







## Article

# Breast Tumor Characterization Using [<sup>18</sup>F]FDG-PET/CT Imaging Combined with Data Preprocessing and Radiomics

Denis Krajnc <sup>1</sup>, Laszlo Papp <sup>1</sup>, Thomas S. Nakuz <sup>2</sup>, Heinrich F. Magometschnigg <sup>3</sup>, Marko Grahovac <sup>2,4</sup>, Clemens P. Spielvogel <sup>2,4</sup>, Boglarka Ecsedi <sup>1</sup>, Zsuzsanna Bago-Horvath <sup>5</sup>, Alexander Haug <sup>2,4</sup>, Georgios Karanikas <sup>2</sup>, Thomas Beyer <sup>1,\*</sup>, Marcus Hacker <sup>2</sup>, Thomas H. Helbich <sup>3</sup> and Katja Pinker <sup>3,6</sup>

- <sup>1</sup> QIMP Team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, 1090 Vienna, Austria; denis.krajnc@meduniwien.ac.at (D.K.); laszlo.papp@meduniwien.ac.at (L.P.); ecsedi.bogi@gmail.com (B.E.)
- <sup>2</sup> Division of Nuclear Medicine, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, 1090 Vienna, Austria; thomas.nakuz@meduniwien.ac.at (T.S.N.); marko.grahovac@meduniwien.ac.at (M.G.); clemens.spielvogel@meduniwien.ac.at (C.P.S.); alexander.haug@meduniwien.ac.at (A.H.); georgios.karanikas@meduniwien.ac.at (G.K.); marcus.hacker@meduniwien.ac.at (M.H.)
- <sup>3</sup> Division of Molecular and Gender Imaging, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, 1090 Vienna, Austria; heinrich.magometschnigg@meduniwien.ac.at (H.F.M.); thomas.helbich@meduniwien.ac.at (T.H.H.); katja.pinker@meduniwien.ac.at or pinkerkd@mskcc.org (K.P.)
- <sup>4</sup> Christian Doppler Laboratory for Applied Metabolomics, Medical University of Vienna, 1090 Vienna, Austria
- <sup>5</sup> Department of Pathology, Medical University of Vienna, 1090 Vienna, Austria; zsuzsanna.horvath@meduniwien.ac.at
- <sup>6</sup> Memorial Sloan Kettering Cancer Center, Breast Imaging Service, Department of Radiology, New York, NY 10065, USA
- \* Correspondence: thomas.beyer@meduniwien.ac.at



**Citation:** Krajnc, D.; Papp, L.; Nakuz, T.S.; Magometschnigg, H.F.; Grahovac, M.; Spielvogel, C.P.; Ecsedi, B.; Bago-Horvath, Z.; Haug, A.; Karanikas, G.; et al. Breast Tumor Characterization Using [<sup>18</sup>F]FDG-PET/CT Imaging Combined with Data Preprocessing and Radiomics. *Cancers* **2021**, *13*, 1249. <https://doi.org/10.3390/cancers13061249>

Academic Editors:  
Andreas Stadlbauer,  
Anke Meyer-Baese and  
Max Zimmermann

Received: 5 February 2021  
Accepted: 9 March 2021  
Published: 12 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Simple Summary:** Breast cancer is the second most common diagnosed malignancy in women worldwide. In this study, we examine the feasibility of breast tumor characterization based on [<sup>18</sup>F]FDG-PET/CT images using machine learning (ML) approaches in combination with data-preprocessing techniques. ML prediction models for breast cancer detection and the identification of breast cancer receptor status, proliferation rate, and molecular subtypes were established and evaluated. Furthermore, the importance of most repeatable features was investigated. Results displayed high performance of malignant/benign tumor differentiation and triple negative tumor subtype ML models. We observed high repeatability of radiomic features for both high performing predictive models.

**Abstract:** *Background:* This study investigated the performance of ensemble learning holistic models for the detection of breast cancer, receptor status, proliferation rate, and molecular subtypes from [<sup>18</sup>F]FDG-PET/CT images with and without incorporating data pre-processing algorithms. Additionally, machine learning (ML) models were compared with conventional data analysis using standard uptake value lesion classification. *Methods:* A cohort of 170 patients with 173 breast cancer tumors (132 malignant, 38 benign) was examined with [<sup>18</sup>F]FDG-PET/CT. Breast tumors were segmented and radiomic features were extracted following the imaging biomarker standardization initiative (IBSI) guidelines combined with optimized feature extraction. Ensemble learning including five supervised ML algorithms was utilized in a 100-fold Monte Carlo (MC) cross-validation scheme. Data pre-processing methods were incorporated prior to machine learning, including outlier and borderline noisy sample detection, feature selection, and class imbalance correction. Feature importance in each model was assessed by calculating feature occurrence by the R-squared method across MC folds. *Results:* Cross validation demonstrated high performance of the cancer detection model (80% sensitivity, 78% specificity, 80% accuracy, 0.81 area under the curve (AUC)), and of the triple negative tumor identification model (85% sensitivity, 78% specificity, 82% accuracy, 0.82 AUC). The individual receptor status and luminal A/B subtype models yielded low performance (0.46–0.68 AUC). SUV<sub>max</sub> model yielded 0.76 AUC in cancer detection and 0.70 AUC in predicting triple negative subtype.

*Conclusions:* Predictive models based on [ $^{18}\text{F}$ ]FDG-PET/CT images in combination with advanced data pre-processing steps aid in breast cancer diagnosis and in ML-based prediction of the aggressive triple negative breast cancer subtype.

**Keywords:** breast cancer; radiomics; machine learning; PET/CT; data pre-processing; triple negative

## 1. Introduction

Breast cancer is the most common cancer in females, with over two million cases per year [1]. Among patients with a suspicious imaging abnormality at screening, image guided biopsy is used to confirm breast cancer diagnosis [2,3]. In breast cancer treatment assessment of receptor status (estrogen (ER), progesterone (PR) and Her2-neu receptor (HER2)) by immunohistochemistry (IHC) from breast biopsy is used for tumor subtype classification. Breast cancer molecular subtypes as determined by IHC (Luminal A, Luminal B, Her2 positive and Triple Negative) guide treatment decisions [4]. Nonetheless, breast cancer subtyping from biopsy sampling has limitations as it is subject to sampling bias and cannot fully capture intra-tumor heterogeneity [5–7]. In addition, there is its inherently invasive nature.

$^{18}\text{F}$ -fluorodeoxyglucose positron emission tomography/computed tomography ([ $^{18}\text{F}$ ]FDG-PET/CT) is a sensitive hybrid imaging method for detecting distant metastases and lymph node metastases in breast cancer patients [8] and for assessing treatment response [9,10]. Recently, dedicated PET/CT breast imaging protocols have shown potential for the classification and initial staging of primary tumors [11–13]. Despite first promising results for the non-invasive characterization of breast tumors, conventional PET/CT image analysis, including the standardized uptake value (SUV), tumor-to-background ratio (TBR), and metabolic tumor volume, remains of limited use for the differentiation of benign and malignant breast tumors and for molecular subtyping of breast cancers [14]. Therefore, several studies have performed radiomic analysis to further the value of [ $^{18}\text{F}$ ]FDG-PET/CT in this context [15–17].

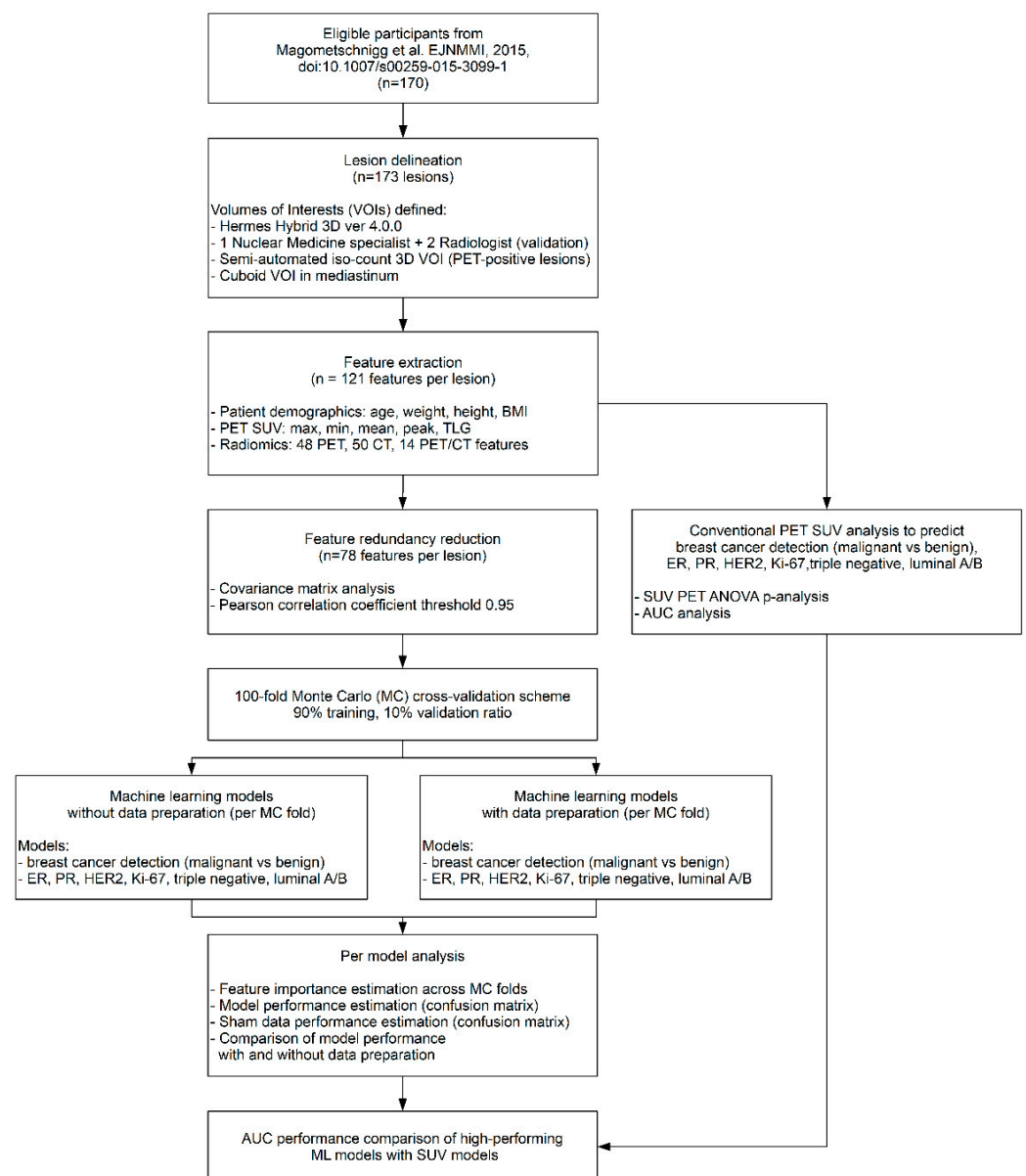
Radiomic analysis combined with machine learning (ML) has shown promise for characterizing tumor heterogeneity [18,19], assessing therapy response [20–22], and improving prognostic stratification of cancer patients [23,24]. However, the lack of repeatability for radiomic models has been noted as a major bottleneck for a clinical adoption [25,26]. The Imaging Biomarker Standardization Initiative (IBSI) [27] as well as optimized radiomics [15] and ComBat feature normalization [28] have been proposed as methodological considerations to support building quantitative radiomic models that can be translated reliably into the clinics. Radiomics combined with ML is prone to challenges originating from the characteristics of the input data itself, such as low sample count [29], imbalanced disease subgroups [30,31], high-dimensionality of data [32,33], and outliers [34,35]. To address these limitations, data preparation steps are necessary [36,37], yet data preparation approaches remain underrepresented in the field of hybrid imaging radiomics. Considering that breast cancer molecular subtypes are naturally imbalanced, with one subgroup of a given subtype, such as more aggressive triple negative (TN) or HER2 positive, being significantly underrepresented than hormone receptor subtypes (ER/PR positive) [20–22], we hypothesize that breast cancer in vivo prediction models benefit from data preparation approaches.

