








Article

A Machine Learning Approach for Breast Cancer Risk Prediction in Digital Mammography [†]

Francesca Angelone ^{1,†}, Alfonso Maria Ponsiglione ^{1,*,†}, Carlo Ricciardi ¹, Maria Paola Belfiore ², Gianluca Gatta ², Roberto Grassi ², Francesco Amato ^{1,§} and Mario Sansone ^{1,§}

¹ Department of Information Technology and Electrical Engineering, University of Naples Federico II, 80125 Naples, Italy; francesca.angelone@unina.it (F.A.); carlo.ricciardi@unina.it (C.R.); framato@unina.it (F.A.); msansone@unina.it (M.S.)

² Department of Precision Medicine Division of Radiology, University of Campania 'Luigi Vanvitelli', 80131 Naples, Italy; mariapaola.belfiore@unicampania.it (M.P.B.); gianluca.gatta@unicampania.it (G.G.); roberto.grassi@unicampania.it (R.G.)

* Correspondence: alfonsonaria.ponsiglione@unina.it

[†] This work extends a previous conference paper presented at the 2022 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering, Rome, Italy, 26–28 October 2022.

[‡] These authors contributed equally to this work.

[§] These authors also contributed equally to this work.

Abstract: Breast cancer is among the most prevalent cancers in the female population globally. Therefore, screening campaigns as well as approaches to identify patients at risk are particularly important for the early detection of suspect lesions. This study aims to propose a workflow for the automatic classification of patients based on one of the most relevant risk factors in breast cancer, which is represented by breast density. The proposed classification methodology takes advantage of the features automatically extracted from mammographic images, as digital mammography represents the major screening tool in women. Textural features were extracted from the breast parenchyma through a radiomics approach, and they were used to train different machine learning algorithms and neural network models to classify the breast density according to the standard Breast Imaging Reporting and Data System (BI-RADS) guidelines. Both binary and multiclass tasks have been carried out and compared in terms of performance metrics. Preliminary results show interesting classification accuracy (93.55% for the binary task and 82.14% for the multiclass task), which are promising compared to the current literature. As the proposed workflow relies on straightforward and computationally efficient algorithms, it could serve as a basis for a fast-track protocol for the screening of mammograms to reduce the radiologists' workload.

Keywords: breast density; digital mammography; artificial intelligence; machine learning; neural networks



Citation: Angelone, F.; Ponsiglione, A.M.; Ricciardi, C.; Belfiore, M.P.; Gatta, G.; Grassi, R.; Amato, F.; Sansone, M. A Machine Learning Approach for Breast Cancer Risk Prediction in Digital Mammography. *Appl. Sci.* **2024**, *14*, 10315. <https://doi.org/10.3390/app142210315>

Academic Editors: Martina Iammarino, Lerina Aversano and Riccardo Pecori

Received: 15 October 2024

Revised: 3 November 2024

Accepted: 6 November 2024

Published: 9 November 2024



Copyright: © 2024 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/>).

1. Introduction

Breast cancer is among the most widespread tumor types in most industrialized countries [1]. However, in recent years, there has been a sharp decline in mortality thanks to the development of diagnostic techniques to diagnose cancer in the early stage, thus ensuring a positive prognosis for the patient.

Full-Field Digital Mammography (FFDM) is the current clinical gold standard in breast cancer screening; indeed, it is widely recognized that proper adherence to mammographic screening protocols in the female population can support the early detection and diagnosis of breast diseases. The United States Preventive Services Task Force recommends that women who are 50 to 74 years old and are at average risk for breast cancer get a mammogram every two years, while women aged 40–49 should talk to their healthcare provider

about when to start and how often to get a mammogram [2–5]. Besides age, there are other different variables correlated with breast cancer risk (e.g., reproductive factors, hormones, diet, metabolism, and breast density) that need to be taken into account for implementing targeted and personalized screening programs.

Among these factors, breast density, which represents the amount of fibrous and glandular tissue in the breast, plays a fundamental role, as women with denser breast tissues can be exposed to a higher risk of developing breast cancer [6,7]. However, it has been shown that the sensitivity of digital mammography is affected by the density of the breast tissue, with values ranging from 86 to 89% in largely fatty breasts and from 62 to 68% in extremely dense breasts. For this reason, the visual inspection of mammograms, despite processing to improve the quality of mammographic images [8–10], does not allow accurate discrimination between dense tissues and tumor lesions, which present similar properties since most tumors absorb X-rays similarly, to an extent, to fibroglandular tissues, as both appear white (dense) on mammograms [11–13]. Breast density represents the amount of fibroglandular tissue compared to fatty tissue and, on a mammography report, is assigned on a qualitative basis, through a visual evaluation, by the radiologists. The classification is based on the international standard BI-RADS (Breast Imaging—Reporting and Data System), based on which it is possible to identify four density classes [14]: (i) class A, predominantly fibro-adipose tissue; (ii) class B, scattered areas of fibroglandular tissue; (iii) class C, heterogeneously dense tissue; and (iv) class D, extremely dense tissue.

