

Article

Machine Learning-Based Prediction of Distant Recurrence in Invasive Breast Carcinoma Using Clinicopathological Data: A Cross-Institutional Study

Shrey S. Sukhadia ^{1,2,*}, Kristen E. Muller ² , Adrienne A. Workman ² and Shivashankar H. Nagaraj ^{1,*} 

¹ Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, QLD 4059, Australia

² Department of Pathology and Laboratory Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH 03766, USA; kristen.e.muller@hitchcock.org (K.E.M.); adrienne.a.workman@hitchcock.org (A.A.W.)

* Correspondence: shreysanjaybhai.sukhadia@hdr.qut.edu.au or shreys.sukhadia@hitchcock.org (S.S.S.); shiv.nagaraj@qut.edu (S.H.N.)

Simple Summary: Breast cancer is a diverse disease with varying prognoses, even within the same subtype. Approximately 30% of breast cancer patients experience distant organ recurrence, known as metastasis, after treatment. The evaluation of breast tumors and surrounding lymph nodes occurs before and after neoadjuvant therapy, which aims to shrink the tumor before surgery. Following resection, residual tumor cells may remain in the breast tissue, lymph nodes, or other areas, necessitating adjuvant therapy. Typically, a follow-up visit is scheduled a year or more after adjuvant therapy, during which metastasis may be detected. By utilizing machine learning techniques, metastasis can be predicted earlier in a clinical setting, allowing for tailored surveillance and treatment strategies. This has the potential to significantly enhance the quality of life for breast cancer patients.



Citation: Sukhadia, S.S.; Muller, K.E.; Workman, A.A.; Nagaraj, S.H. Machine Learning-Based Prediction of Distant Recurrence in Invasive Breast Carcinoma Using Clinicopathological Data: A Cross-Institutional Study. *Cancers* **2023**, *15*, 3960. <https://doi.org/10.3390/cancers15153960>

Academic Editor: Andreas Stadlbauer

Received: 9 June 2023

Revised: 19 July 2023

Accepted: 19 July 2023

Published: 3 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Breast cancer is the most common type of cancer worldwide. Alarmingly, approximately 30% of breast cancer cases result in disease recurrence at distant organs after treatment. Distant recurrence is more common in some subtypes such as invasive breast carcinoma (IBC). While clinicians have utilized several clinicopathological measurements to predict distant recurrences in IBC, no studies have predicted distant recurrences by combining clinicopathological evaluations of IBC tumors pre- and post-therapy with machine learning (ML) models. The goal of our study was to determine whether classification-based ML techniques could predict distant recurrences in IBC patients using key clinicopathological measurements, including pathological staging of the tumor and surrounding lymph nodes assessed both pre- and post-neoadjuvant therapy, response to therapy via standard-of-care imaging, and binary status of adjuvant therapy administered to patients. We trained and tested four clinicopathological ML models using a dataset (144 and 17 patients for training and testing, respectively) from Duke University and validated the best-performing model using an external dataset (8 patients) from Dartmouth Hitchcock Medical Center. The random forest model performed better than the C-support vector classifier, multilayer perceptron, and logistic regression models, yielding AUC values of 1.0 in the testing set and 0.75 in the validation set ($p < 0.002$) across both institutions, thereby demonstrating the cross-institutional portability and validity of ML models in the field of clinical research in cancer. The top-ranking clinicopathological measurement impacting the prediction of distant recurrences in IBC were identified to be tumor response to neoadjuvant therapy as evaluated via SOC imaging and pathology, which included tumor as well as node staging.

Keywords: radiogenomics; invasive breast carcinoma; machine learning; neoadjuvant therapy

1. Introduction

Breast cancer—which has surpassed lung cancer as the most common type of cancer worldwide—accounts for nearly one-third of cancers in women [1,2]. Despite considerable advances in detection and treatment, nearly 30% of breast cancer patients will develop distant recurrences (i.e., metastasis), often years after diagnosis and treatment of the primary tumor [3]. Although clinicopathological features and molecular biomarkers have been used previously to guide therapeutic decisions in breast cancer clinical workflows [4], forecast outcomes [5–7], and prediction of distant metastases [8–10], long-term survival rates and prognoses vary widely, even within the same histologic and molecular subtype [11,12]. Thus, the high degree of heterogeneity both within and between tumors, combined with myriad other factors that affect the evolution of breast cancer within each individual, presents substantial challenges in treatment and clinical management [13].

Neoadjuvant therapy (i.e., any therapy that precedes surgery) has been used widely in breast cancer to improve outcomes by downstaging inoperable primary tumors to facilitate tumor resection [14,15]. Tumor response to neoadjuvant therapy is evaluated using standard-of-care (SOC) radiological imaging techniques such as MRI, PET-CT, or ultrasound [16,17] and/or biopsies of tumor tissue or surrounding lymph nodes [18], and can be a valuable predictor of survival after therapy [15]. Notably, a complete pathologic response with no residual tumor is indicative of an excellent prognosis in patients with HER2-positive and triple-negative tumors [19,20]. After surgical removal of the primary tumor, adjuvant therapy can be administered to eliminate any residual tumor in breast tissues or possibly in any other part of the body [21,22].

Distant recurrences post adjuvant therapy are common in patients with certain subtypes of breast cancer, such as invasive breast carcinoma (IBC) [23]. Breast cancer most frequently metastasizes to the bone, lung, liver, brain, and distant lymph nodes, of which bone is the most prominent metastatic site [23,24]. As metastatic disease—the primary reason for breast cancer-related deaths—is known to have poor outcomes and is generally considered to be incurable, predicting distant recurrences is a crucial aspect of personalized monitoring and treatment strategies [3]. While breast cancer subtype can be predictive of the preferred site of distant metastasis [9], accurately predicting the risk of metastasis is more complex. Although OncotypeDX (ODX)—a 21-gene molecular screening assay—is widely used in clinical practice to predict the distant recurrence and benefits from adjuvant chemotherapy for patients with breast cancer, its utility is limited as the ODX recurrence score is not consistent across all ages and relevant only in women with estrogen receptor (ER)-positive breast cancer, which limits its scope [25–27]. The benefits of adjuvant chemotherapy vary depending on the range of the ODX recurrence score when combined with patient age, for which the ideal conditions are women ≤ 50 years with an ODX recurrence score range of 16–25 [28], therefore limiting the patient group for ODX-based tumor recurrence evaluation to younger to mid-aged women only.

Machine learning (ML) has become an increasingly popular tool for making objective clinical predictions and has been used to predict a wide variety of clinical outcomes using biological or clinical input data [29–35]. Metastasis-risk calculators (i.e., “nomograms”), which are based on logistic regression models, have been known to predict synchronous metastases in breast cancer based on the clinicopathological profile (i.e., tumor size, nodal status, and estrogen and progesterone receptor status); however, this methodology does not involve the usage of adjuvant (i.e., second-line) therapy usually given to patients that may not have a complete response to neoadjuvant therapy (Boutros et al. 2015) [36]. A mechanistic model developed by Nicolò et al. [37] predicts time to metastases in patients with early-stage breast cancer using several clinicopathological diagnoses, but it neither considers patients that received adjuvant therapy nor conducts validation of the model using an external cohort. Another study used various clinicopathological features, including serum HER2 levels, to predict metastases; however, low AUC values in the testing set (i.e., AUC < 0.80) in conjunction with lack of validation of the model using an independent (or external) validation set highlights the low clinical utility of the model [38].