Therefore, the objectives of this study are: (a) to establish prediction models for breast cancer detection and the identification of breast cancer receptor status, proliferation rate, and molecular subtypes from [ $^{18}\text{F}$ ]FDG-PET/CT images with ML, (b) to investigate the effect of data pre-processing on breast tumor characterization ML models, and (c), to compare ML-based prediction models with conventional SUV-based approaches.

## 2. Materials and Methods

### 2.1. Patients

One hundred and seventy patients (median age, 57.6 years; range, 18–86 years) were examined with [<sup>18</sup>F]FDG-PET/CT imaging between 2009 and 2014 as part of a prospective study, which has been previously reported [11,13,38] and approved by the institutional review board of the Medical University of Vienna (EK 510-2009). Written informed consent was obtained from all patients prior to the imaging examinations. The inclusion criteria were as follows: age 18 years or older; and an abnormality at mammography or breast ultrasound (asymmetric density, architectural distortion, suspicious microcalcifications, or breast mass classified as Breast Imaging Reporting and Data System (BI-RADS category 0 or 4–5). Exclusion criteria included pregnancy, lactation, prior treatment (e.g., breast biopsy before PET/CT, neoadjuvant chemotherapy), or inadequate patient positioning resulting in considerably compressed or deformed imaging. For all patients, the following clinical information was recorded: height, weight, body mass index (BMI), and age. See Figure 1 for the study design of our analysis.



**Figure 1.** The analysis workflow of the collected dataset. Prospective study conducted between

2009 and 2014, approved by the institutional review board provided data records for 170 patients. [<sup>18</sup>F]FDG-PET/CT of the breast was performed with a dedicated breast imaging protocol using a combined whole-body PET/CT system. 173 lesions were delineated and extracted following the imaging biomarker standardization initiative (IBSI) guidelines combined with optimized feature extraction principles. Feature redundancy reduction was performed resulting in 77 features. Monte Carlo cross validation was utilized to generate 100 training vs. validation folds. Pre-processing steps were performed over training data. Ensemble learning scheme was utilized to establish predictive models. All machine learning models underwent confusion matrix analytics, sham data analysis, and Area Under the Receiver Operator Characteristics Curve (AUC) analysis across MC folds and the conventional PET SUV analysis. VOI = Volume of Interest; BMI = Body Mass Index; ER = Estrogen; PR = Progesterone; HER2 = Human Epidermal Growth Receptor 2; PET-Positron Emission Tomography; CT-Computed Tomography.

## 2.2. Histopathologic Analysis

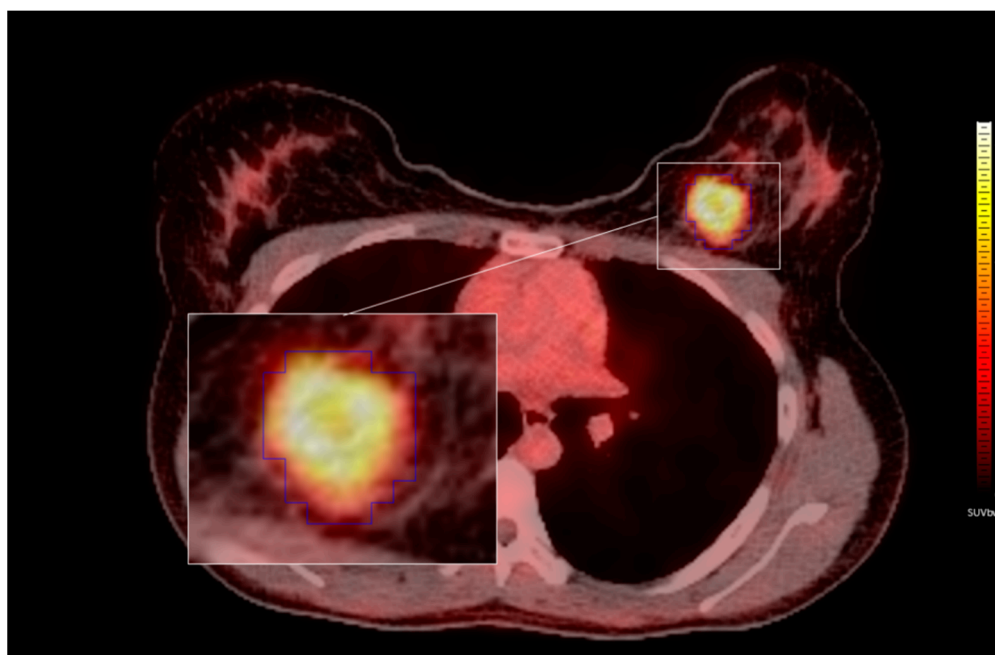
Diagnosis was established by an experienced specialized breast pathologist (ZBH). All lesions were verified by image-guided needle biopsy or surgery. For all invasive breast cancers, histopathology results were reviewed for tumor subtype according to the World Health Organization (WHO) classification [39], and tumor stage and grade according to Elston and Ellis [40]. Breast cancer intrinsic subtype was determined by immunohistochemistry based on estrogen receptor (ER), progesterone receptor (PR), human epidermal growth receptor 2 (HER2) status, and Ki-67 expression according to current guidelines [41], and defined as luminal A (ER/PR positive, Ki67 < 15%), luminal B (ER/PR positive, HER2 negative, Ki-67 ≥ 15% or ER/PR positive, HER2 positive), HER2 positive (ER/PR negative, HER2 positive), or triple negative (TN, ER/PR negative, HER2 negative) [42,43]. Patients with equivocal HER2 status were evaluated using chromogenic in situ hybridization to detect gene amplification. Patients with amplified genes were considered HER2 positive and patients whose genes were not amplified were considered as HER2 negative. In terms of Ki-67 expression, patients with ≥15% proliferation were considered as positive, while patients with <15% proliferation were classified as negative. HER2 positive and TN breast cancers were considered more aggressive breast cancers with a worse prognosis than luminal A/B breast cancers.

## 2.3. PET/CT

[<sup>18</sup>F]FDG-PET/CT of the breast was performed with a dedicated breast imaging protocol using a combined whole-body PET/CT system (Biograph 64 TruePoint®; Siemens Healthineers, Erlangen, Germany) with a high-resolution PET and a 64-row detector CT system. Patients were required to fast for at least 5 h before receiving an intravenous bolus injection of 200–350 MBq [<sup>18</sup>F]FDG based on body weight with blood glucose level < 150 mg/dL (8.3 mmol/L). After an uptake time of 60 min, PET/CT imaging was performed over one PET bed position with the patient consistently in the prone position [11,13,38]. The low-dose CT scan without CT contrast administration was acquired for attenuation correction covering a region from the base of the skull to the upper abdomen. Then, the PET acquisition was performed over the same region with 5 min acquisition time per bed position. CT images were reconstructed with 2 mm slice thickness. PET images were reconstructed using the iterative TrueX algorithm (Siemens), which incorporated resolution recovery [44,45]. Four iterations per 21 subsets were used with a matrix size of 168 × 168, a transaxial field of view (FOV) of 605 mm (pixel size of 3.6 mm), and a section thickness of 5 mm.