To overcome the inter- and intra-observer variability to which visual evaluations are subject, in recent years, many methods of automatic detection on bio-images have been developed, through the extraction of radiomic features [15] combined with artificial intelligence (AI) techniques, for different clinical purposes [16–22]. In this study, the extracted features will represent the textural characteristics of the breast parenchyma [23–25]. Radiomics for the study of breast parenchyma is widely investigated in the literature not only to evaluate risk factors such as breast density but also for the early diagnosis of different types of breast cancer [26–28]. Furthermore, the analysis is extended to different imaging modalities [29], from magnetic resonance (MR) imaging [30] to computed tomography (CT) [17,31] to ultrasound (US) [32] and digital tomosynthesis [33], but to date, the gold standard for the early diagnosis of breast cancer is considered digital mammography. The possibility of correlating the extracted features with breast density is then investigated. The first preliminary step, in Medio-Lateral-Oblique (MLO) images, is to segment the breast parenchyma with respect to the background but above all to the pectoral muscle, which can be interpreted as fibroglandular tissue, if not correctly segmented, leading to an overestimation of the breast density. Extending the results previously presented by the authors [23], this study is primarily intended to evaluate automatic segmentation methods to remove pectoral muscle from MLO mammographic images among those present in the scientific literature. Next, it is meant to test the capability of machine learning and neural networks in classifying breast density based on textural features extracted from only breast parenchyma. Therefore, the main novelties and advances compared to the previous work are (i) the evaluation of automatic segmentation methods to remove pectoral muscle from mediolateral oblique (MLO) mammographic images; (ii) the introduction of a multiclass classification based on BI-RADS guidelines; and (iii) the test of the capability of machine learning and neural networks in classifying breast density on imbalanced dataset.

2. Materials and Methods

The methodological workflow followed in this study is shown in Figure 1.

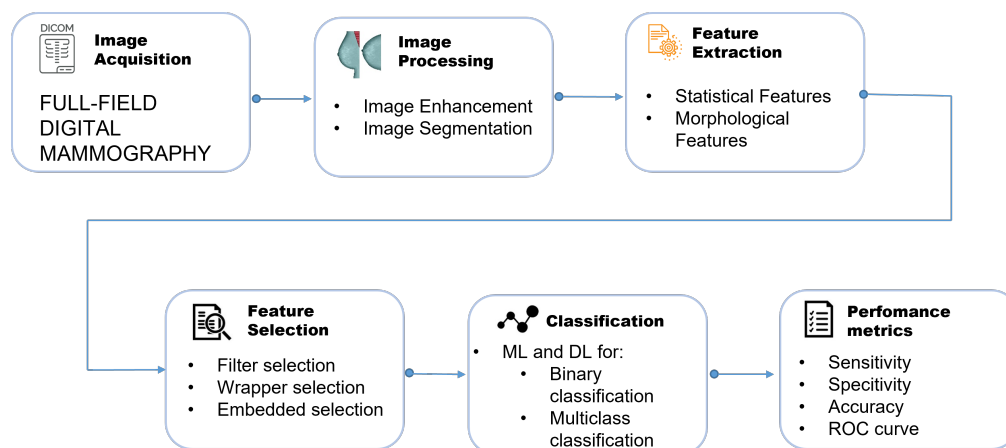


Figure 1. Operational workflow.

2.1. Dataset

The dataset includes images from 161 women patients, whose characteristics are summarized in Table 1, who underwent bilateral, two-view (both cranio-caudal, CC, and MLO) FFDM under the standard protocol, using Giotto Class (IMS GIOTTO S.p.A., Sasso Marconi, Bologna, Italy). The patients considered underwent screening mammography examinations at “Luigi Vanvitelli” University Hospital.

Table 1. Study population.

Number of women	161
Age (mean \pm SD)	56.2 \pm 9.0
Age at first menstrual period (mean \pm SD)	12.0 \pm 1.5
Women in menopause	116
Age menopause (mean \pm SD)	49.8 \pm 5.0
BMI (mean \pm SD)	25.0 kg/mm ² \pm 4.0

For the extraction of the quantitative features, only the “FOR PROCESSING” images were taken into consideration, i.e., raw images, as the intensities detected are proportional to the attenuation of the X-rays on the breast tissue, thus being more suitable for quantitative analysis and eliminating any bias due to post-processing software. Four images for each patient were obtained. Image processing and training and testing of the AI algorithms were performed via a laptop with Intel Core i5-9300H CPU 2.40 GHz and 8 GB RAM and NVIDIA GeForce GTX 1650 Graphic Card.

2.2. Segmentation

To segment the breast, the interface of the air–tissue and the breast–pectoral muscle had to be identified. For segmentation, there are various strategies (from thresholding to more sophisticated algorithms). Accurate segmentation is an important preliminary step to extract quantitative features from the breast parenchyma and be sure of not having an overestimation of the density due to an inaccurate segmentation of the pectoral muscle, which, by an automatic classification algorithm, would be classified as fibroglandular tissue. In this study, it was therefore decided to validate two software tools for automatic segmentation of the breast [34]: OpenBreast (<https://github.com/spertuz/openbreast>, accessed on 15 September 2024) [35] and Libra (<https://www.med.upenn.edu/sbia/libra.html>, accessed on 15 September 2024) [36]. The objective of the two software is different: in the first case, a risk score is produced, and in the second case, only a density assessment. Therefore, only the first part relating to breast segmentation alone was considered in this study. Manual segmentation by expert radiologists represents the ground truth. The

most accurate segmentation method was evaluated on the basis of the calculation of the parameters Dice index and Cohen's kappa.

2.3. Textural Features

Having established the best segmentation software among those previously described, we moved on to the extraction of the textural features. When attributing features to an image, it is essential to examine its pixel-by-pixel structure and consider the interactions among different pixels as observed through the lens of grayscale intensity. Each region of a given image can be depicted by both external and internal features, representing the characteristics of edges and the constituent pixels of the region, respectively. In the case of FFDM images, the primary objective is to extract information related to pixel intensity. Texture analysis, lacking a formal definition, relies on intuitive descriptions for features such as smoothness, coarseness, and regularity. All these properties were related to the intensity of the pixels present within a certain region or of the entire image, highlighting areas of non-uniformity. Therefore, the estimation and classification of different textures were performed thanks to the texture analysis, which consists of the extraction of textural features. Texture analysis has been applied in the analysis of digital images in many fields, enjoying particular success in the field of radiomics [15]. The literature offers a large number of features that can be extracted from an entire image or by appropriate ROIs. "Texture detection" means a process that separates, finding edges, and areas with different textures on the image; the number of features taken into consideration affects how accurately the distinction is made between pixels, and therefore, with it, existing diversities can be recognized in different breasts. As already mentioned, the extraction methods for texture analysis can be divided into two types:

- Statistical/stochastic approach;
- Structural approach.