Another factor that affects the clinical utility of predictive ML models in breast cancer is the availability and accessibility of code. For example, the aforementioned studies do not share the complete code (i.e., a package of programming scripts) used to train and test the respective ML models, thereby placing the burden on other researchers to unpack the black box of ML programming and/or attempt to translate the methodology into code that can be applied to their own research. Moreover, these studies do not provide the final (i.e., tested) model pickle files (i.e., model.pkl) that a researcher could potentially use to validate their model using their own (i.e., external) dataset. ML models do not contain sensitive personal health information; rather, they contain mathematical equations that are produced from the training conducted on deidentified patient data. These equations consist of correlation weights or the feature importance scores of features used to train the model, both of which are an indication of each feature's individual contribution towards a collective prediction of a particular outcome [39]. These weights or importance scores must be tested using the test dataset, independently of the training set, to arrive at a model worthy of validation.

Further, sharing ML models across institutions could enable meta-learning, potentially yielding superiorly performing models [40]. Our recently released version (v3.2) of the ML software program ImaGene [41] exposes its codebase and automates ML operations for an expanded list of classification-based models such as random forest (RF), C-support vector classifier (SVC), multilayer perceptron (MLP), and logistic regression (LogitR) (i.e., in addition to the previously existing regression-based models). ImaGene also offers various customizable parameters to conduct multiple simultaneous ML experiments and provides result metrics such as R-square, RMSE:Stdev ratio, and AUC, the majority of which are plotted as graphs and included in resulting reports. These reports, along with the subordinate tabular output files, aid users in tracking feature-level performances for the respective models.

Early predictions of the risk of metastases in breast cancer patients post adjuvant-therapy and pre-follow-up (for follow-ups of greater than a year) in clinics using an ML model could allow early clinical intervention and guide surveillance and treatment, which could ultimately improve the quality of life for the respective patients. Furthermore, increasing the availability and accessibility of such an ML model across various hospital sites could facilitate the model's validation and make it clinically viable sitewide. Thus, the objective of our study was twofold: (i) to create a novel, high-fidelity ML model to predict the risk of metastasis in IBC patients using clinicopathologic measurements pre- and post-neoadjuvant therapy, such as clinical tumor [T] and node [N] staging pre-neoadjuvant therapy, response evaluation (via SOC imaging, and T and N staging) post-neoadjuvant therapy, and administration status ('Yes' or 'No') and type of adjuvant therapy (i.e., chemotherapy or Anti-Neu HER2) obtained from one hospital site (i.e., Duke University Hospital [DUH]), and (ii) to validate that model externally (i.e., Dartmouth-Hitchcock Medical Center [DHMC]). To accomplish this, we used ImaGene to train and test four classification-based ML/AI models (i.e., RF, SVC, MLP, and LogitR) using the DUH cohort and validate the best-performing model using the DHMC cohort. To the best of our knowledge, this is the first study to test and validate ML models for the prediction of distant recurrences using clinicopathological profiles across various therapies in IBC patient cohorts from two different medical institutions.

2. Materials and Methods

The present study used clinicopathological data from a retrospective study of 900 IBC patients at DUH, available via the TCIA portal [42,43]. Of these, 312 patients received neoadjuvant therapies such as chemotherapy, radiation, endocrine hormone-based, or anti-neu/HER2 therapy. Responses to therapy were evaluated using imaging and pathological (i.e., T and N) staging. Additionally, 304 of these patients received adjuvant therapies (i.e., chemo-, radiation-, hormone-, or anti-Neu/HER2-based therapies) depending on their responses to neoadjuvant therapy. As our focus was on patients who received neoadjuvant chemotherapy and had either a partial or a complete response, patient entries

pertaining to ungraded tumor responses to therapies or those labeled as ‘Not applicable (NA)’ for any of the clinicopathological features were excluded from further analyses, yielding 161 patient entries.

Specifically, the following eight clinicopathological features (Figure 1A) were considered: (a) tumor stage or size (T0–T4, i.e., in the numerical format: [0,4]) pre-neoadjuvant therapy, (b) staging of lymph nodes (N0–N3, i.e., [0,3]) pre-neoadjuvant therapy, (c) clinical response to neoadjuvant therapy as evaluated using radiological images (i.e., complete response (i.e., ‘1’), incomplete response (i.e., ‘2’), or assessment unavailable (i.e., ‘3’), (d) pathologic response to neoadjuvant therapy (i.e., complete: ‘1’, incomplete: ‘2’, or DCIS-only remaining: ‘3’), (e) pathologic response to neoadjuvant therapy as evaluated using pathology stage “T” tumor size (T0–T4, i.e., [0–4]), (f) pathologic response to neoadjuvant therapy evaluated using node status “N” (N0–N3, i.e., [0,3]), (g) status of adjuvant chemotherapy administered (i.e., yes: ‘1’ or no: ‘0’), and (h) status of adjuvant anti-HER2/Neu therapy administered (i.e., yes: ‘1’ or no: ‘0’). Patients with no response to neoadjuvant chemotherapy were excluded from this study, as this patient population has a poor prognosis and a higher risk of recurrence [44]. Model training was conducted using ImaGene software [41].

The 161 patient entries were split further into training and testing sets at a 90:10 ratio, in which 90% of entries were used for training (i.e., $n_{\text{train}} = 144$) and 10% for testing ($n_{\text{test}} = 17$) four classification-based ML models (i.e., RF, SVC, MLP, and LogitR) for binary predictions of distant recurrences (i.e., yes or no). For logistic regression, the train:test ratio was modified to 85:15 to achieve optimal performance. Experiments for the four ML models were performed using ImaGene [41], which yielded operational reports and supporting text files to aid in the interpretation of the results, thereby facilitating the selection of the best-performing model for validation (Figure 1B).

2.1. Model Development and Testing Using ImaGene

The multimodal feature file consisting of the aforementioned clinicopathological features was set as “data”, while a binary column containing a feature file reporting disease recurrence at a distant site (i.e., distant recurrence flag; 0 for “no” and 1 for “yes”) for the respective patients ($n = 161$) was set as “label” for ImaGene. The model type was first set to RF in “Train” mode. Test size was set to “0.1 (i.e., 10% of dataset allocated to test)”, which partitioned the dataset into training ($n_{\text{Train}} = 144$) and testing ($n_{\text{Test}} = 17$) sets. The K-fold cross-validation splitter parameter (i.e., ‘cv’) was set to “2”. Grid search was set to “True” to enable the execution of a grid search through the hyperparameters: $\text{max_depth} = [6, 9, 10, 12, 15, 20]$ and $\text{cv} = 4$. The “data” normalization method was set to “StandScaler”, while the “label” normalization was set to “none” owing to its binary nature. Furthermore, the absolute correlation threshold parameter pre-ML-training was set to “−1.0” to silence the filtering of features based on Pearson’s correlation co-efficient threshold, thereby considering all the clinicopathological features for further training of the RF model. The run took approximately two minutes to yield a report on model performance (Supplementary Materials Report 1, Figure 2A, and Table 1).