#### 2.4. Lesion Delineation

PET/CT images were delineated in the Hybrid 3D software (ver. 4.0.0., Hermes Medical Solutions, Stockholm, Sweden). PET-based SUV values were normalized by a cubic volume of interest (VOI) over the mediastinum to serve as background reference for TBR calculations [46]. Three-dimensional isocount-based lesion delineations were performed semi-automatically on the PET images by a nuclear medicine specialist, and then reviewed by two radiologists (Figure 2). Based on previously suggested minimum voxel count for radiomic analysis [47] the smallest analyzed lesion size was 1.56 cm<sup>3</sup>. Overall, 167 patients had one primary lesion delineated, while three patients had two delineated lesions, resulting in overall 173 lesions.



**Figure 2.** 18F-fluorodeoxyglucose positron emission tomography/computed tomography ([<sup>18</sup>F]FDG-PET/CT) view of a breast cancer patient with semi-automatically delineated volume of interest (VOI) in the PET image. Windowing: hot iron palette with SUV body weight (SUV<sub>bw</sub>) of 6.5 for PET and range of −100 to 200 Hounsfield units (HU) for CT. The patient underwent imaging procedure in prone position and view is shown following the radiological convention.

#### 2.5. Feature Extraction

Patient demographics (age, height, weight, body mass index (BMI)), conventional SUV PET (SUV<sub>mean</sub>, SUV<sub>max</sub>, SUV<sub>min</sub>, SUV<sub>peak</sub> and SUV<sub>TLC</sub>) and radiomic PET/CT features were combined to form a holomics dataset [25,48]. In order to support reproducibility of our study, radiomic features with “strong” and “very strong” consensus were extracted following the IBSI guidelines [27], combined with optimized feature extraction principles [15] from the 173 lesions. For each lesion, 48 PET features, 50 CT features, and 14 fusion PET/CT features were extracted and merged with patient demographics and SUV features, resulting in 121 features per lesion. See Supplemental Table S1 for the list of IBSI-conform radiomic features.

#### 2.6. Feature Redundancy Reduction

Covariance matrix analysis [49] was performed across the 120 features where features with absolute Pearson correlation coefficient greater than 0.95 were considered as redundant, resulting in 77 features for further analysis.



### 2.7. Predictive Model Establishment

Mixed ensemble learning of five Random Forest (RF) algorithms with various hyperparameter values [50] was utilized for model establishment to minimize the effect of method bias and to increase the predictive performance (Supplemental Table S2). The final model decision was obtained by majority vote across the five model predictions. The ensemble model scheme was utilized to establish breast cancer detection (malignant vs. benign), ER, PR, HER2, Ki-67, triple negative, and luminal A/B predictive models.

### 2.8. Model Performance Estimation

Hundred-fold Monte Carlo (MC) cross-validation with a training-to-validation ratio of 90%–10% was utilized for each model [51]. To estimate the performance of the established models compared with random guesses, sham data analysis was performed by random label permutations as done previously [37,52]. Confusion matrix (CM) analyses were employed to estimate model performance including accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operator characteristics curve (AUC) across the MC folds.

### 2.9. Estimating the Effect of Data Preparation

This study utilized data preparation methods prior to ML over the training dataset of each MC fold. These methods covered a range of preprocessing steps including outlier and borderline sample detection [34,53], feature ranking and selection [54,55], and class imbalance correction [56–58]. Feature ranking was performed by R-squared approach [59] where the 15 highest-ranking feature per MC fold were selected from the training set for ML analysis. Methods were utilized in a predefined order of steps (Supplemental Table S3). In order to estimate the effect of these methods on ML predictive performance, each model was established twice within the Monte Carlo cross-validation scheme: with and without data preparation.

### 2.10. Feature Importance Estimation

To estimate the feature importance per predictive model, the feature occurrences as selected by the R-squared ranking were calculated across the individual MC folds.

### 2.11. Conventional PET Correlation Analyses

Conventional PET correlation analyses were performed for each patient subgroup according to malignant/benign tumor status, receptor status, proliferation rate, and molecular subtype.  $SUV_{mean}$ ,  $SUV_{max}$ ,  $SUV_{min}$ ,  $SUV_{peak}$  and  $SUV_{TLG}$  PET-based features were analyzed by using the ANOVA  $p$ -value test method (Microsoft Excel 2016 software) with significance threshold of  $p < 0.05$ .

## 3. Results

### 3.1. Patients

Our cohort demonstrated highly imbalanced disease subgroups (Table 1). Out of 170 patients, 132 patients had a malignant breast tumor (78%) and 38 patients had a benign tumor (22%); 11 patients were classified as triple negative (6%), 22 as HER2 positive (13%), and 14 as luminal A (9%) vs. 81 as luminal B (81%). Furthermore, 88 patients were ER positive (52%), 78 were PR positive (46%), and 73 had a high number of Ki-67 positive cells (43%).

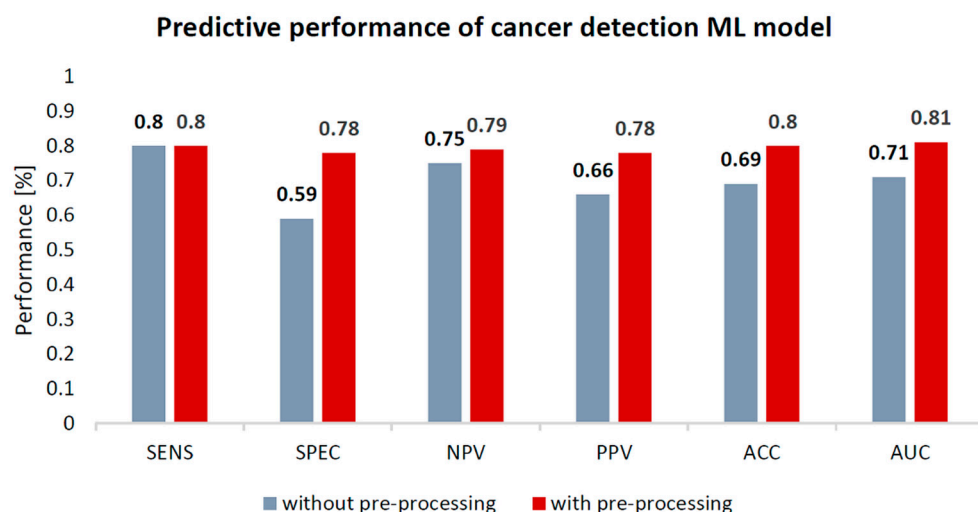
**Table 1.** Patient cohort characteristics for malignancy, estrogen (ER), progesterone (PR), human epidermal growth receptor 2 (HER2), Ki-67 protein expression, triple negative, and luminal A/B status. NA = Not Available.

Patient Characteristics (n = 170)	Value
Age (years), median (IQR)	57.6 (18–86)
Lesion volume (cm <sup>3</sup> ), median (IQR)	12.8 (6.2–26.9)
Malignancy	n (%)
Malignant	132 (78)
Benign	38 (22)
Estrogen (ER)	n (%)
–	17 (10)
+	88 (52)
NA	65 (38)
Progesterone (PR)	n (%)
–	27 (16)
+	78 (46)
NA	65 (38)
Ki-67	n (%)
–	26 (15)
+	73 (43)
NA	71 (42)
HER2	n (%)
–	84 (49)
+	22 (13)
NA	64 (38)
Triple negative	n (%)
Yes	11 (6)
No	95 (56)
NA	64 (38)
Luminal A/B	n (%)
A	14 (8)
B	81 (48)
NA	75 (44)

### 3.2. Model Performance Estimation

#### 3.2.1. Breast Cancer Detection

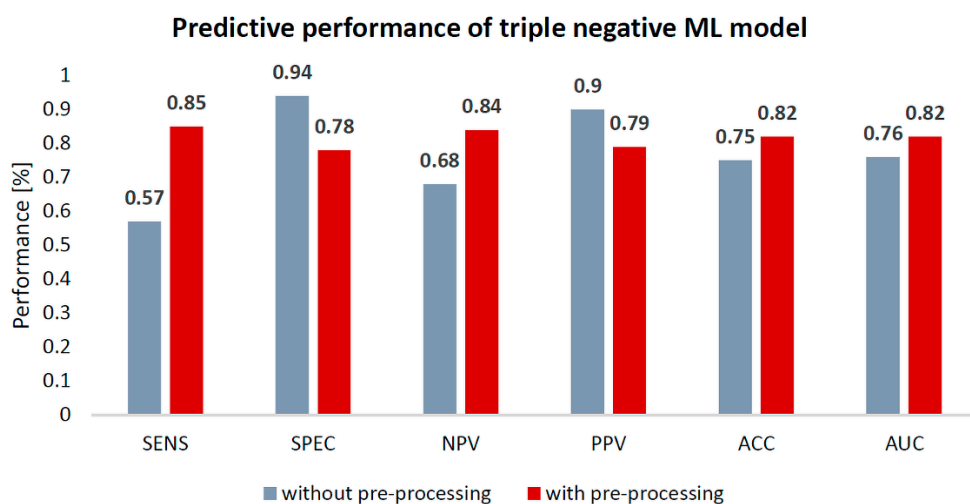
The model for differentiation of benign and malignant breast tumors/breast cancer detection with data preparation yielded 80% sensitivity, 78% specificity, 80% accuracy and 0.81 AUC, compared to the same model without data preparation (80% sensitivity, 59% specificity, 69% accuracy and 0.71 AUC). See Figure 3 for the performance comparison of the breast cancer detection models.



**Figure 3.** Performance comparison of breast cancer detection machine learning (ML) predictive models, with and without data pre-processing. ACC = Accuracy; SENS = Sensitivity; SPEC = Specificity; NPV = Negative Predictive Value; PPV = Positive Predictive Value. Performance is expressed in percentages (%).