The first, as the name suggests, treats textures as statistical phenomena, i.e., the formation of a texture on the image is described by statistical properties such as pixel intensity and location. The level of intensity at a certain point of the image is strongly dependent on the intensities of nearby points, unless the image is not made up only of random noise [37]. Statistics based on co-occurrence matrices and histograms [38,39] are the simplest examples of the statistical measurements of textural features. The second category, which is based on a structural approach, introduces the concept of texel, which is considered the fundamental unit of the map textures. This means that with this second approach, more stress is placed on the structure, the spatial dimension of a texture, considering textures as vectors of texels that make up the texture space, just as we consider the image to consist of vectors of pixels. In any case, each texture contains both structural and statistical characteristics; therefore, in this study, both approaches were considered in parallel [40]. The features to be extracted from the breast, chosen based on the literature by the most recent studies [41], can be summarized as follows:

- 1–112 features of Haralick;
- 113–157 features of law;
- 158–185 features run length;
- 186–215 features of wavelet;
- 216–227 features of the histogram;
- 228 fractal dimension;
- 229 local binary pattern.

Then, a total of 229 statistical and morphological features describing the texture of the region of interest were extracted.

2.4. Pre-Processing Dataset and Features Selection

Different data visualization tools were used to study the datasets. In particular, it was found that the features are located in different ranges. Therefore, to be treated equally, it was necessary to report them all in the same range. Then, the normalization of the dataset was

performed. Before selecting the features, the correlation of the features was evaluated using a correlation matrix, and through the visualization of the boxplots in relation to the feature density categories, the most discriminating features were also highlighted. There was, therefore, a first reduction in the dataset. In particular, it was noted that the characteristics of wavelets are not very discriminating, while Haralick characteristics turn out to be strongly correlated with each other. There was, therefore, an initial reduction to 187 total features. Feature selection is an indispensable step for selecting only non-redundant and discriminating features.

To eliminate redundant features, this study employed a filter feature selection. Wrapper feature selection treats the task of choosing an appropriate set of features as a search problem, exploring different combinations of features. On the other hand, filter feature selection utilizes statistical measures to assign a specific score to each feature. In the first scenario, the features are categorized based on their scores, and in accordance with the classification, they are either selected or excluded from the dataset. In the second scenario, a prediction model evaluates each combination using various techniques. In this study, a correlation-based filter method for feature selection is chosen because, being not dependent on any machine learning algorithm implemented and given the unbalance of the dataset in multiclass classification, it appears as the best choice for eliminating the non-discriminative features.

2.5. Machine Learning Algorithms and Neural Networks

For classification, different machine learning models were tested: Support Vector Machine (SVM) [42], Decision Tree [43] using the recursive partition (rpart) function [44], neural network (Nnet) [45], linear discriminant analysis (LDA) [46], and random forest (RF) [47]. To tune the hyperparameters of the defined models, a grid-search method [48] was used to test all possible combinations of the specific hyperparameters for each model. All the ML analysis were performed in R. Finally, neural networks were used for the multiclass classification problem, to evaluate its performance, even on a small dataset. The network used in this work presents a relu activation function of the layers, an optimization function rmsprop, and the evaluation is based on accuracy. The division into testing and training was made using 80% of the data for training and 20% for testing. The training of the models was performed using a batch size equal to 1 (having a reduced dataset available is not necessary to further divide the training set) and the insertion of a callback that allows to store the model that has layer weights such that the maximum accuracy is reached. An empirical evaluation was made on the number of epochs.

2.6. Dataset Imbalance

The classes (A, B, C, D) of the dataset are distributed as follows among the 161 patients: 16 class A; 78 class B; 34 class C; and 33 class D.

A dataset thus constituted is said to be unbalanced, as the patients are not equally distributed in the classes to which they belong, and this significantly influences the overall accuracy of the machine learning algorithms to be applied to the data. Therefore, the possibility of generating a more balanced dataset was investigated. The most intuitive solution was to transform into a binary classification problem by associating classes A and B as “non-dense breasts” and classes C and D as “dense-breasts” as performed in previous works [23,24,49]. Subsequently, a multiclass classification was carried out, and oversampling methods were applied to overcome the problem of the unbalanced dataset. ROSE techniques (Random Over-Sampling Example), through the *ovun.sample* function of the ROSE library in R, and SMOTE (Synthetic Minority Over-sampling TEchnique), through the *SMOTE* function of the *DMwR* library in R, were used.

2.7. Performance Metrics

The overall performance of the multiclass classification was evaluated in terms of accuracy, with its 95% confidence interval (CI) and Cohen's kappa. The classification evaluation performance for the individual four-density classes was evaluated in terms of sensitivity and specificity. In particular, sensitivity is the portion of true positives correctly classified, specificity is the portion of true negatives correctly classified, accuracy is the parameter that summarizes how well the classification went, evaluating the proportion of correct predictions (both positive and negative) on the total number of cases examined while the Cohen's kappa coefficient is calculated by comparing the observed agreement (true positives and true negatives) between the model inferences and the ground truths with the randomly expected classifications based on the marginal frequencies of each class. The latter is a particularly robust metric particularly when there is a significant class imbalance as in the present case.