Secondly, the model type was set to SVC and run with default parameters: $\text{cv} = 2$, $\text{test_size} = 0.1$, pre-ML correlation-threshold set to “−1.0”, data-normalization to “Stand Scaler”, and label-normalization to “none”. Furthermore, the grid_search was set to “True” to perform a grid search for the SVC model through polynomial degree hyperparameters (i.e., $\text{kernel} = [\text{‘poyl’}]$, $\text{degree} = [3, 4, 5, 6, 7, 8, 9]$ and $\text{cv} = 2$; Supplementary Materials Report 2, Figure 2B, and Table 1).

Thirdly, a supervised neural network, the MLP classifier, was employed with default parameters. Subsequently, a grid search was executed setting test_size to 0.1 and using the following hyperparameters: $\text{Solver} = [\text{‘sgd’}, \text{‘bfgs’}, \text{‘adam’}]$, $\text{alpha} = [0.0001, 0.001, 0.01, 0.02, 0.05]$, $\text{hidden_layer_sizes} = [(9),(9,9),(9,9,9)]$ and $\text{cv} = 6$ (Supplementary Materials Report 3, Figure 2C, and Table 1).

Lastly, the model type was set to LogitR and run with both default model parameters and in grid search mode. To achieve optimal performance, the train:test ratio was modified

to 85:15 followed by the grid search performed on the following hyperparameters: Solver: ['liblinear', 'newton-cg', 'lbfgs', 'sag', 'saga'] and cv = 4 (Supplementary Materials Report 4, Figure 2D, and Table 1).

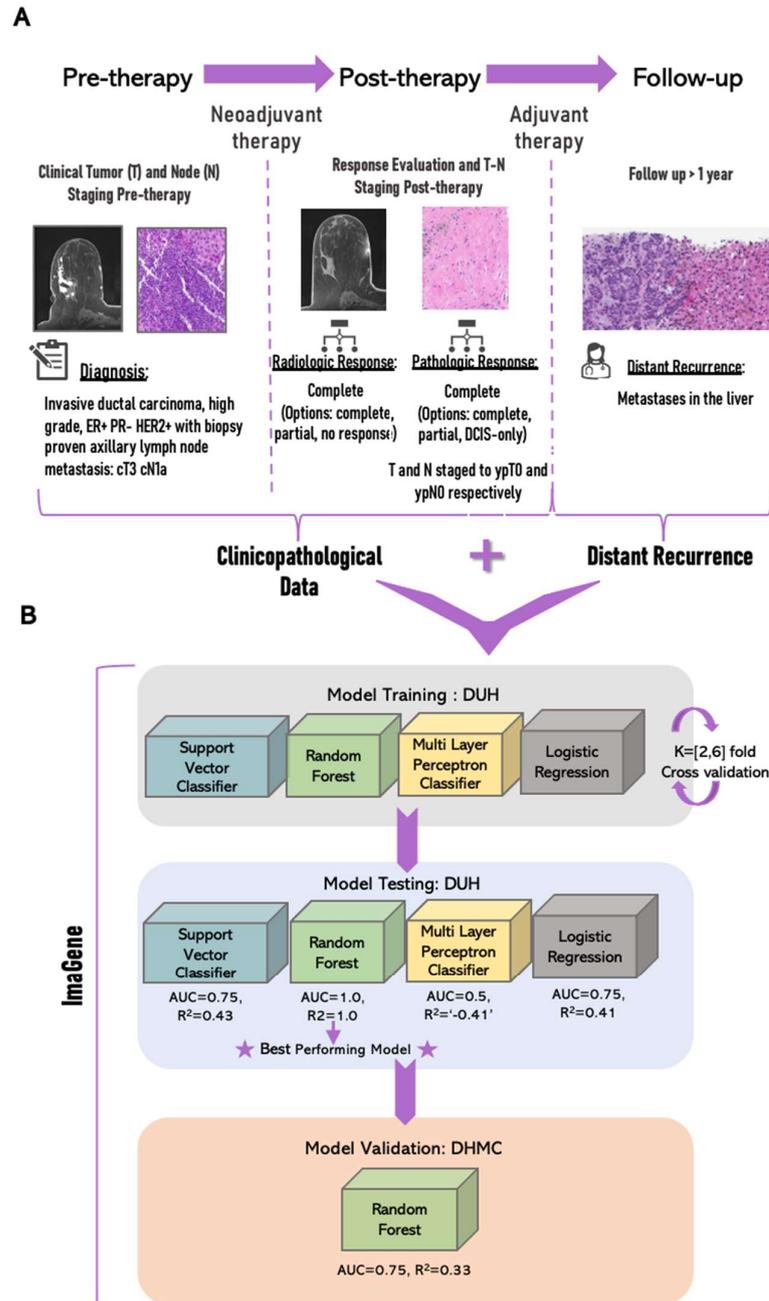


Figure 1. Modeling clinicopathological data to predict distant recurrences in IBC, with clinicopathological workflow for invasive breast cancer (IBC) from initial diagnosis to follow-up represented in (A) followed by training, testing, and validation of machine learning (ML) models for the prediction of distant recurrences recorded post-follow-up in IBC using clinicopathological data recorded pre- and post-therapy and pre-follow-up represented in (B). The area under the receiver operating curve (AUROC/AUC) and R^2 values are mentioned for each model, where a negative R^2 value and AUC = 0.5 indicate model failure. DUH: Duke University Hospital; DHMC: Dartmouth Hitchcock Medical Center.

