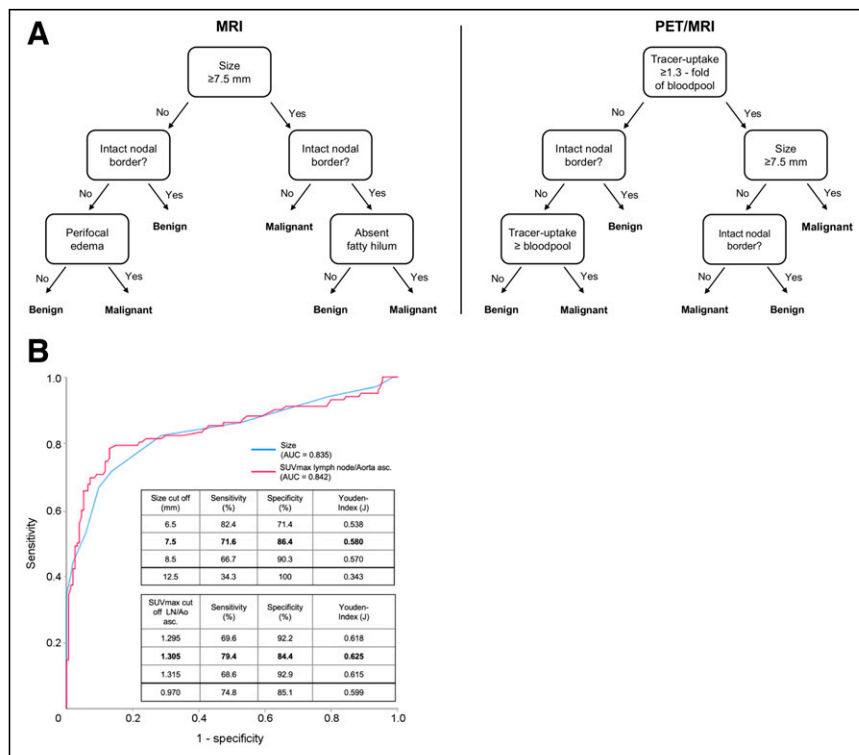


**FIGURE 6.** (A) Decision tree for predicting lymph node status in MRI and PET/MRI. (B) ROC AUC for size and for $SUV_{max}$ ratio of lymph node to mediastinal blood pool for prediction of lymph node status. Ao = aorta; LN = lymph node.

## TABLE 6
### Performance Metrics for Decision Trees

| Metric | Data |
|---|---|
| **MRI** | |
| Sensitivity | 88.5% (69.9%–97.6%) |
| Specificity | 90.9% (70.8%–98.9%) |
| PPV | 92.0% (75.3%–97.8%) |
| NPV | 87.0% (69.5%–95.1%) |
| Accuracy | 89.6% (77.3%–96.5%) |
| **PET/MRI** | |
| Sensitivity | 84.6% (65.1%–95.6%) |
| Specificity | 95.5% (77.2%–99.9%) |
| PPV | 95.7% (76.3%–99.3%) |
| NPV | 84.0% (68.0–92.9. %) |
| Accuracy | 89.6% (77.3%–96.5%) |

Data in parentheses are ranges.

suitable for prediction of axillary lymph node involvement by breast cancer, whereas a diameter of at least 12.5 mm can be seen even as evidence of malignancy (Fig. 6B).

As with other cancer entities, there is no consensus about uptake thresholds in breast cancer to define a lymph node as benign or malignant (19), but an $SUV_{max}$ threshold of 1.8–2.0 has reported to be a helpful criterion to diagnose malignancy (20,21). Our study demonstrated that uptake in the lymph node below that in the mediastinal blood pool is a reliable feature of benignity, whereas uptake at least 1.3 times that in the mediastinal blood pool should be considered malignant.