3. Results

3.1. Segmentation Results

An overlay of the ground truth mask with the masks respectively extracted with Openbreast and Libra software is shown in Figure 2. The areas in green are those which, despite being breasts (based on the ground truth), are considered pectoral muscle by the software; vice versa, the areas in pink are those which, despite being pectoral muscle (based on the ground truth), are considered breast by the software. Consequently, the green areas represent false negatives, and the pink areas represent false positives. A better visualization of the segmentation can be made by showing, as in Figure 3, only the segmentation edges not on the masks but on the original breast images. Cohen's kappa and Dice indices were calculated for a quantitative evaluation. Thus distinguishing the case MLO and CC, it is possible to show in Figures 4 and 5 the histogram of occurrence of the Dice indices. From the histograms, it can already be seen that in the CC configuration, the Dice indices are more or less the same for both software, while instead of the MLO configuration, two different situations are envisaged: the histogram of the indices of comparison between ground truth and Openbreast appears more dispersed and has index values smaller on average than the histogram of the indices extracted from the comparison between the ground truth and Libra. This observation is even clearer if we consider the boxplots in Figure 6. It can therefore be concluded that the similarity index with the ground truth is better for the segmentations performed with the Libra software, certifying the average value of the Dice index at 0.98. A further evaluation is made based on Cohen's kappa. By comparing the two software tools in question, i.e., Libra and Openbreast, the histogram of the occurrences of Cohen's kappa is evaluated to see the index of similarity between the two. In Figure 7, it can be seen that the degree of agreement in the CC configuration is approximately $k = 1$, i.e., the two software create masks that are practically identical. This occurs, as CC images do not suffer segmentation of the pectoral muscle. The MLO images instead have a more dispersed kappa value but still show an excellent degree of agreement. Of more interest is the comparison that is made respectively between the software tools and the ground truth. The histograms Figures 8 and 9 show indices of agreement with kappa high for both comparisons. Finally, we can show the boxplots in Figure 10, based on which we can conclude that the highest k , i.e., the highest level of agreement between the truth and classification, is that of the segmentations obtained using the Libra software. Then, the masks obtained with the Libra software were used for the extraction of textural features.

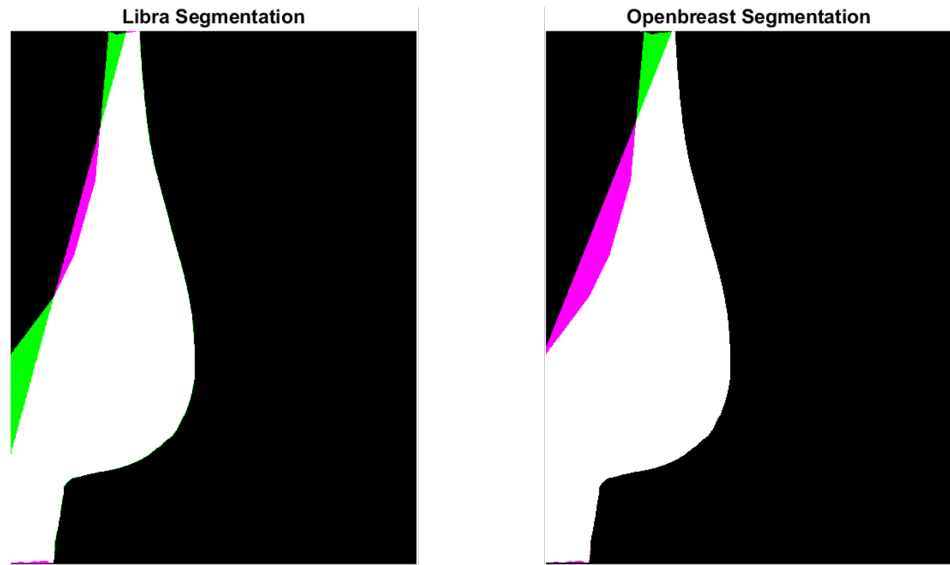


Figure 2. Overlay of the ground truth mask with the masks respectively extracted with Openbreast and Libra software.

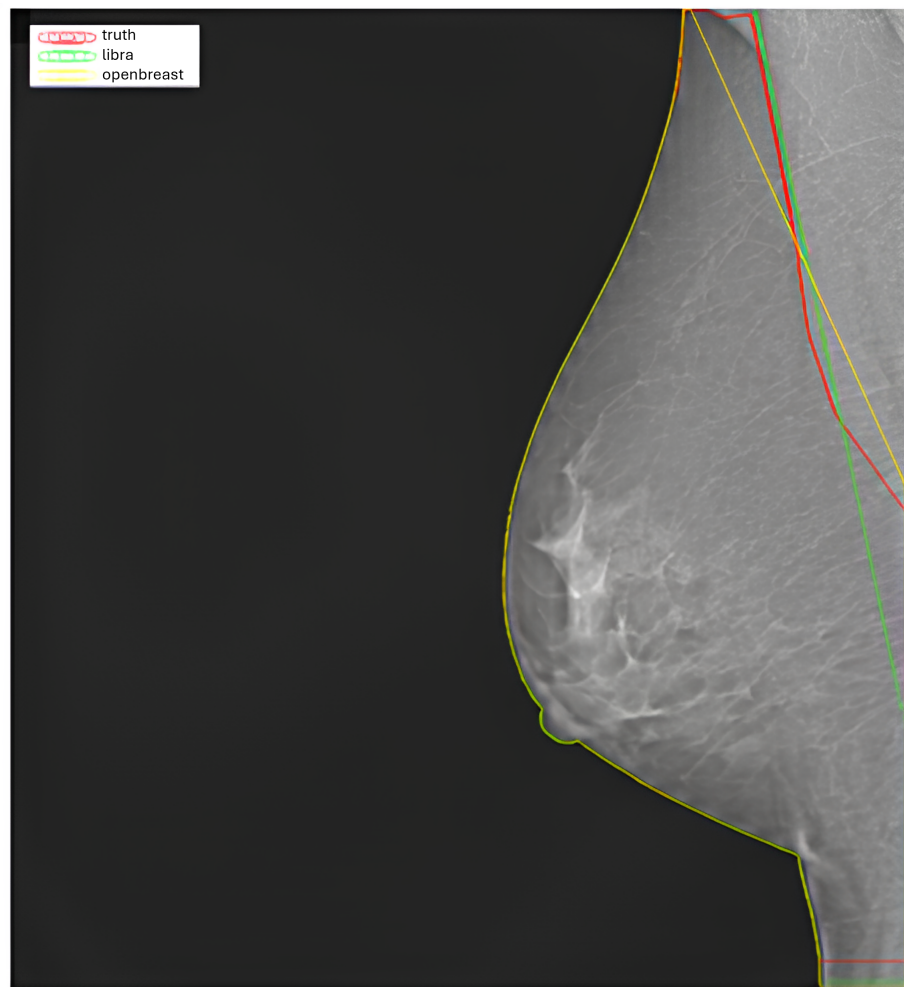


Figure 3. Segmentation edges on the original breast images.