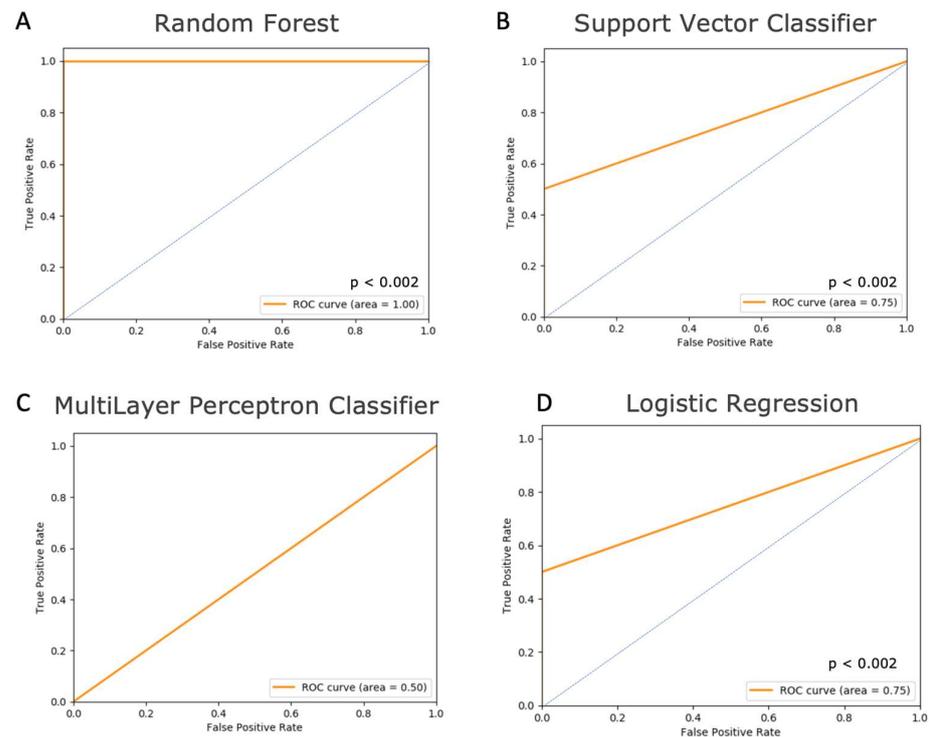


Figure 2. Receiver operating curves (ROC) for the three machine learning models assayed to predict distant metastases in the testing set of breast cancer patients from Duke University Hospital. (A) The random forest model yielded an area under the receiver operating curve (AUROC/AUC) of 1.0 for the testing dataset. (B) The ROC for the Support Vector Classifier yielded an AUC of 0.75 for the testing dataset. (C) The ROC for the multilayer perceptron classifier yielded an AUC of 0.5 for the testing dataset. No *p*-value was calculated as the model failed. (D) The ROC for the logistic regression model yielded an AUC of 0.75 for the testing dataset. Note: The blue line indicates the ROC curve of a random classifier.

Table 1. Results for various classifiers for the test dataset at DUH.

Models	AUC	R ²	RMSE:Stdev
Random Forest	1.0	1.0	0.0
Support Vector Classifier	0.75	0.43	0.75
Multilayer Perceptron Classifier	0.5	−0.41	1.19
Logistic Regression	0.75	0.4	0.77

2.2. Validation Using *ImaGene*

The model that performed best on the testing dataset from DHC was chosen for validation using the external IBC cohort at DHMC (Figure 1B). The validation study included a total of 67 IBC patients that had the same eight clinicopathological features collected from diagnosis to follow-up at DHMC. This validation dataset (i.e., *nvalidate* = 67) was screened for the following two inclusion criteria: (a) Neoadjuvant therapy administered and (b) onsite (i.e., DHMC-based) follow-up 24–45 months post adjuvant therapy. Of these 67 patients, only 15 matched the inclusion criteria. Out of these 15, only three patients exhibited distant recurrences, leading us to select eight patients at random (i.e., two with and six without distant recurrences) to balance the dataset and arrive at a more realistic validation AUC. The validation report from *ImaGene* showcased the performance of the validation dataset through the model (Supplementary Materials Report 5, Figure 3, and Table 2).

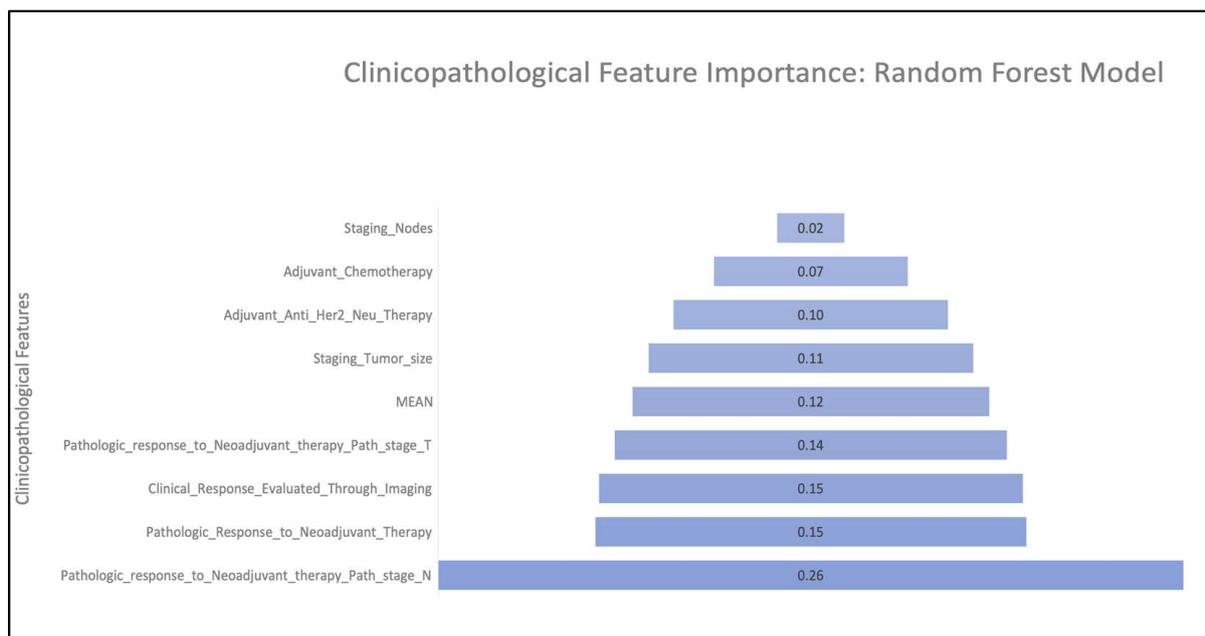


Figure 3. Feature importance scores of clinicopathological features from the random forest model used to predict distant recurrences in invasive breast carcinoma.

Table 2. Validation of the random forest model at external site: DHMC.

Model	AUC	R ²	RMSE:Stdev
Random Forest	0.75	0.33	0.82

3. Results

3.1. Model Performance

The training (nTrain = 144: DUH) and testing (nTest = 17: DUH) of the models (i.e., RF, SVC, MLP, and LogitR) using ImaGene yielded a detailed performance report in html format, along with the supporting tables in csv format to facilitate the evaluation of these models (Supplementary Materials Reports 1–4, Figure 2, and Table 1). These reports and the supporting tabular output files showcased several metrics and plots collectively, which included cross-validation (CV) score, grid search CV score (when grid search is set to “True”), actual-vs-predicted-values scatter plots, and area under the receiver operating curve (AUROC or AUC), and the respective *p*-value for predicting a distant recurrence flag. Additionally, the Mean Square Error (MSE) and R-square (R2) from the model performance on the test dataset were also reported. The MSE, R2, and AUC metrics from all models facilitated the comparison of their performance through the test dataset.