### 3.2.2. Breast Cancer Subtyping

The highest cross-validation performance was achieved with the molecular subtyping ML model for triple negative breast cancer with data preparation which yielded 85% sensitivity, 78% specificity, 82% accuracy and 0.82 AUC. In contrast, the same model without data preparation yielded 59% sensitivity, 94% specificity, 75% accuracy and 0.76 AUC. See Figure 4 for the performance comparison of the triple negative models.



**Figure 4.** Performance comparison of triple negative subtype machine learning (ML) predictive models, with and without data pre-processing. ACC = Accuracy; SENS = Sensitivity; SPEC = Specificity; NPV = Negative Predictive Value; PPV = Positive Predictive Value. Performance is expressed in percentages (%).



Data preparation did not impact ML model performance for the prediction of individual receptor status and proliferation rate (0.46–0.68 AUC).

Table 2 summarizes the Monte Carlo cross-validation performance of all ensemble predictive models with and without data preparation. Predictive performance of all models over sham data yielded 0.47–0.59 AUC (Supplemental Table S4).

**Table 2.** Monte Carlo cross-validation performance of all ensemble predictive models with and without data preparation. Confusion matrix values are expressed in percentages (%). AUC is expressed in ratio.

Model	Data Preprocessing	SENS	SPEC	NPV	PPV	ACC	AUC
ER	No	83	40	70	58	62	0.63
	Yes	82	56↑	78↑	65↑	69↑	0.68↑
PR	No	74	36	58	54	55	0.56
	Yes	78↑	35	61↑	54	56↑	0.55
Ki-67	No	68	39	55	53	53	0.63
	Yes	65	45↑	56↑	54↑	55↑	0.65↑
HER2	No	17	84	50	51	50	0.46
	Yes	17	84	50	51	50	0.46
Luminal A/B	No	17	87	51	57	52	0.62
	Yes	16	89↑	51	59↑	53↑	0.52
Triple negative	No	57	94	68	90	75	0.76
	Yes	85↑	78	84↑	79	82↑	0.82↑
Breast Cancer Detection (Malignant vs. Benign)	No	80	59	75	66	69	0.71
	Yes	80	78↑	79↑	78↑	80↑	0.81↑

ACC = Accuracy, AUC = Area under the receiver operator characteristic curve, SENS = Sensitivity, SPEC = Specificity, NPV = Negative Predictive Value, PPV = Positive Predictive Value, ER = Estrogen, HER2 = Human Epidermal Growth Receptor 2, PR = Progesterone. Sign ↑ indicates performance increase in pre-processed training datasets compared to original datasets.

### 3.3. Feature Importance Estimation

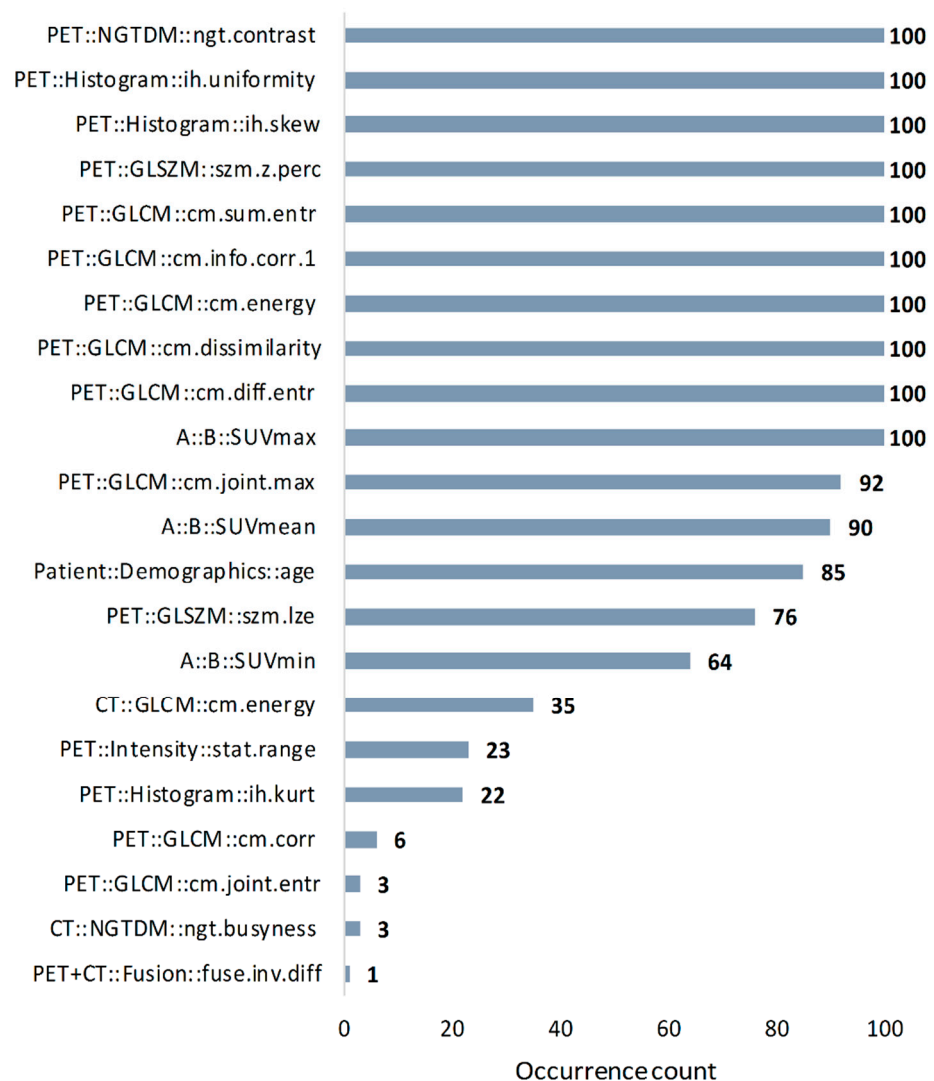
#### 3.3.1. Breast Cancer Detection

In the cancer detection predictive model, nine out of ten most relevant features for breast cancer detection originated from PET images. Features with highest occurrence number ( $n = 100$ ) were five PET gray level co-occurrence matrix (GLCM) features (sum entropy, energy, difference entropy, information correlation 1, dissimilarity), two PET histogram features (skewness, uniformity), PET neighborhood grey tone difference matrix (NGTDM) contrast and  $SUV_{max}$  feature. High occurrence was also observed in PET GLCM joint maximum ( $n = 92$ ),  $SUV_{mean}$  ( $n = 90$ ) and patient demographics age feature ( $n = 85$ ). See Figure 5 for all selected features in the breast cancer detection ML model.

#### 3.3.2. Breast Cancer Subtyping

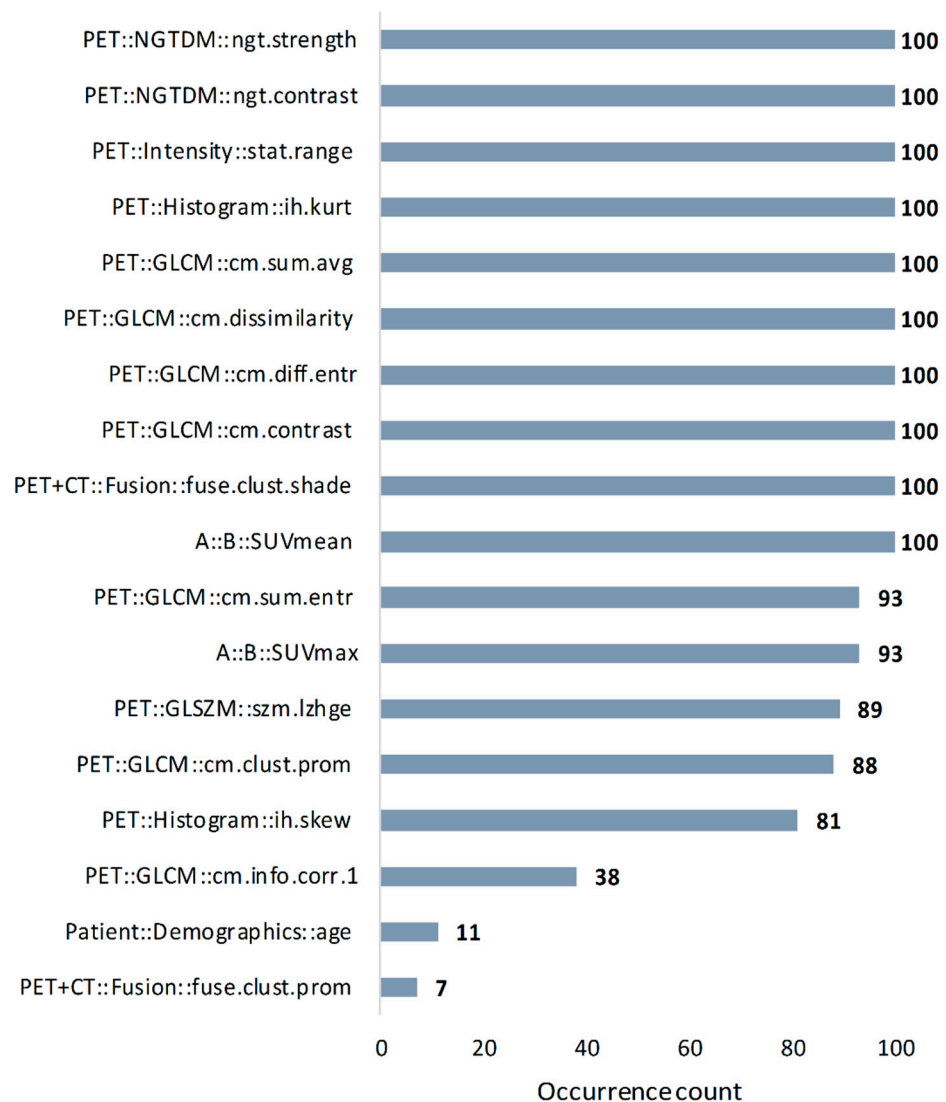
In the triple negative predictive model eight out of ten most relevant features originated from PET images. Features with highest occurrence number ( $n = 100$ ) were four PET GLCM features (contrast, difference entropy, dissimilarity, sum average), two PET NGTDM features (contrast, strength), PET histogram kurtosis, PET intensity range, PET + CT fusion cluster shade and  $SUV_{mean}$  feature. High occurrence was also observed in PET GLCM sum entropy ( $n = 93$ ),  $SUV_{max}$  ( $n = 93$ ), PET GLSZM large zone high grey level emphasis ( $n = 89$ ), PET GLCM cluster prominence ( $n = 88$ ) and PET histogram skewness feature ( $n = 81$ ). See Figure 6 for all selected features in the triple negative ML model.