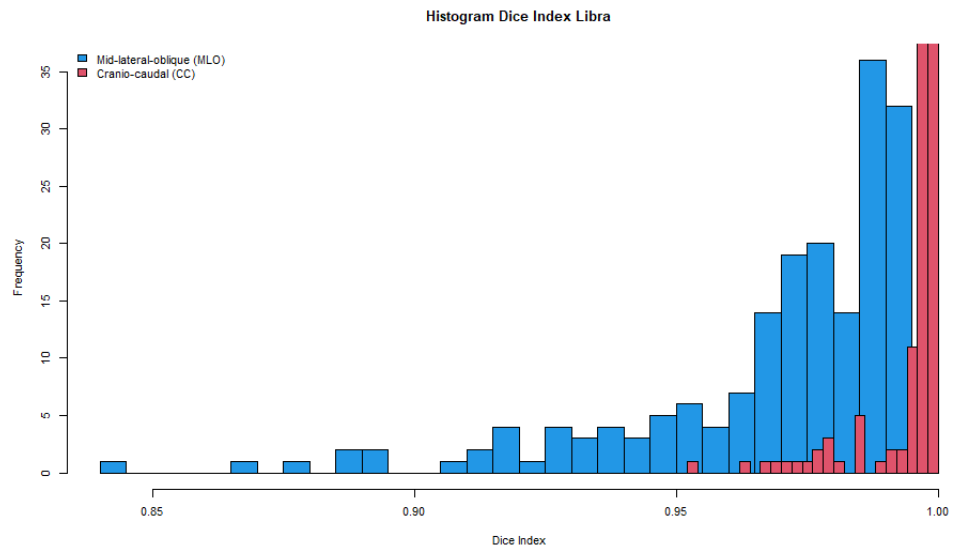


Figure 4. Dice indices distribution histogram for mid-lateral oblique (MLO) and cranio-caudal (CC) configuration breast images using Libra software.

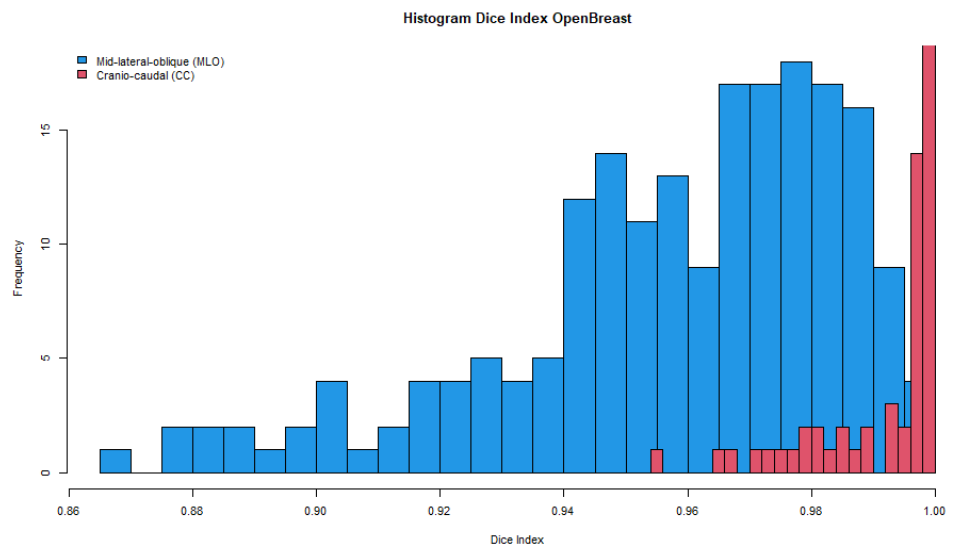


Figure 5. Dice indices distribution histogram for mid-lateral oblique (MLO) and cranio-caudal (CC) configuration breast images using OpenBreast software.

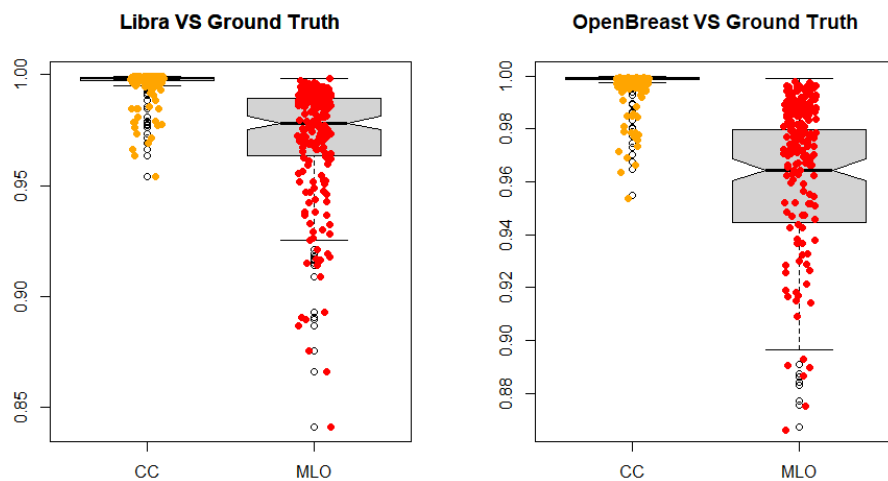


Figure 6. Box plots of Dice indices for CC and MLO images.