Training the RF model yielded a best grid search CV score of 0.8 (Supplementary Materials Report 1), indicating the best score obtained from the 4-fold cross-validation performed during the tuning of the model through the aforementioned hyperparameters. Testing the RF model yielded MSE, R2, and AUC values of 0.0, 1.0, and 1.0 respectively, indicating a perfect prediction by the RF model (Supplementary Materials Report 1, Figure 2A, and Table 1). ImaGene ran permutations (i.e., shuffling) of labels (i.e., distant recurrence binary values) across test samples ($n = 17$) to infer the statistical significance of the prediction of the R2 and AUC reported by the model, indicating a strong prediction (i.e., $p < 0.002$).

Training the SVC model yielded a best grid search CV score of 0.78 (Supplementary Materials Report 2). Testing the SVC model yielded MSE, R2, and AUC values of 0.06, 0.43, and 0.75 respectively, indicating that the model performed considerably well (Supplementary Materials Report 2, Figure 2B, and Table 1). The performance of the SVC model

was also deemed to be significant ($p < 0.002$); however, this model did not perform as well as the RF model, indicating the superiority of the ensemble of decision tree algorithms (i.e., RF) compared to support vectors for the classification of distant recurrence flags using the clinicopathological profiles of the IBC patients in the present study.

Training of the MLP model yielded a best grid search CV score of 0.83 (Supplementary Materials Report 3). However, testing this model yielded a very high MSE (i.e., 0.29), and AUC and R2 values of 0.5 and -0.41 respectively, indicating complete failure of the model (Supplementary Materials Report 3, Figure 2C, and Table 1). Furthermore, this demonstrates that supervised neural networks are not superior to RF and SVC methods for the classification of distant recurrences using clinicopathological profiles of IBC in the present study.

Lastly, training of the LogitR model yielded a best grid search CV core of 0.75 (Supplementary Materials Report 4). The testing of this model yielded MSE, R2, and AUC values of 0.08, 0.41, and 0.75 respectively, indicating that the model performed considerably well (Supplementary Materials Report 4, Figure 2D, and Table 1). The performance of the LogitR model was also deemed to be significant ($p < 0.002$). LogitR showed similar performance to SVC but lower compared to RF, reinstating the superiority of ensemble decision trees (i.e., RF) compared to the logit method (i.e., LogitR) for the prediction of distant recurrence flags using clinicopathological profiles in IBC patients.

The RF model was the best-performing model in the present study (Figure 1B). The feature importance scores of clinicopathological features for predicting distant recurrences using the RF model indicate that the top four features exhibited scores greater than or equal to the mean of all scores (Figure 3). These features are (in order of importance) (1) pathologic response to neoadjuvant therapy evaluated using node status “N” (i.e., N0–N3), (2) pathologic response to neoadjuvant therapy (i.e., complete, incomplete, or DCIS-only remaining), (3) clinical response to neoadjuvant therapy as evaluated using SOC radiological images (i.e., complete response, incomplete response, or assessment unavailable), and (4) pathologic response to neoadjuvant therapy as evaluated using pathology stage “T” tumor size (T0–T4). The RF model was selected further for validation using the external cohort from DHMC (Figure 1B).

3.2. Model Validation

The validation of the RF model was conducted on 8 out of 67 patients matching the inclusion criteria of having received therapy and followed up at DHMC 24–45 months after adjuvant therapy. The validation MSE of the model was found to be 0.125, with an AUC of 0.75 and an R2 of 0.33 (Supplementary Materials Report 5, Figure 4, and Table 2). This indicates that the RF model performed considerably well given the small size of the clinical dataset after screening for inclusion criteria as previously mentioned, highlighting a common issue in clinical datasets for lengthy studies that include clinicopathological results collected from the same site. Moreover, low AUC could also be attributed to interobserver variability in tumor and node evaluation pre- and post-therapy by clinicians and pathologists across two different hospital sites (i.e., DUH and DHMC).

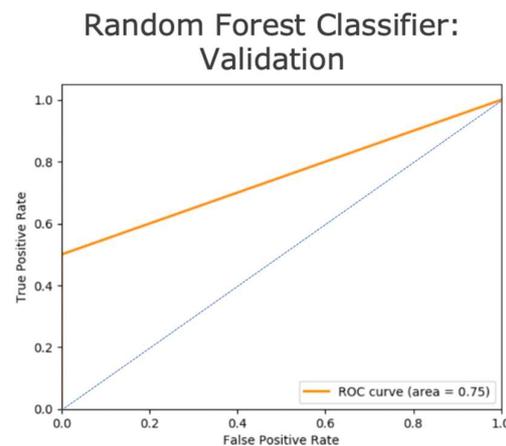


Figure 4. Receiver operating curve (ROC) for random forest model validated on the external Dartmouth Hitchcock Medical Center cohort, depicting an area under the curve (AUC) of 0.75. The AUC represents the validation of random forest model for the prediction of distant recurrence flags from the clinicopathological profiles in patients at the external site. Note: The blue line indicates the ROC curve of a random classifier.

4. Discussion

To the best of our knowledge, our study is the first ML/AI-based study to predict distant recurrences or metastasis in IBC patients using clinicopathological data collected from initial diagnosis through follow-up in the clinic and subsequently validated in an external dataset (i.e., cross-institutional). The random forest (RF) technique outperformed other models using the testing and validation sets across two hospital sites (i.e., DUH and DHMC). The key clinicopathological features impacting the prediction of metastasis are tumor responses to neoadjuvant therapy (NAT) evaluated using SOC imaging and pathology (i.e., Tumor (T) and Node (N) staging after neoadjuvant therapy) (Figure 4). The pathological evaluation of tumors and nodes post-neoadjuvant therapy is the gold standard in evaluating tumor response to neoadjuvant therapy [45] and is a recommendation of the 8th edition of the American Joint Committee of Cancer [46]. The use of SOC imaging is vital for the gross examination of tumors both before and after NAT (Viale and Fusco 2022) [46]. In our study, we found that the contribution of SOC imaging after NAT (i.e., for evaluating the tumor response to NAT) was more impactful than imaging before NAT in the prediction of metastasis (Figure 4). In contrast, we found that the binary status of adjuvant therapy did not make a substantial contribution to the prediction of metastases (Figure 4). This is in agreement with previous studies in an international breast cancer cohort, in which the adjuvant therapy could result in increased metastasis to the liver and central nervous system but decreased metastasis to bones [47–49], which supports the reduced contribution of adjuvant therapy to predict metastasis in our study as well (Figure 4).

Several ML/AI studies have been conducted in oncology [29–35,50,51]. Some of these have attempted to predict 5–10-year breast cancer recurrences using both structured and unstructured clinicopathological data from Electronic Health Records (EHRs) [50,51]; however, these studies focused primarily on predicting local rather than distant recurrences. A previous study [52] predicted distant recurrences in breast cancer from both unstructured and structured clinical data in EHR using natural language processing and deep learning algorithms but lacked structured clinicopathological predictors such as those used in our study. Furthermore, the same study [52] also lacked advanced feature engineering. In contrast, the present study used structured clinicopathological data from EHRs that were further curated by trained pathologists and medical residents at the hospital prior to being fed to the ML model. Moreover, the present study provides us with an advanced understanding of the contribution of each clinicopathological feature towards the prediction of distant recurrence in the form of feature importance scores (which measures the weights of the features used to predict the label, i.e., risk of distant recurrence), which could guide fea-

ture engineering efforts in the future (Figure 3). Additionally, several previous ML studies have predicted disease recurrences in a variety of other cancers, such as in non-metastatic renal cell carcinoma [53] and in early-stage endometrial cancer [54] which only predicted recurrences at an AUC of 0.53 owing to the small size of the dataset. In contrast, the present study achieves AUC values of 1.0 and 0.75 in the intra- and inter-institutional tests respectively, despite the small sample size (i.e., $n_{\text{test}} = 17$ and $n_{\text{validate}} = 8$), highlighting the importance of highly curated data in ML studies.