Using the adjusted threshold of the random forest classifier, the rate of false-negatives might be substantially decreased to a range that would be acceptable for clinical purposes. The single participant missed by our machine-learning algorithm after adjustment of the threshold had a histopathologically proven micrometastasis (1 mm). The clinical impact of micrometastases does not appear to be comparable to that of macrometastases, with micrometastasis outcome being comparable to that of node-negative patients (22). Thus, machine-learning algorithms may be expected to play a crucial role in reducing invasive procedures in the future.

This study had some limitations. Because only therapy-naïve patients were examined at baseline staging, no general statements can be made on regressively altered lymph nodes after therapy or on response to therapy. The reference standard was in part based on posttherapeutic specimens from axillary nodes and different methods of sample acquisition, including axillary dissection and ultrasound-guided biopsy. These differences may have had an impact on definition of the reference standard. The imaging features used as input for the machine-learning–based prediction models still rely on subjective assessments of radiologists. Nevertheless, we could show that these imaging features are easy assessable and have a high interrater reliability. In addition, the size of the validation cohort was only moderate; further studies with a larger population are needed.

## CONCLUSION

This study showed, first, that a random forest classifier based on simple imaging features provides diagnostic performance comparable to that of an experienced radiologist; second, that $^{18}$F-FDG PET uptake and lymph node size assessed on MRI are the most informative features in determining the metastatic status of an axillary lymph node; third, that a combination of 3 features can be helpful for differentiating between malignant and benign axillary lymph nodes in newly diagnosed breast cancer in daily routine; and fourth, that—accepting a low specificity—a sensitivity of more than 95% can be achieved with an adjusted random forest classifier on $^{18}$F-FDG PET/MRI data, which can exclude lymph node involvement with high confidence and might play a central role in reducing invasive procedures in the future. Thus, the combination of the 3 imaging features, in particular, may be applied for daily use by the radiologist, as these can be determined and evaluated quickly and reliably, although the decision tree should not be the only basis for therapy planning. For therapy decision making, the adjusted random forest model is more reliable for differentiation between malignant and benign lymph nodes because of its higher sensitivity. Nevertheless, the adjusted random forest model needs to be confirmed in large, prospective studies to minimize the number of unnecessary invasive procedures and, if successful, will then have great impact.

## DISCLOSURE

### KEY POINTS

**QUESTION:** Can machine-learning prediction models perform comparably to experienced radiologists in determining nodal status on PET/MRI examinations of patients with newly diagnosed breast cancer?

**PERTINENT FINDINGS:** Machine learning performed comparably to experienced radiologists in identifying axillary lymph node metastases on PET/MRI in patients with primary breast cancer. The most important lymph node feature was tracer uptake, followed by lymph node size. A combination of 3 features was helpful for differentiation between malignant and benign axillary lymph nodes in newly diagnosed breast cancer, leading to an easily applicable decision tree in everyday clinical routine.

**IMPLICATIONS FOR PATIENT CARE:** With the help of machine learning, axillary lymph node metastases can reliably be excluded on PET/MRI, sparing 68.2% of the patients an invasive procedure such as sentinel lymph node biopsy.

## REFERENCES

1. Breast cancer. World Health Organization website. https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Published March 26, 2016. Accessed November 30, 2022.
2. Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary nodal evaluation in breast cancer: state of the art. *Radiology.* 2020;295:500–515.
3. Giuliano AE, Ballman KV, McCall L, et al. Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (Alliance) randomized clinical trial. *JAMA.* 2017;318:918–926.
4. Giuliano AE, Hunt KK, Ballman KV, et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA.* 2011;305:569–575.

5. Duraes M, Guillot E, Seror J, Pouget N, Rouzier R. Sentinel lymph node biopsy and neoadjuvant treatment in breast cancer [in French]. *Bull Cancer.* 2017;104:892–901.

6. Yan M, Abdi MA, Falkson C. Axillary management in breast cancer patients: a comprehensive review of the key trials. *Clin Breast Cancer.* 2018;18:e1251–e1259.