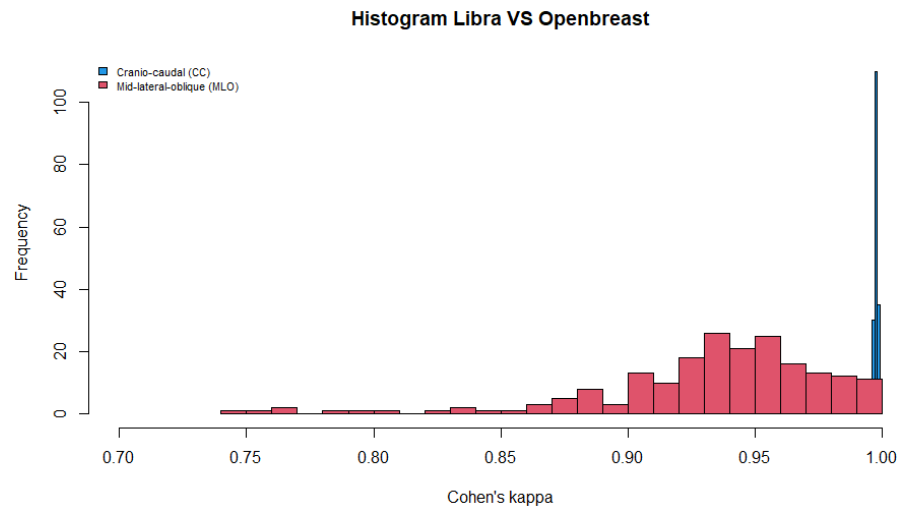


Figure 7. Cohen’s kappa distribution histogram for mid-lateral oblique (MLO) and cranio-caudal (CC) configuration breast images in the comparison between Libra and OpenBreast.

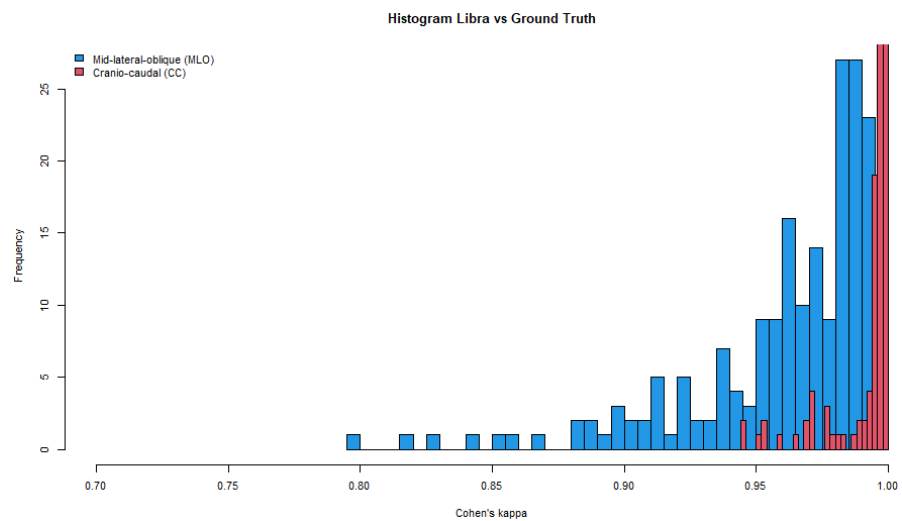


Figure 8. Cohen’s kappa distribution histogram for mid-lateral oblique (MLO) and cranio-caudal (CC) configuration breast images in the comparison between Libra and Ground Truth.

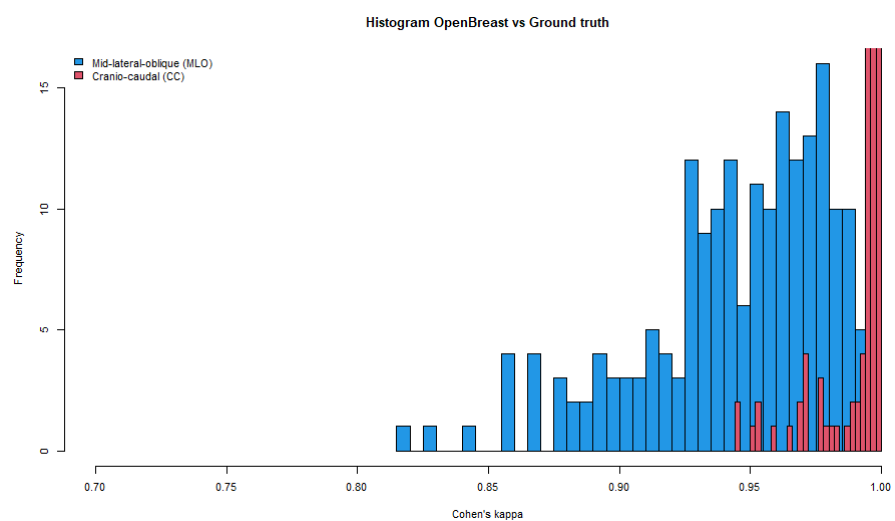


Figure 9. Cohen’s kappa distribution histogram for mid-lateral oblique (MLO) and cranio-caudal (CC) configuration breast images in the comparison between OpenBreast and Ground Truth.