Although the use of ML in predictive models can increase our ability to predict outcomes in individual cancer patients, validation is vital to the incorporation of ML into clinical workflows (Kourou 2015) [55]. A previous study that predicted distant recurrences in breast cancer used both classification-based and deep learning models but repurposed approximately 10% of their training data for validation [52]. Another study that predicted distant recurrence in breast cancer using clinicopathological and serum HER2 profiles yielded an AUC of 0.8 for the testing set but lacked validation in an external cohort, calling its clinical utility into question [38]. Thus, although the present study was based solely on classification models, it did use a blinded clinical validation set from a completely different hospital site than the sites from which training and test originated, thereby highlighting the strength of the study and predictions.

However, the present study has several limitations. Notably, the number of samples in the testing and validation sets was relatively low. In the near future, we plan to enhance the testing and validation datasets by collating data from multiple institutions to encompass a wider range of heterogenous IBC tumors and institutions. This could be performed using an open-source platform such as ImaGene, which enables the democratization of multi-omic analyses and gives open access to results [41]. Another limitation of the present study is the lack of information regarding the site of distant recurrence, which was not provided by the dataset shared by DUH via the TCIA platform. With this additional information, the present study could be used as a blueprint for predicting the site of distant recurrence as well.

Using the RF model to predict the possibility of distant recurrences in breast cancer patients could provide clinicians with the ability to foresee distant recurrences and tailor treatment and management plans accordingly to improve outcomes. Furthermore, our study extends the capability of ImaGene to utilize several clinicopathological features of each patient's tumor throughout the diagnostic and therapeutic journey. Using ImaGene, a patient's unique pathologic, radiologic, and therapeutic information can be leveraged to predict distant recurrences using various ML/AI models in IBC. ImaGene is an open-access software that shares the code for the automated operation of ML/AI models, supporting the repeatability of their training, testing, and validation with datasets at any institution worldwide, unlike the code used by previous studies that sought to predict distant recurrences using clinicopathological features of cancer [36–38]. Our study also proves the cross-validity of the RF model across two distinct hospital sites. A similar model could potentially be trained, tested, and validated for the prediction of disease progression in other cancer types based on the clinicopathological profiles of the respective tumors in the future. Our study advances the field of non-invasive predictions of cancer metastasis. Future research in this field could aid researchers and clinicians in identifying the risk and sites of disease recurrence, thereby optimizing cancer treatment and ultimately reducing cancer mortality.

5. Conclusions

This study explores ML models for predicting metastasis risk in IBC patients using clinicopathological features of their tumor and lymph nodes measured pre- and post-neoadjuvant therapy, including adjuvant therapy status. Classification-based ML models were trained and tested on one hospital's (DUH) datasets. The best model (RF) was further validated using another hospital's (DHMC) dataset, demonstrating significant AUC and R^2 values for cross-validity in heterogenous IBC tumors sitewide. Tumor response

to neoadjuvant therapy, evaluated through SOC imaging and pathology (including the tumor and node staging), contributed most to metastasis prediction. ML models hold the potential for stratifying patients into high- and low-risk categories for metastasis, enabling the regulation of surveillance and treatments to improve their quality of life.

Supplementary Materials: The following supporting information can be downloaded at: <https://github.com/skr1/Imagene>. Reference [56] are cited in the supplementary materials.

Author Contributions: Conceptualization, S.S.S.; methodology, S.S.S. software, S.S.S.; validation, S.S.S., K.E.M. and A.A.W.; formal analysis, S.S.S.; investigation, S.S.S., K.E.M. and A.A.W.; resources, S.S.S. and S.H.N.; data curation, S.S.S. and K.E.M.; writing—original draft preparation, S.S.S.; writing—review and editing: S.S.S., K.E.M., A.A.W. and S.H.N.; visualization, S.S.S. and S.H.N.; supervision, S.H.N.; project administration, S.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The validation data used in study was from Dartmouth-Hitchcock Medical Center (or Dartmouth Health) and was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Dartmouth Health (protocol code STUDY00031179 and 28 April 2023).

Informed Consent Statement: Patient consent was waived as it was not required per IRB.

Data Availability Statement: Data are available at: <https://github.com/skr1/Imagene> (accessed on 1 January 2023).

Acknowledgments: We thank The Cancer Imaging Archive (TCIA) portal and the Duke University Hospital for making the clinicopathological data available at TCIA.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics. *CA A Cancer J. Clin.* **2022**, *72*, 7–33. [[CrossRef](#)] [[PubMed](#)]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
3. Redig, A.J.; McAllister, S.S. Breast cancer as a systemic disease: A view of metastasis. *J. Intern. Med.* **2013**, *274*, 113–126. [[CrossRef](#)]
4. Van Poznak, C.; Somerfield, M.R.; Bast, R.C.; Cristofanilli, M.; Goetz, M.P.; Gonzalez-Angulo, A.M.; Hicks, D.G.; Hill, E.G.; Liu, M.C.; Lucas, W.; et al. Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **2015**, *33*, 2695–2704. [[CrossRef](#)]
5. Phung, M.T.; Tin, S.T.; Elwood, J.M. Prognostic models for breast cancer: A systematic review. *BMC Cancer* **2019**, *19*, 230. [[CrossRef](#)]
6. Huang, E.; Cheng, S.H.; Dressman, H.; Pittman, J.; Tsou, M.H.; Horng, C.F.; Bild, A.; Iversen, E.S.; Liao, M.; Chen, C.M.; et al. Gene expression predictors of breast cancer outcomes. *Lancet* **2003**, *361*, 1590–1596. [[CrossRef](#)]
7. Engelhardt, E.G.; Garvelink, M.M.; Haes, J.C.J.M.D.; Hoeven, J.J.M.V.D.; Smets, E.M.A.; Pieterse, A.H.; Stiggelbout, A.M. Predicting and Communicating the Risk of Recurrence and Death in Women with Early-Stage Breast Cancer: A Systematic Review of Risk Prediction Models. *J. Clin. Oncol.* **2014**, *32*, 238–250. [[CrossRef](#)]
8. Xiao, W.; Zheng, S.; Yang, A.; Zhang, X.; Zou, Y.; Tang, H.; Xie, X. Breast cancer subtypes and the risk of distant metastasis at initial diagnosis: A population-based study. *Cancer Manag. Res.* **2018**, *10*, 5329–5338. [[CrossRef](#)] [[PubMed](#)]
9. Wu, Q.; Li, J.; Zhu, S.; Wu, J.; Chen, C.; Liu, Q.; Wei, W.; Zhang, Y.; Sun, S. Breast cancer subtypes predict the preferential site of distant metastases: A SEER based study. *Oncotarget* **2017**, *8*, 27990–27996. [[CrossRef](#)]
10. Gupta, G.P.; Massagué, J. Cancer Metastasis: Building a Framework. *Cell* **2006**, *127*, 679–695. [[CrossRef](#)]
11. Howlader, N.; Cronin, K.A.; Kurian, A.W.; Andridge, R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiol. Biomark. Prev.* **2018**, *27*, 619–626. [[CrossRef](#)] [[PubMed](#)]
12. Liao, H.-Y.; Zhang, W.-W.; Sun, J.-Y.; Li, F.-Y.; He, Z.-Y.; Wu, S.-G. The Clinicopathological Features and Survival Outcomes of Different Histological Subtypes in Triple-negative Breast Cancer. *J. Cancer* **2018**, *9*, 296–303. [[CrossRef](#)] [[PubMed](#)]
13. Polyak, K. Heterogeneity in breast cancer. *J. Clin. Investig.* **2011**, *121*, 3786–3788. [[CrossRef](#)] [[PubMed](#)]
14. Selli, C.; Sims, A.H. Neoadjuvant Therapy for Breast Cancer as a Model for Translational Research. *Breast Cancer* **2019**, *13*, 1178223419829072. [[CrossRef](#)]
15. Wang, H.; Mao, X. Evaluation of the Efficacy of Neoadjuvant Chemotherapy for Breast Cancer. *Drug Des. Dev. Ther.* **2020**, *14*, 2423–2433. [[CrossRef](#)]