### Occurrence of high-ranking features across the 100 Monte Carlo folds in cancer detection ML model



**Figure 5.** Occurrence of high-ranking features across the 100 Monte Carlo folds in cancer detection predictive model. NGTDM = neighborhood grey tone difference matrix; GLSZM = gray level size zone matrix; GLCM = gray level co-occurrence matrix;  $SUV_{max}$  = maximum standard uptake value;  $SUV_{mean}$  = mean standard uptake value;  $SUV_{min}$  = minimal standard uptake value; skew = skewness; z.perc = zone percentage; entr = entropy; info.corr.1 = information correlation 1; joint.max = joint maximum; lze = large zone emphasis; kurt = kurtosis; corr, correlation; joint.entr = joint entropy; inv.diff = inversed difference.

### Occurrence of high-ranking features across the 100 Monte Carlo folds in triple negative ML model



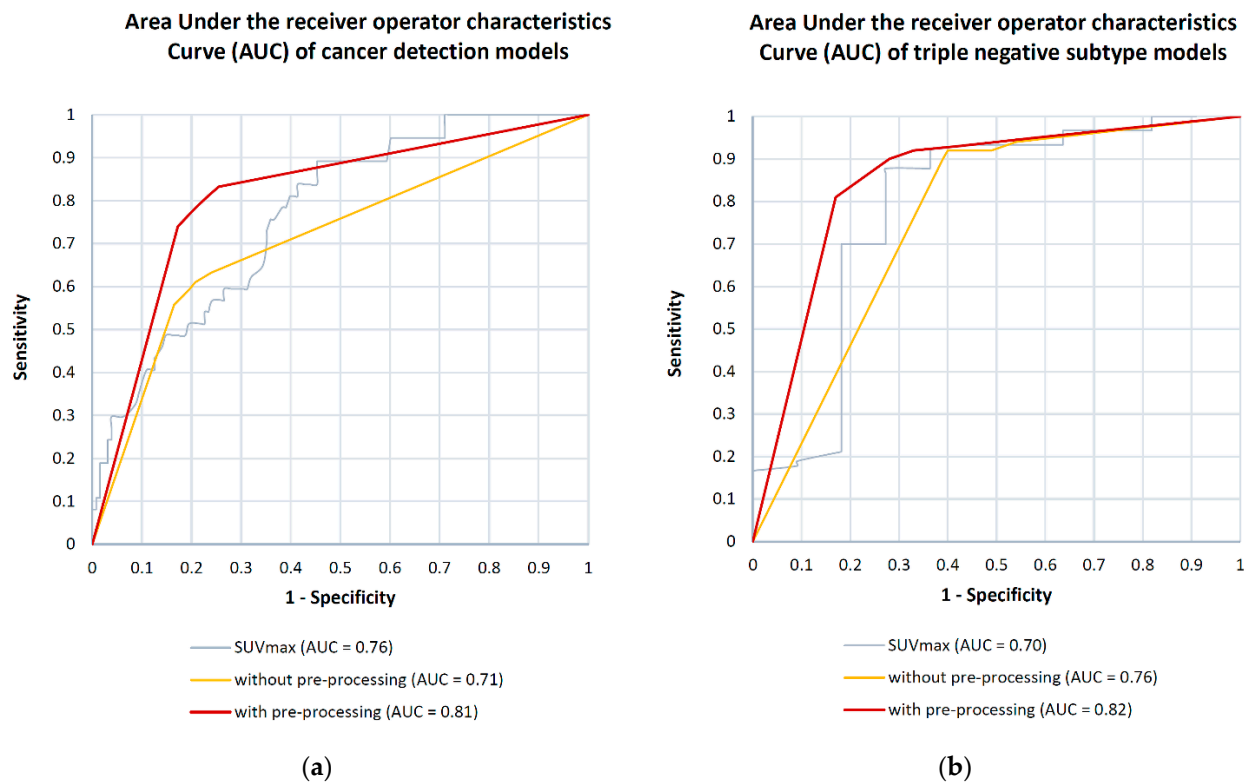
**Figure 6.** Occurrence of high-ranking features across the 100 Monte Carlo folds in triple negative predictive model. NGTDM = neighborhood grey tone difference matrix; GLCM = gray level co-occurrence matrix; GLSZM = gray level size zone matrix;  $SUV_{max}$  = maximum standard uptake value;  $SUV_{mean}$  = mean standard uptake value; kurt = kurtosis; sum.avg = sum average; diff.entr = difference entropy; clust.shade = cluster shade; sum.entr = sum entropy; lzhge = large zone high grey level emphasis; clust.prom = cluster prominence; skew = skewness; info.corr.1 = information correlation 1.

#### 3.4. Conventional PET Correlation Analysis

Supplemental Table S5 summarizes SUV correlation metrics for malignant-vs-benign breast tumors ( $SUV_{max}$   $p = 0.0002$ ) as well as malignant tumors stratified by receptor status and molecular triple negative subtype ( $SUV_{max}$   $p = 0.000001$ ). Significant differences in  $SUV_{max}$  distributions were present in ER ( $SUV_{max}$   $p = 0.00016$ ), PR ( $SUV_{max}$   $p = 0.003$ ), and Ki-67 ( $SUV_{max}$   $p = 0.003$ ) subgroups. In contrast, HER2 ( $SUV_{max}$   $p = 0.54$ ) and luminal A/B subgroups ( $SUV_{max}$   $p = 0.81$ ) demonstrated low correlation.

The highest performance of the SUV models was demonstrated by  $SUV_{max}$  in cancer detection (0.76 AUC) and predicting triple negative subtype (0.70 AUC). See Figure 7 for comparison of AUC performance of  $SUV_{max}$  and holomics-based predictive models in

cancer detection and triple negative subtype. See Supplemental Table S6 for comparison of holomics-based and SUV-based ML predictive performance across all models.



**Figure 7.** Comparison of area under the receiver operator characteristics curve (AUC) performance of maximum standard uptake value ( $SUV_{max}$ ) and holomics-based ensemble models with and without data pre-processing for (a) breast cancer detection (b) triple negative subtype.

#### 4. Discussion

This study investigated the performance of ML predictive models based on [ $^{18}F$ ]FDG-PET/CT ML analysis of 173 breast tumors in 170 patients with and without data preparation. Our study shows that data pre-processing contributes to model performance of the breast cancer detection ML model (80% vs. 69% accuracy, 0.81 vs. 0.77 AUC) and the aggressive triple-negative breast cancer subtype ML model (82% vs. 75% accuracy, 0.82 vs. 0.76 AUC). Nonetheless, our findings regarding the molecular subtype ML models also imply that data pre-processing alone does not warrant performance improvement, unless there is already an identifiable pattern in the imbalanced subgroups. The low performance observed in our molecular subtype ML models is in agreement with radiomics studies investigating the predictability of molecular subtypes in breast cancer with dynamic contrast-enhanced magnetic resonance imaging (MRI) and diffusion-weighted imaging [60–62]. We consider that the low performance of these models is due to the fact that compared with individual receptors, molecular subtypes such as the triple negative subtype are determined by the information from all receptors and carry distinct radiomics signatures.

Feature importance estimation analysis in our two highest performing models (breast cancer detection and triple negative, both with data preparation) revealed that PET is the most important information source to establish these models. Specifically, 16 out of 18 prominent features and 18 out of 20 prominent features selected across MC folds were from PET in the two models respectively (Figures 5 and 6). Furthermore, only two and three of PET features were conventional SUV parameters respectively. The prominent role of  $SUV_{max}$  was identifiable in both models ( $n = 100$  for breast cancer detection and  $n = 93$  for triple negative) implying, that radiomics and conventional SUV metrics in

combination can maximize the predictive performance of these models instead of building on only one of these feature sets. These findings are in alignment with recent studies investigating the importance of PET radiomic and SUV parameters in characterizing tumors in vivo [37,50]. Radiomic PET feature types represented a wide range in both models, where most prominent features were extracted from the neighborhood gray tone difference (NGTDM) and gray level co-occurrence (GLCM) matrices. These matrices are both designed to describe heterogeneity characteristics of lesions [27]. As an example, the NGTDM contrast feature which was identified as high-ranking ( $n = 100$ ) in both models, reflects on spatial intensity changes in between neighboring voxels.

The low importance of CT features in both models needs to be interpreted with caution. While CT may not represent heterogeneity patterns on its own, it contributes to PET attenuation correction as part of the PET/CT hybrid scanner [63], therefore, any prominent role identified in PET is inherently influenced by the presence of CT as well. Furthermore, the GLCM cluster shade PET/CT fusion feature was identified with high importance ( $n = 100$ ) in the breast cancer detection. This feature implies that the co-occurrence pattern of spatially-overlapping PET and CT voxels can contribute to differentiate malignant and benign breast lesions.