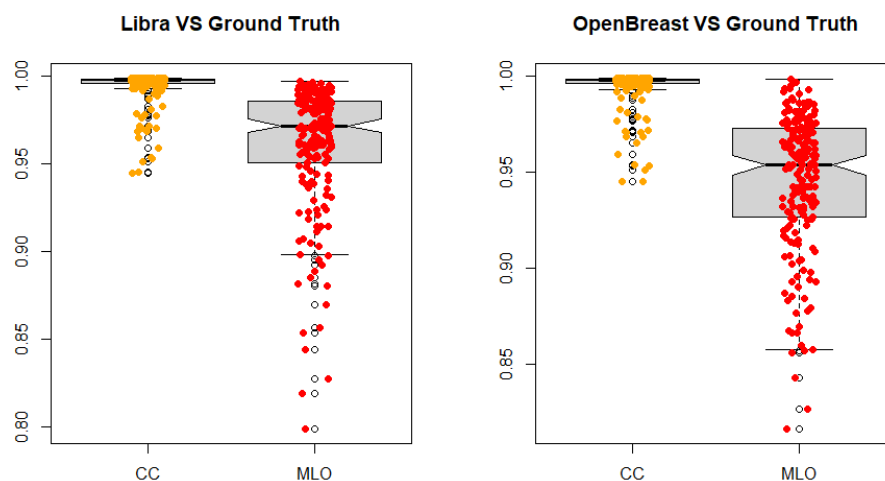


Figure 10. Box plots of Cohen's Kappa indices for CC and MLO images.

3.2. Machine Learning

The results of the classification using ML algorithms relating are shown below. In particular, to solve the unbalanced dataset problem, we tested two different methods and we show the results in each of them. The transformation into a binary classification problem has already been addressed in previous work with satisfactory results [23,24].

Oversampling Techniques

One of the proposed solutions to solve the imbalance problem is to apply oversampling techniques such as ROSE or SMOTE on the datasets. For the feature selection, the redundant features were initially removed with the use of the correlation matrix. As with the other selection techniques, some features fail to discriminate on an unbalanced dataset since they are based on the use of a classifier. Therefore, the features selected are the 35 features obtained by applying the filter features selection methods. SMOTE was applied on a dataset consisting of the largest class, that is, B, and the least numerous class, that is, A, generating an oversampling of class A and an undersampling of class B in order to obtain a number of elements per class, comparable with classes C and D. The results of the classification by applying SMOTE can be observed in Table 2, for the overall accuracy, and in Table 3, for the accuracy relating to the individual classes. As can be seen from these results, the portion of TPP and TNP for the class that was less numerous in the starting dataset (class A) has higher values, while class B exhibits a lower value, which means that it improves the accuracy of class A, at the expense of class B; moreover, classes C and D do not seem to show improvements related to the fact that the dataset is balanced. As a result, the overall accuracy is not improved.

Once ROSE is applied, all classes are oversampled to match the number of elements of the largest class, which is B. The results of the classification by applying ROSE can be observed in Table 2 for the overall accuracy, and in Table 3 for the accuracy relating to the individual classes. In this case, it can be seen that although class A was increased in the number of samples, as well as classes C and D, this does not allow the algorithms to acquire new information; therefore, the results in accuracy remain low both on the single class and overall. In conclusion, by applying the ROSE and SMOTE techniques, it can be seen that ROSE does not solve the problem of low accuracy on less numerous classes, while SMOTE increases the accuracy of class A without significantly raising the overall accuracy, whose highest value obtained is equal to 63.33%, with the linear discriminant analysis model.

Finally, the binarization described in [23,24] turns out to be the best solution to solve the unbalancing problem, and this allows reaching a maximum of 93.55% accuracy, obtained with a Support Vector Machine model, substantially improving accuracy compared to multiclass classification on the original dataset.

Table 2. Overall classification evaluation metrics among four density classes applying SMOTE and ROSE oversampling.

	Algorithm	Accuracy (95% CI)	Kappa
SMOTE	SVM	0.43 (0.25–0.62)	0.27
	Rpart tree	0.33 (0.17–0.53)	0.11
	Nnet	0.50 (0.31–0.69)	0.32
	LDA	0.63 (0.44–0.80)	0.49
	RF	0.50 (0.31–0.69)	0.33
ROSE	SVM	0.37 (0.20–0.56)	0.13
	Rpart tree	0.50 (0.31–0.69)	0.23
	Nnet	0.57 (0.37–0.54)	0.36
	LDA	0.47 (0.28–0.66)	0.29
	RF	0.53 (0.34–0.72)	0.25

Table 3. Classification evaluation metrics relating individual four density classes applying SMOTE and ROSE oversampling.

	Algorithm	Sensitivity				Specificity			
		A	B	C	D	A	B	C	D
SMOTE	SVM	1.00	0.27	0.50	0.50	0.85	0.93	0.67	0.83
	Rpart	0.33	0.20	0.50	0.50	0.85	0.73	0.70	0.79
	Nnet	0.67	0.40	0.67	0.50	0.92	0.87	0.79	0.75
	LDA	1.00	0.53	0.67	0.67	0.96	0.93	0.75	0.87
	RF	1.00	0.33	0.67	0.50	0.89	0.87	0.75	0.83
ROSE	SVM	0.33	0.50	0.33	0.33	0.85	0.73	0.75	0.79
	Rpart	0.00	0.67	0.33	0.50	1.00	0.67	0.79	0.79
	Nnet	0.33	0.67	0.50	0.50	0.92	0.80	0.83	0.83
	LDA	0.67	0.33	0.83	0.33	0.85	0.87	0.67	0.92
	RF	0.00	0.73	0.33	0.50	0.96	0.53	0.92	0.83

3.3. Neural Networks

Since machine learning does not provide good results in multiclass classification, it was thought to apply a neural network to the extracted features. First of all, the comparison between two structures was evaluated: a simple structure composed of only two dense layers and a more complex structure composed of many dense layers. For a network of only two dense layers and 500 nodes, the overfitting was reached after about 50 epochs and reached a maximum accuracy around epoch 30. Instead, for a network of six dense layers, after 60 epochs, the overfitting was not yet reached, but the epoch that performed best was around iteration 55. Based on the results obtained, it makes no sense to prefer a network with many layers; therefore, we tried to reduce the number of nodes of a structure made of only two layers. In particular, for a number of nodes equal to 250 for a single-layer network, it can be seen that after epoch 50, the overfitting on the data started testing, and that before this iteration, the maximum accuracy level was reached. A comparison of the overall accuracies of two methods with the same number of layers but a different number of nodes is shown in Table 4. The values of specificity, sensitivity, and accuracy of the two methods for the different classes is shown in detail in Table 5. Based on the results obtained, it can be concluded that better accuracy is achieved for a model with only two layers with 500 and 4 nodes respectively.