16. Fowler, A.M.; Mankoff, D.A.; Joe, B.N. Imaging Neoadjuvant Therapy Response in Breast Cancer. *Radiology* **2017**, *285*, 358–375. [[CrossRef](#)]
17. Sadeghi-Naini, A.; Sannachi, L.; Tadayyon, H.; Tran, W.T.; Slodkowska, E.; Trudeau, M.; Gandhi, S.; Pritchard, K.; Kolios, M.C.; Czarnota, G.J. Chemotherapy-Response Monitoring of Breast Cancer Patients Using Quantitative Ultrasound-Based Intra-Tumour Heterogeneities. *Sci. Rep.* **2017**, *7*, 10352. [[CrossRef](#)]
18. Bossuyt, V.; Spring, L. Pathologic evaluation of response to neoadjuvant therapy drives treatment changes and improves long-term outcomes for breast cancer patients. *Breast J.* **2020**, *26*, 1189–1198. [[CrossRef](#)]
19. Provenzano, E.; Bossuyt, V.; Viale, G.; Cameron, D.; Badve, S.; Denkert, C.; MacGrogan, G.; Penault-Llorca, F.; Boughey, J.; Curigliano, G.; et al. Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: Recommendations from an international working group. *Mod. Pathol.* **2015**, *28*, 1185–1201. [[CrossRef](#)]
20. von Minckwitz, G.; Untch, M.; Blohmer, J.U.; Costa, S.D.; Eidtmann, H.; Fasching, P.A.; Gerber, B.; Eiermann, W.; Hilfrich, J.; Huober, J.; et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J. Clin. Oncol.* **2012**, *30*, 1796–1804. [[CrossRef](#)]
21. Anampa, J.; Makower, D.; Sparano, J.A. Progress in adjuvant chemotherapy for breast cancer: An overview. *BMC Med.* **2015**, *13*, 195. [[CrossRef](#)]
22. Flatley, M.J.; Dodwell, D.J. Adjuvant treatment for breast cancer. *Surgery* **2019**, *37*, 176–180. [[CrossRef](#)]
23. Riggio, A.I.; Varley, K.E.; Welm, A.L. The lingering mysteries of metastatic recurrence in breast cancer. *Br. J. Cancer* **2021**, *124*, 13–26. [[CrossRef](#)] [[PubMed](#)]
24. Xin, J.; Ping, M. Targeting breast cancer metastasis. *Breast Cancer Basic. Clin. Res.* **2015**, *9*, 23–34.
25. Albain, K.S.; Barlow, W.E.; Shak, S.; Hortobagyi, G.N.; Livingston, R.B.; Yeh, I.T.; Ravdin, P.; Bugarini, R.; Baehner, F.L.; Davidson, N.E.; et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: A retrospective analysis of a randomised trial. *Lancet Oncol.* **2010**, *11*, 55–65. [[CrossRef](#)]
26. Roberts, M.C.; Bryson, A.; Weinberger, M.; Dusetzina, S.B.; Dinan, M.A.; Reeder-Hayes, K.; Wheeler, S.B. Oncologists' Barriers and Facilitators for Oncotype dx Use: Qualitative Study. *Int. J. Technol. Assess. Health Care* **2016**, *32*, 355–361. [[CrossRef](#)]
27. Iles, K.; Roberson, M.L.; Spanheimer, P.; Gallagher, K.; Ollila, D.W.; Strassle, P.D.; Downs-Canner, S. The impact of age and nodal status on variations in oncotype DX testing and adjuvant treatment. *npj Breast Cancer* **2022**, *8*, 27. [[CrossRef](#)]
28. Sparano, J.A.; Gray, R.J.; Makower, D.F.; Pritchard, K.I.; Albain, K.S.; Hayes, D.F.; Geyer, C.E.; Dees, E.C.; Goetz, M.P.; Olson, J.A.; et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N. Engl. J. Med.* **2018**, *379*, 111–121. [[CrossRef](#)] [[PubMed](#)]
29. Osareh, A.; Shadgar, B. Machine learning techniques to diagnose breast cancer. In Proceedings of the 2010 5th International Symposium on Health Informatics and Bioinformatics, Ankara, Turkey, 20–22 April 2010; pp. 114–120.
30. Amrane, M.; Oukid, S.; Gagaoua, I.; Ensari, T. Breast cancer classification using machine learning. In Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineering's Meeting (EBBT), Istanbul, Turkey, 18–19 April 2018; pp. 1–4.
31. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Andrew, H.B. Deep Learning for Identifying Metastatic Breast Cancer. *arXiv* **2016**, arXiv:1606.05718.
32. Ganggayah, M.D.; Taib, N.A.; Har, Y.C.; Lio, P.; Dhillon, S.K. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 48. [[CrossRef](#)]
33. Zhao, S.; Wang, L.; Ding, W.; Ye, B.; Cheng, C.; Shao, J.; Liu, J.; Zhou, H. Crosstalk of disulfidptosis-related subtypes, establishment of a prognostic signature and immune infiltration characteristics in bladder cancer based on a machine learning survival framework. *Front. Endocrinol.* **2023**, *14*, 1180404. [[CrossRef](#)] [[PubMed](#)]
34. Zou, Y.; Xie, J.; Zheng, S.; Liu, W.; Tang, Y.; Tian, W.; Deng, X.; Wu, L.; Zhang, Y.; Wong, C.W.; et al. Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery. *Int. J. Surg.* **2022**, *107*, 106936. [[CrossRef](#)]
35. Xie, J.; Luo, X.; Deng, X.; Tang, Y.; Tian, W.; Cheng, H.; Zhang, J.; Zou, Y.; Guo, Z.; Xie, X. Advances in artificial intelligence to predict cancer immunotherapy efficacy. *Front. Immunol.* **2022**, *13*, 1076883. [[CrossRef](#)] [[PubMed](#)]
36. Boutros, C.; Mazouni, C.; Lerebours, F.; Stevens, D.; Lei, X.; Gonzalez-Angulo, A.M.; Delalogue, S. A preoperative nomogram to predict the risk of synchronous distant metastases at diagnosis of primary breast cancer. *Br. J. Cancer* **2015**, *112*, 992–997. [[CrossRef](#)] [[PubMed](#)]
37. Nicolò, C.; Périer, C.; Prague, M.; Bellera, C.; MacGrogan, G.; Saut, O.; Benzekry, S. Machine Learning and Mechanistic Modeling for Prediction of Metastatic Relapse in Early-Stage Breast Cancer. *JCO Clin. Cancer Inform* **2020**, *4*, 259–274. [[CrossRef](#)]
38. Tseng, Y.-J.; Huang, C.-E.; Wen, C.-N.; Lai, P.-Y.; Wu, M.-H.; Sun, Y.-C.; Wang, H.-Y.; Lu, J.-J. Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. *Int. J. Med. Inform.* **2019**, *128*, 79–86. [[CrossRef](#)]
39. Tuladhar, A.; Gill, S.; Ismail, Z.; Forkert, N.D. Building machine learning models without sharing patient data: A simulation-based analysis of distributed learning by ensembling. *J. Biomed. Inform.* **2020**, *106*, 103424. [[CrossRef](#)]
40. Dluhoš, P.; Schwarz, D.; Cahn, W.; van Haren, N.; Kahn, R.; Španiel, F.; Horáček, J.; Kašpárek, T.; Schnack, H. Multi-center machine learning in imaging psychiatry: A meta-model approach. *NeuroImage* **2017**, *155*, 10–24. [[CrossRef](#)]