Patient age was the only demographics feature identified as highly important ( $n = 85$ ) in the breast cancer detection model [64], while it was present with negligible importance ( $n = 11$ ) in the triple negative predictive model. To date, no studies of cancer detection or triple negative subtype prediction based on PET/CT radiomics in compliance with IBSI has been performed, hence the comparison of our findings on feature repeatability to other studies remains of interest.

To date, studies that have analyzed radiomic features based on [ $^{18}\text{F}$ ]FDG-PET/CT in patients with breast cancer have focused on building models to predict pathological complete response to neoadjuvant chemotherapy or to differentiate breast carcinoma from breast lymphoma [20,22,65]. Antunovic et al. reported AUCs of 0.70–0.73 across all predictive models [22]; Li et al. reported AUCs of 0.72 and 0.73, respectively, without and with patient age incorporated [20]; and Ou et al. reported AUCs of 0.81 and 0.76 for PET and CT models, respectively [65]. Huang et al. [66] reported mean AUCs of 0.75 and 0.68 for one-year and two-year recurrence-free survival, respectively, from using PET/MRI-based models. The sample size of the patient cohort in all these studies ranged from 44–113 patients, which is lower compared to that used in our work ( $n = 170$ ). In addition, none of the above studies utilized data preprocessing approaches and they also did not build on mixed ensemble learning to minimize method selection bias in their relatively small patient cohorts.

To date, data pre-processing steps have been rarely discussed in PET-based radiomic studies. Zhou et al. [67] implemented class imbalance correction in a breast cancer cohort of 55 patients in their MRI-based radiomics study; specifically, the minority subclass was corrected using the Synthetic Minority Oversampling Technique (SMOTE). They reported AUCs of 0.81–0.87 across six different predictive MRI-based models to predict response to neoadjuvant chemotherapy [67]. Cysouw et al. [37] performed imbalance correction applying the SMOTE algorithm, and in addition they utilized principal component analysis to reduce the high number of features while retaining 95% of the observed variance to characterize prostate cancer in [ $^{18}\text{F}$ ]DCFPyL PET. Xie et al. [36] investigated class imbalance solutions in a cohort of head and neck cancer patients in their [ $^{18}\text{F}$ ]FDG-PET/CT-based radiomics study, by testing various resampling techniques for generating minority subclass samples and for cleaning noisy and redundant data. They reported performance increase of 0.32 (AUC) with applying data resampling techniques.

Our study differed in several aspects to the aforementioned studies. First, Zhou et al. and Cysouw et al. did not consider the presence of noisy/borderline data samples which may decrease the overall performance of the established models [68–70]. Second, none of the prior studies handled outliers in their training datasets.

In our study, conventional PET-based correlation analysis showed that  $SUV_{max}$ ,  $SUV_{mean}$ , and  $SUV_{min}$  were significantly different between malignant and benign tumors, which is in agreement with prior publications [11,71,72]. For predicting individual breast cancer receptor status and proliferation rate, conventional PET correlation analysis showed that  $SUV_{max}$  and  $SUV_{mean}$  were significantly different according to ER and PR status, and according to Ki-67 protein expression. For predicting molecular subtype, significance differences in standard SUV metrics were found only for the triple negative subtype ( $SUV_{max}$  and  $SUV_{mean}$ ,  $p < 0.001$  for both) while for the HER2 and luminal A/B subtype, standard SUV metrics showed no significant differences. Although there were significant differences in SUV metrics for ER and PR expression and the triple negative subtype, SUV-based models resulted in poor predictive performance.

Compared with the radiomics ML models, the SUV-based models had a lower AUC performance for differentiating between malignant and benign tumors (0.76 AUC vs. 0.81 AUC) as well as triple negative subtypes (0.70 AUC vs. 0.82 AUC). Performance difference is even more expressed in other confusion matrix analytics metrics, where we observed lower accuracy (ACC) performance of SUV-based models in cancer detection (56% vs. 80%) as well as in triple negative subtype prediction (74% vs. 82%).

Our study had limitations: our analysis was based on data from a single center only. Nevertheless, we extracted highly-repeatable radiomic features from our data as of the IBSI-standard together with optimized radiomic parameter sets [15,27]. In addition, we utilized Monte Carlo cross-validation scheme in combination with ensemble learning to minimize the effect of selection bias in both our data and ML methods. Last, we performed sham data analysis by random label permutation to estimate the performance compared to random guess.

Considering the high repeatability of our identified high-ranking features and the high reproducibility nature of RF classifiers [73], we consider that our findings could be reproduced by other centers building on our methodological approaches. The results of our study indicate that future radiomic studies can benefit from data pre-processing steps before conducting ML analyses especially if the given disease subgroups are highly imbalanced.

Of note, this study did not aim to investigate solid vs. invasive tumor sensitivities independently, rather, the effect of data preparation on overall ML performance which analyzes all types of tumors. Nevertheless, lesion delineation in our study was PET-driven. PET has high sensitivity, but at the same time it is prone to partial volume effects which naturally results in overestimating lesion volumes. Therefore, we assume that boundary regions of invasive tumors were not underrepresented in our analysis.

## 5. Conclusions

The diagnostic accuracy of [ $^{18}F$ ]FDG-PET/CT of breast cancer detection and prediction of the aggressive triple negative molecular subtype of breast cancer improved following the use of advanced data pre-processing in radiomic models. Radiomics analysis of [ $^{18}F$ ]FDG-PET/CT aid the differentiation of benign and malignant tumors in patients that cannot be assessed sufficiently with conventional breast imaging and who are not candidates for MRI. Results indicate that aggressive triple negative breast cancers that often require intensified or presurgical treatment carry distinct radiomics signatures and can be separated from less aggressive subtypes. However, radiomics analysis of [ $^{18}F$ ]FDG-PET/CT is limited in value for the prediction of individual receptor status and proliferation rate.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2072-6694/13/6/1249/s1>. Supplemental Table S1: Imaging Biomarker Standardization Initiative (IBSI) reporting structure of the study. The information presented herein is based on the IBSI guidelines, Supplemental Table S2: Algorithms settings of the 5 RF models employed in the ensemble learning scheme, Supplemental Table S3: Data preparation pipelines across all machine learning predictive models, Supplemental Table S4: Machine learning results of best performing models (per reference label) over sham data, Supplemental Table S5: Conventional positron emission tomography (PET)-



based correlation analysis, Supplemental Table S6: Holomics-based vs. standard uptake value (SUV)-based ML performance comparison across all predictive models.

**Author Contributions:** Concept and design: D.K., L.P., T.B., K.P., Data acquisition: K.P., H.F.M., T.H.H., G.K., A.H., Z.B.-H., T.S.N., Data analysis/interpretation: D.K., T.S.N., M.G., C.P.S., B.E., K.P., Drafting of the manuscript: D.K., L.P., K.P., Critical revision of the manuscript: All, Statistical analysis: D.K., Funding acquisition: K.P., Administration, financial, or material support: T.B., Supervision: K.P., T.B., M.H., L.P., T.H.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by Senologie-Forschungsfoerderungsprieses, Jubiläumsfonds of the Austrian National Bank (OENB#13652), and Scientific Funds of the Mayor of Vienna. Katja Pinker was also supported in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

**Institutional Review Board Statement:** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Ethik Kommission der Medizinischen Universität Wien (EK 510-2009).

**Informed Consent Statement:** Informed consent was obtained from all individual participants included in the study.

**Data Availability Statement:** Available upon reasonable request.

**Conflicts of Interest:** Katja Pinker reported payment for service on speakers bureaus from the European Society of Breast Imaging, Siemens Healthineers, and IDKD 2019, and membership on the advisory board of Merantix Healthcare GmbH. Laszlo Papp, Marcus Hacker, and Thomas Beyer are co-founders of Dedicaid GmbH.