7. Morawitz J, Bruckmann N-M, Dietzel F, et al. Determining the axillary nodal status with four current imaging modalities including ^{18}F-FDG PET/MRI in newly diagnosed breast cancer: a comparative study using histopathology as reference standard. *J Nucl Med.* 2021;62:1677–1683.

8. Elsholtz FHJ, Asbach P, Haas M, et al. Introducing the node reporting and data system 1.0 (Node-RADS): a concept for standardized assessment of lymph nodes in cancer. *Eur Radiol.* 2021;31:6116–6124.

9. Zhao M, Wu Q, Guo L, Zhou L, Fu K. Magnetic resonance imaging features for predicting axillary lymph node metastasis in patients with breast cancer. *Eur J Radiol.* 2020;129:109093.

10. Atallah D, Moubarak M, Arab W, Kassis NE, Chahine G, Salem C. MRI-based predictive factors of axillary lymph node status in breast cancer. *Breast J.* 2020;26:2177–2182.

11. Ehteshami Bejnordi BE, Veta M, van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* 2017;318:2199–2210.

12. Bruckmann NM, Kirchner J, Umutlu L, et al. Prospective comparison of the diagnostic accuracy of ^{18}F-FDG PET/MRI, MRI, CT, and bone scintigraphy for the detection of bone metastases in the initial staging of primary breast cancer patients. *Eur Radiol.* 2021;31:8714–8724.

13. Morawitz J, Kirchner J, Martin O, et al. Prospective correlation of prognostic immunohistochemical markers with SUV and ADC derived from dedicated hybrid breast ^{18}F-FDG PET/MRI in women with newly diagnosed breast cancer. *Clin Nucl Med.* 2021;46:201–205.

14. Kirchner J, Grueneisen J, Martin O, et al. Local and whole-body staging in patients with primary breast cancer: a comparison of one-step to two-step staging utilizing ^{18}F-FDG-PET/MRI. *Eur J Nucl Med Mol Imaging.* 2018;45:2328–2337.

15. Newman LA, Pernick NL, Adsay V, et al. Histopathologic evidence of tumor regression in the axillary lymph nodes of patients treated with preoperative chemotherapy correlates with breast cancer outcome. *Ann Surg Oncol.* 2003;10:734–739.

16. Takahashi Y, Soh J, Shien K, et al. Fibrosis or necrosis in resected lymph node indicate metastasis before chemoradiotherapy in lung cancer patients. *Anticancer Res.* 2020;40:4419–4423.

17. Golden JA. Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. *JAMA.* 2017;318:2184–2186.

18. Ramírez-Galván YA, Cardona-Huerta S, Elizondo-Riojas G, Álvarez-Villalobos NA, Campos-Coy MA, Ferrara-Chapa CM. Does axillary lymph node size predict better metastatic involvement than apparent diffusion coefficient (ADC) value in women with newly diagnosed breast cancer? *Acta Radiol.* 2020;61:1494–1504.

19. Yu C, Xia X, Qin C, Sun X, Zhang Y, Lan X. Is SUVmax helpful in the differential diagnosis of enlarged mediastinal lymph nodes? A pilot study. *Contrast Media Mol Imaging.* 2018;2018:3417190.

20. Rosen EL, Eubank WB, Mankoff DA. FDG PET, PET/CT, and breast cancer imaging. *Radiographics.* 2007;27(suppl 1):S215–S229.

21. Carkaci S, Adrada BE, Rohren E, et al. Semiquantitative analysis of maximum standardized uptake values of regional lymph nodes in inflammatory breast cancer: is there a reliable threshold for differentiating benign from malignant? *Acad Radiol.* 2012;19:535–541.

22. Wada N, Imoto S. Clinical evidence of breast cancer micrometastasis in the era of sentinel node biopsy. *Int J Clin Oncol.* 2008;13:24–32.