41. Sukhadia, S.S.; Tyagi, A.; Venkataraman, V.; Mukherjee, P.; Prasad, P.; Gevaert, O.; Nagaraj, S.H. ImaGene: A web-based software platform for tumor radiogenomic evaluation and reporting. *Bioinform. Adv.* **2022**, *2*, vbac079. [[CrossRef](#)]
42. Saha, A.; Harowicz, M.R.; Grimm, L.J.; Kim, C.E.; Ghate, S.V.; Walsh, R.; Mazurowski, M.A. A machine learning approach to radiogenomics of breast cancer: A study of 922 subjects and 529 DCE-MRI features. *Br. J. Cancer* **2018**, *119*, 508–516. [[CrossRef](#)]
43. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)] [[PubMed](#)]
44. Symmans, W.F.; Peintinger, F.; Hatzis, C.; Rajan, R.; Kuerer, H.; Valero, V.; Assad, L.; Poniecka, A.; Hennessy, B.; Green, M.; et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J. Clin. Oncol.* **2007**, *25*, 4414–4422. [[CrossRef](#)] [[PubMed](#)]
45. Korde, L.A.; Somerfield, M.R.; Carey, L.A.; Crews, J.R.; Denduluri, N.; Hwang, E.S.; Khan, S.A.; Loibl, S.; Morris, E.A.; Perez, A.; et al. Neoadjuvant Chemotherapy, Endocrine Therapy, and Targeted Therapy for Breast Cancer: ASCO Guideline. *J. Clin. Oncol.* **2021**, *39*, 1485–1505. [[CrossRef](#)] [[PubMed](#)]
46. Viale, G.; Fusco, N. Pathology after neoadjuvant treatment—How to assess residual disease. *Breast* **2022**, *62*, S25–S28. [[CrossRef](#)] [[PubMed](#)]
47. Fink, M.K. Does adjuvant therapy reduce postmetastatic survival? *Ann. Oncol.* **2019**, *30*, 1184–1188. [[CrossRef](#)]
48. Valagussa, P.; Tess, J.D.; Rossi, A.; Tancini, G.; Banfi, A.; Bonadonna, G. Adjuvant CMF effect on site of first recurrence, and appropriate follow-up intervals, in operable breast cancer with positive axillary nodes. *Breast Cancer Res. Treat.* **1981**, *1*, 349–356. [[CrossRef](#)]
49. Hölzel, D.; Eckel, R.; Bauerfeind, I.; Baier, B.; Beck, T.; Braun, M.; Ettl, J.; Hamann, U.; Kiechle, M.; Mahner, S.; et al. Improved systemic treatment for early breast cancer improves cure rates, modifies metastatic pattern and shortens post-metastatic survival: 35-year results from the Munich Cancer Registry. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 1701–1712. [[CrossRef](#)]
50. González-Castro, L.; Chávez, M.; Dufлот, P.; Bleret, V.; Martin, A.G.; Zobel, M.; Nateqi, J.; Lin, S.; Pazos-Arias, J.J.; Del Fiol, G.; et al. Machine Learning Algorithms to Predict Breast Cancer Recurrence Using Structured and Unstructured Sources from Electronic Health Records. *Cancers* **2023**, *15*, 2741. [[CrossRef](#)]
51. Lou, S.J.; Hou, M.F.; Chang, H.T.; Chiu, C.C.; Lee, H.H.; Yeh, S.J.; Shi, H.Y. Machine Learning Algorithms to Predict Recurrence within 10 Years after Breast Cancer Surgery: A Prospective Cohort Study. *Cancers* **2020**, *12*, 3817. [[CrossRef](#)]
52. Wang, H.; Li, Y.; Khan, S.A.; Luo, Y. Prediction of breast cancer distant recurrence using natural language processing and knowledge-guided convolutional neural network. *Artif. Intell. Med.* **2020**, *110*, 101977. [[CrossRef](#)]
53. Khene, Z.-E.; Bigot, P.; Doumerc, N.; Ouzaid, I.; Boissier, R.; Nouhaud, F.-X.; Albiges, L.; Bernhard, J.-C.; Ingels, A.; Borchiellini, D.; et al. Application of Machine Learning Models to Predict Recurrence After Surgical Resection of Nonmetastatic Renal Cell Carcinoma. *Eur. Urol. Oncol.* **2022**, *6*, 323–330. [[CrossRef](#)] [[PubMed](#)]
54. Akazawa, M.; Hashimoto, K.; Noda, K.; Yoshida, K. The application of machine learning for predicting recurrence in patients with early-stage endometrial cancer: A pilot study. *Obstet. Gynecol. Sci.* **2020**, *64*, 266–273. [[CrossRef](#)]
55. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [[CrossRef](#)] [[PubMed](#)]
56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, É. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.