## References

- World Health Organization. Estimated Age-Standardized Incidence and Mortality Rates (World) in 2020, Worldwide, Both Sexes, All Ages. Available online: [https://gco.iarc.fr/today/online-analysis-multi-bars?v=2020&mode=cancer&mode\\_population=countries&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population\\_group=0&ages\\_group%5B%5D=0&ages\\_group%5B%5D=17&nb\\_items=10&group\\_cancer=1&include\\_nmssc=1&include\\_nmssc\\_other=1&type\\_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type\\_sort=0&type\\_nb\\_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D](https://gco.iarc.fr/today/online-analysis-multi-bars?v=2020&mode=cancer&mode_population=countries&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmssc=1&include_nmssc_other=1&type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type_sort=0&type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D) (accessed on 18 January 2021).
- Loughran, C.F.; Keeling, C.R. Seeding of tumour cells following breast biopsy: A literature review. *Br. J. Radiol.* **2011**, *84*, 869–874. [\[CrossRef\]](#)
- White, R.R.; Halperin, T.J.; Olson, J.A., Jr.; Soo, M.S.; Bentley, R.C.; Seigler, H.F. Impact of Core-Needle Breast Biopsy on the Surgical Management of Mammographic Abnormalities. *Ann. Surg.* **2001**, *233*, 769–777. [\[CrossRef\]](#)
- Zaha, D.C. Significance of immunohistochemistry in breast cancer. *World J. Clin. Oncol.* **2014**, *5*, 382. [\[CrossRef\]](#)
- Boba, M.; Kołtun, U.; Bobek-Billewicz, B.; Chmielik, E.; Eksner, B.; Olejnik, T. False-negative results of breast core needle biopsies—Retrospective analysis of 988 biopsies. *Pol. J. Radiol.* **2011**, *76*, 25–29. [\[PubMed\]](#)
- Haynes, B.; Sarma, A.; Nangia-Makker, P.; Shekhar, M.P. Breast cancer complexity: Implications of intratumoral heterogeneity in clinical management. *Cancer Metastasis Rev.* **2017**, *36*, 547–555. [\[CrossRef\]](#)
- Cajal, S.R.Y.; Sesé, M.; Capdevila, C.; Aasen, T.; De Mattos-Arruda, L.; Diaz-Cano, S.J.; Hernández-Losa, J.; Castellví, J. Clinical implications of intratumor heterogeneity: Challenges and opportunities. *J. Mol. Med.* **2020**, *98*, 161–177. [\[CrossRef\]](#)
- Garg, P.K.; Deo, S.V.S.; Kumar, R.; Shukla, N.K.; Thulkar, S.; Gogia, A.; Sharma, D.N.; Mathur, S.R. Staging PET–CT Scanning Provides Superior Detection of Lymph Nodes and Distant Metastases than Traditional Imaging in Locally Advanced Breast Cancer. *World J. Surg.* **2016**, *40*, 2036–2042. [\[CrossRef\]](#) [\[PubMed\]](#)
- Humbert, O.; Riedinger, J.M.; Vrigneaud, J.M.; Kanoun, S.; Dygai-Cochet, I.; Berriolo-Riedinger, A.; Toubreau, M.; Depardon, E.; Lassere, M.; Tisserand, S.; et al. 18F-FDG PET-Derived Tumor Blood Flow Changes After 1 Cycle of Neoadjuvant Chemotherapy Predicts Outcome in Triple-Negative Breast Cancer. *J. Nucl. Med.* **2016**, *57*, 1707–1712. [\[CrossRef\]](#) [\[PubMed\]](#)
- Basu, S.; Alavi, A. PET-Based Personalized Management in Clinical Oncology. *PET Clin.* **2016**, *11*, 203–207. [\[CrossRef\]](#) [\[PubMed\]](#)
- Magometschnigg, H.F.; Baltzer, P.A.; Fueger, B.; Helbich, T.H.; Karanikas, G.; Dubsy, P.; Rudas, M.; Weber, M.; Pinker, K. Diagnostic accuracy of 18F-FDG PET/CT compared with that of contrast-enhanced MRI of the breast at 3 T. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 1656–1665. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xiao, Y.; Wang, L.; Jiang, X.; She, W.; He, L.; Hu, G. Diagnostic efficacy of 18F-FDG-PET or PET/CT in breast cancer with suspected recurrence. *Nucl. Med. Commun.* **2016**, *37*, 1180–1188. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Pinker, K.; Bogner, W.; Baltzer, P.; Karanikas, G.; Magometschnigg, H.; Brader, P.; Gruber, S.; Bickel, H.; Dubsky, P.; Bago-Horvath, Z.; et al. Improved differentiation of benign and malignant breast tumors with multiparametric 18fluorodeoxyglucose positron emission tomography magnetic resonance imaging: A feasibility study. *Clin. Cancer Res.* **2014**, *20*, 3540–3549. [[CrossRef](#)]
14. Visvikis, D.; Hatt, M.; Tixier, F.; Le Rest, C.C. The age of reason for FDG PET image-derived indices. *Eur. J. Nucl. Med. Mol. Imaging* **2012**, *39*, 1670–1672. [[CrossRef](#)]
15. Papp, L.; Rausch, I.; Grahovac, M.; Hacker, M.; Beyer, T. Optimized feature extraction for radiomics analysis of 18 F-FDG-PET imaging. *J. Nucl. Med.* **2018**. [[CrossRef](#)]
16. Castiglioni, I.; Gallivanone, F.; Soda, P.; Avanzo, M.; Stancanello, J.; Aiello, M.; Interlenghi, M.; Salvatore, M. AI-based applications in hybrid imaging: How to build smart and truly multi-parametric decision models for radiomics. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 2673–2699. [[CrossRef](#)]
17. Veit-Haibach, P.; Buvat, I.; Herrmann, K. EJNMMI supplement: Bringing AI and radiomics to nuclear medicine. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 2627–2629. [[CrossRef](#)]
18. Chicklore, S.; Goh, V.; Siddique, M.; Roy, A.; Marsden, P.K.; Cook, G.J.R. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur. J. Nucl. Med. Mol. Imaging* **2013**, *40*, 133–140. [[CrossRef](#)]
19. Groheux, D.; Majdoub, M.; Tixier, F.; Le Rest, C.C.; Martineau, A.; Merlet, P.; Espié, M.; de Roquancourt, A.; Hindié, E.; Hatt, M.; et al. Do clinical, histological or immunohistochemical primary tumour characteristics translate into different 18F-FDG PET/CT volumetric and heterogeneity features in stage II/III breast cancer? *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 1682–1691. [[CrossRef](#)]
20. Li, P.; Wang, X.; Xu, C.; Liu, C.; Zheng, C.; Fulham, M.J.; Feng, D.; Wang, L.; Song, S.; Huang, G. 18F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1116–1126. [[CrossRef](#)]
21. Ha, S.; Park, S.; Bang, J.-I.; Kim, E.-K.; Lee, H.-Y. Metabolic Radiomics for Pretreatment 18F-FDG PET/CT to Characterize Locally Advanced Breast Cancer: Histopathologic Characteristics, Response to Neoadjuvant Chemotherapy, and Prognosis. *Sci. Rep.* **2017**, *7*, 1556. [[CrossRef](#)]
22. Antunovic, L.; De Sanctis, R.; Cozzi, L.; Kirienko, M.; Sagona, A.; Torrissi, R.; Tinterri, C.; Santoro, A.; Chiti, A.; Zelic, R.; et al. PET/CT radiomics in breast cancer: Promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 1468–1477. [[CrossRef](#)] [[PubMed](#)]
23. Groheux, D.; Giacchetti, S.; Moretti, J.-L.; Porcher, R.; Espié, M.; Lehmann-Che, J.; de Roquancourt, A.; Hamy, A.-S.; Cuvier, C.; Vercellino, L.; et al. Correlation of high 18F-FDG uptake to clinical, pathological and biological prognostic factors in breast cancer. *Eur. J. Nucl. Med. Mol. Imaging* **2011**, *38*, 426–435. [[CrossRef](#)] [[PubMed](#)]
24. Koo, H.R.; Park, J.S.; Kang, K.W.; Han, W.; Park, I.A.; Moon, W.K. Correlation between 18F-FDG uptake on PET/CT and prognostic factors in triple-negative breast cancer. *Eur. Radiol.* **2015**, *25*, 3314–3321. [[CrossRef](#)] [[PubMed](#)]
25. Papp, L.; Spielvogel, C.P.; Rausch, I.; Hacker, M.; Beyer, T. Personalizing Medicine Through Hybrid Imaging and Medical Big Data Analysis. *Front. Phys.* **2018**, *6*. [[CrossRef](#)]
26. Vallières, M.; Zwanenburg, A.; Badic, B.; Le Rest, C.C.; Visvikis, D.; Hatt, M. Responsible Radiomics Research for Faster Clinical Translation. *J. Nucl. Med.* **2018**, *59*, 189–193. [[CrossRef](#)]
27. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)] [[PubMed](#)]
28. Orhac, F.; Boughdad, S.; Philippe, C.; Stalla-Bourdillon, H.; Nioche, C.; Champion, L.; Soussan, M.; Frouin, F.; Frouin, V.; Buvat, I. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. *J. Nucl. Med.* **2018**, *59*, 1321–1328. [[CrossRef](#)]
29. Raudys, S.J.; Jain, A.K. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 252–264. [[CrossRef](#)]
30. Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [[CrossRef](#)]
31. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [[CrossRef](#)]
32. Aggarwal, C.C.; Hinneburg, A.; Keim, D.A. *On the Surprising Behavior of Distance Metrics in High Dimensional Space*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 420–434.
33. Zhao, H.; Wang, Z.; Nie, F. A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 629–640. [[CrossRef](#)]
34. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation forest. In Proceedings of the IEEE International Conference on Data Mining, ICDM, Pisa, Italy, 15–19 December 2008; pp. 413–422. [[CrossRef](#)]
35. Hadi, A.S.; Imon, A.H.M.R.; Werner, M. Detection of outliers. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 57–70. [[CrossRef](#)]
36. Xie, C.; Du, R.; Ho, J.W.K.; Pang, H.H.; Chiu, K.W.H.; Lee, E.Y.P.; Vardhanabhuti, V. Effect of machine learning re-sampling techniques for imbalanced datasets in 18F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**. [[CrossRef](#)] [[PubMed](#)]

37. Cysouw, M.C.F.; Jansen, B.H.E.; van de Brug, T.; Oprea-Lager, D.E.; Pfaehler, E.; de Vries, B.M.; van Moorselaar, R.J.A.; Hoekstra, O.S.; Vis, A.N.; Boellaard, R. Machine learning-based analysis of [18F] DCFPyL PET radiomics for risk stratification in primary prostate cancer. *Eur. J. Nucl. Med. Mol. Imaging* **2020**. [[CrossRef](#)]
38. Leithner, D.; Baltzer, P.A.; Magometschnigg, H.F.; Wengert, G.J.; Karanikas, G.; Helbich, T.H.; Weber, M.; Wadsak, W.; Pinker, K. Quantitative assessment of breast parenchymal uptake on 18F-FDG PET/CT: Correlation with age, background parenchymal enhancement, and amount of fibroglandular tissue on MRI. *J. Nucl. Med.* **2016**, *57*, 1518–1522. [[CrossRef](#)]
39. Hoon Tan, P.; Ellis, I.; Allison, K.; Brogi, E.; Fox, S.B.; Lakhani, S.; Lazar, A.J.; Morris, E.A.; Sahin, A.; Salgado, R.; et al. The 2019 WHO classification of tumours of the breast. *Histopathology* **2020**. [[CrossRef](#)]
40. Elston, C.W.; Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* **1991**, *19*, 403–410. [[CrossRef](#)]
41. Perry, N.; Broeders, M.; de Wolf, C.; Törnberg, S.; Holland, R.; von Karsa, L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition—Summary document. *Ann. Oncol.* **2008**, *19*, 614–622. [[CrossRef](#)]
42. Allison, K.H.; Hammond, M.E.H.; Dowsett, M.; McKernin, S.E.; Carey, L.A.; Fitzgibbons, P.L.; Hayes, D.F.; Lakhani, S.R.; Chavez-MacGregor, M.; Perlmutter, J.; et al. Estrogen and Progesterone Receptor Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Guideline Update. *Arch. Pathol. Lab. Med.* **2020**. [[CrossRef](#)] [[PubMed](#)]
43. Wolff, A.C.; Hammond, M.E.H.; Allison, K.H.; Harvey, B.E.; Mangu, P.B.; Bartlett, J.M.S.; Bilous, M.; Ellis, I.O.; Fitzgibbons, P.; Hanna, W.; et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline focused update. *J. Clin. Oncol.* **2018**, *36*, 2105–2122. [[CrossRef](#)]
44. Knäusel, B.; Hirtl, A.; Dobrozemsky, G.; Bergmann, H.; Kletter, K.; Dudczak, R.; Georg, D. PET based volume segmentation with emphasis on the iterative TrueX algorithm. *Z. Med. Phys.* **2012**, *22*, 29–39. [[CrossRef](#)] [[PubMed](#)]
45. Rapisarda, E.; Bettinardi, V.; Thielemans, K.; Gilardi, M.C. Image-based point spread function implementation in a fully 3D OSEM reconstruction algorithm for PET. *Phys. Med. Biol.* **2010**, *55*, 4131–4151. [[CrossRef](#)] [[PubMed](#)]
46. Hofheinz, F.; Hoff, J.V.D.; Steffen, I.G.; Lougovski, A.; Ego, K.; Amthauer, H.; Apostolova, I. Comparative evaluation of SUV, tumor-to-blood standard uptake ratio (SUR), and dual time point measurements for assessment of the metabolic uptake rate in FDG PET. *EJNMMI Res.* **2016**, *6*. [[CrossRef](#)] [[PubMed](#)]
47. Nioche, C.; Orhac, F.; Boughdad, S.; Reuzé, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.; Buvat, I. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [[CrossRef](#)]
48. Gatta, R.; Depeursinge, A.; Ratib, O.; Michielin, O.; Leimgruber, A. Integrating radiomics into holomics for personalised oncology: From algorithms to bedside. *Eur. Radiol. Exp.* **2020**, *4*, 1–9. [[CrossRef](#)] [[PubMed](#)]
49. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)]
50. Papp, L.; Spielvogel, C.P.; Grubmüller, B.; Grahovac, M.; Krajnc, D.; Ecsedi, B.; Sareshgi, R.A.M.; Mohamad, D.; Hamboeck, M.; Rausch, I.; et al. Supervised machine learning enables non-invasive lesion characterization in primary prostate cancer with [68Ga]Ga-PSMA-11 PET/MRI. *Eur. J. Nucl. Med. Mol. Imaging* **2020**. [[CrossRef](#)]
51. Papp, L.; Pötsch, N.; Grahovac, M.; Schmidbauer, V.; Woehrer, A.; Preusser, M.; Mitterhauser, M.; Kiesel, B.; Wadsak, W.; Beyer, T.; et al. Glioma Survival Prediction with Combined Analysis of In Vivo 11 C-MET PET Features, Ex Vivo Features, and Patient Features by Supervised Machine Learning. *J. Nucl. Med.* **2018**, *59*, 892–899. [[CrossRef](#)]
52. Lacroix, M.; Frouin, F.; Dirand, A.-S.; Nioche, C.; Orhac, F.; Bernaudin, J.-F.; Brillet, P.-Y.; Buvat, I. Correction for Magnetic Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential of MR Radiomic Features in Lung Cancer. *Front. Oncol.* **2020**, *10*. [[CrossRef](#)]
53. Elhassan, T.; Aljurf, M.; Al-Mohanna, F.; Shoukri, M.M. Classification of Imbalance Data Using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as A Data Reduction Method Sampling-based Methods Basic Sampling Methods. *J. Inform. Data Min.* **2016**, *1*, 1–12.
54. Marcano-Cedeno, A.; Quintanilla-Dominguez, J.; Cortina-Januchs, M.G.; Andina, D. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In Proceedings of the IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; Institute of Electrical and Electronics Engineers (IEEE): Los Alamitos, CA, USA, 2010; pp. 2845–2850. [[CrossRef](#)]
55. Vanaja, S.; Kumar, K.R. Analysis of Feature Selection Algorithms on Classification: A Survey. *Int. J. Comput. Appl.* **2014**, *96*, 29–35. [[CrossRef](#)]
56. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
57. Nguyen, H.M.; Cooper, E.W.; Kamei, K. Borderline over-sampling for imbalanced data classification. *Int. J. Knowl. Eng. Soft Data Paradig.* **2011**, *3*, 4–21. [[CrossRef](#)]
58. Barua, S.; Islam, M.M.; Yao, X.; Murase, K. MWMOTE—Majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 405–425. [[CrossRef](#)]
59. van Timmeren, J.E.; Leijenaar, R.T.H.; van Elmpt, W.; Reymen, B.; Oberije, C.; Monshouwer, R.; Bussink, J.; Brink, C.; Hansen, O.; Lambin, P. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiother. Oncol.* **2017**, *123*, 363–369. [[CrossRef](#)]

60. Leithner, D.; Mayerhoefer, M.E.; Martinez, D.F.; Jochelson, M.S.; Morris, E.A.; Thakur, S.B.; Pinker, K. Non-Invasive Assessment of Breast Cancer Molecular Subtypes with Multiparametric Magnetic Resonance Imaging Radiomics. *J. Clin. Med.* **2020**, *9*, 1853. [[CrossRef](#)] [[PubMed](#)]
61. Leithner, D.; Horvat, J.V.; Marino, M.A.; Bernard-Davila, B.; Jochelson, M.S.; Ochoa-Albiztegui, R.E.; Martinez, D.F.; Morris, E.A.; Thakur, S.; Pinker, K. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: Initial results. *Breast Cancer Res.* **2019**, *21*, 106. [[CrossRef](#)]
62. Leithner, D.; Bernard-Davila, B.; Martinez, D.F.; Horvat, J.V.; Jochelson, M.S.; Marino, M.A.; Avendano, D.; Ochoa-Albiztegui, R.E.; Sutton, E.J.; Morris, E.A.; et al. Radiomic Signatures Derived from Diffusion-Weighted Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes. *Mol. Imaging Biol.* **2020**, *22*, 453–461. [[CrossRef](#)]
63. Beyer, T.; Townsend, D.W.; Brun, T.; Kinahan, P.; Charron, M.; Roddy, R.; Jerin, J.; Young, J.; Byars, L.; Nutt, R. A Combined PET/CT scanner for clinical oncology. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **2000**, *41*, 1369–1379.
64. McGuire, A.; Brown, J.; Malone, C.; McLaughlin, R.; Kerin, M. Effects of Age on the Detection and Management of Breast Cancer. *Cancers* **2015**, *7*, 908–929. [[CrossRef](#)] [[PubMed](#)]
65. Ou, X.; Zhang, J.; Wang, J.; Pang, F.; Wang, Y.; Wei, X.; Ma, X. Radiomics based on 18 F-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. *Cancer Med.* **2020**, *9*, 496–506. [[CrossRef](#)]
66. Huang, S.-Y.; Franc, B.L.; Harnish, R.J.; Liu, G.; Mitra, D.; Copeland, T.P.; Arasu, V.A.; Kornak, J.; Jones, E.F.; Behr, S.C.; et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer* **2018**, *4*, 1–13. [[CrossRef](#)]
67. Zhou, J.; Lu, J.; Gao, C.; Zeng, J.; Zhou, C.; Lai, X.; Cai, W.; Xu, M. Predicting the response to neoadjuvant chemotherapy for breast cancer: Wavelet transforming radiomics in MRI. *BMC Cancer* **2020**, *20*, 100. [[CrossRef](#)]
68. Xiong, H.; Gaurav, P.; Steinbach, M.; Vipin, K. Enhancing data analysis with noise removal. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 304–319. [[CrossRef](#)]
69. Nazari, Z.; Nazari, M.; Sayed, M.; Danish, S. Evaluation of Class Noise Impact on Performance of Machine Learning Algorithms. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2018**, *18*, 149.
70. Zhu, X.; Wu, X. Class Noise vs. Attribute Noise: A Quantitative Study of Their Impacts. *Artif. Intell. Rev.* **2004**, *22*, 177–210. [[CrossRef](#)]
71. Moy, L.; Noz, M.E.; Maguire, G.Q.; Ponzo, F.; Deans, A.E.; Murphy-Walcott, A.D.; Kramer, E.L. Prone MammoPET Acquisition Improves the Ability to Fuse MRI and PET Breast Scans. *Clin. Nucl. Med.* **2007**, *32*, 194–198. [[CrossRef](#)]
72. Imbriaco, M.; Caprio, M.G.; Limite, G.; Pace, L.; De Falco, T.; Capuano, E.; Salvatore, M. Dual-Time-Point 18 F-FDG PET/CT Versus Dynamic Breast MRI of Suspicious Breast Lesions. *Am. J. Roentgenol.* **2008**, *191*, 1323–1330. [[CrossRef](#)]
73. Misra, S.; Wu, Y. Chapter 10—Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. In *Machine Learning for Subsurface Characterization*; Misra, S., Li, H., He, J., Eds.; Gulf Professional Publishing: Houston, TX, USA, 2020; pp. 289–314.