Table 4. Overall accuracies of two dense layers networks with different numbers of nodes.

<i>Overall Statistics</i>	<i>Accuracy (95% CI)</i>	<i>Kappa</i>
2 layers/500 nodes	0.82 (0.63–0.93)	0.71
2 layers/250 nodes	0.71 (0.51–0.97)	0.50

Table 5. Evaluations metrics of the two methods for the different classes.

<i>Statistics</i>	<i>Class</i>	<i>Specificity</i>	<i>Sensitivity</i>	<i>Balanced Accuracy</i>
500 nodes	A	0.33	1.00	0.67
	B	0.93	0.84	0.89
	C	0.50	0.96	0.73
	D	1.00	0.91	0.95
250 nodes	A	0.33	0.96	0.65
	B	0.93	0.61	0.77
	C	0.25	1.00	0.62
	D	0.67	0.91	0.79

4. Discussion and Conclusions

The automatic classification of patients based on breast density can represent an important result for carrying out targeted screening programs based on early diagnosis. The study aimed to evaluate the possibility of classifying the breast density on a small dataset of patients, who were undergoing mammography screening programs, thanks to the extraction of textural features from images. The classification of breast density through CAD systems based on textural features extracted from the image has already been investigated in the recent scientific literature. Kriti and Virmani [50] made a binary classification based on textural features obtained through Laws' texture energy, applying principal component analysis (PCA) to reduce the dimensionality of feature vectors and KNN and NN classifiers, obtaining a maximum accuracy of 95%. The NN classifier was also used by Carneiro et al. [51], using textural features based on histograms and Haralick texture descriptors. In this case, a maximum accuracy of 92.9% was obtained when considering only Haralick texture descriptors, which reached 98.95% when histogram features were also considered. A multiclass classification was also made in [52], which resorted to a fuzzy classification based on textural features extracted through the co-occurrence matrix, reaching an accuracy of 84.2%. In all these cases, the dataset imbalance problem was not addressed.

Our results showed an agreement with the radiologist's report depending on the cases: for a binary classification aimed only at the distinction between dense and non-dense breasts, the highest accuracy of 93.55% was obtained with an SVM method, with $k = 0.8675$, and percentage of true positives and negative equal to $TPP = 94.44\%$ and $TNP = 92.31\%$.

For multiclass classification, the best accuracy in terms of OCA (Overall Classifier Accuracy) was equal to 82.14%, with $k = 0.7089$, obtained with a two-layer neural network. This value is considered encouraging compared with the results obtained from the application of machine learning methods for multiclass classification because we followed a deep learning approach. In this case, the maximum accuracy achieved was equal to 60%, with $k = 0.3583$, a value that has not shown improvements with the application of oversampling techniques, which improved the estimate on the minus class significantly, reaching a maximum overall accuracy of 63.33% with $k = 0.4931$. The application of neural networks provided the most encouraging results for acting on unbalanced datasets for multiclass classification, despite the low number of patients.

Author Contributions: Conceptualization, F.A. (Francesco Amato) and M.S.; methodology, A.M.P., F.A. (Francesca Angelone), M.S. and F.A. (Francesco Amato); validation, A.M.P., F.A. (Francesca Angelone), M.S., F.A. (Francesco Amato) and R.G.; formal analysis, A.M.P. and F.A. (Francesca

Angelone); investigation, A.M.P., F.A. (Francesca Angelone) and C.R.; data curation, A.M.P., F.A. (Francesca Angelone), M.P.B. and G.G.; writing—original draft preparation, A.M.P., F.A. (Francesca Angelone), M.S. and F.A. (Francesco Amato); writing—review and editing, A.M.P., F.A. (Francesca Angelone), M.S., F.A. (Francesco Amato), C.R. and R.G.; visualization, A.M.P., F.A. (Francesca Angelone), M.P.B., G.G., R.G. and C.R.; supervision, M.S. and F.A. (Francesco Amato); F.A. (Francesca Angelone) and A.M.P. equally contributed to the work. F.A. (Francesco Amato) and M.S. equally contributed to the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This retrospective study was approved by the ethics committee of the University “Luigi Vanvitelli”, Naples, Italy, with deliberation n. 469 of 23 July 2019 and, informed consent was waived by the ethics committee. All methods were carried out according to National regulations and guidelines.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: This work was carried out as part of a wider project (“Senologia per te”) promoted by the Italian Society of Medical Radiology and it was performed in collaboration with IMS GIOTTO S.p.A. Sasso Marconi (BO), Italy.

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

References

1. GlobalCancerObservatory. Breast Cancer. 2023. Available online: <https://gco.iarc.fr/en> (accessed on 15 September 2024).
2. Ray, K.M.; Joe, B.N.; Freimanis, R.I.; Sickles, E.A.; Hendrick, R.E. Screening mammography in women 40–49 years old: Current evidence. *Am. J. Roentgenol.* **2018**, *210*, 264–270. [[CrossRef](#)] [[PubMed](#)]
3. Pitman, J.A.; McGinty, G.B.; Soman, R.R.; Drotman, M.B.; Reichman, M.B.; Arleo, E.K. Screening mammography for women in their 40s: The potential impact of the American Cancer Society and US Preventive Services Task Force breast cancer screening recommendations. *Am. J. Roentgenol.* **2017**, *209*, 697–702. [[CrossRef](#)]
4. Hendrick, R.E.; Helvie, M.A. United States preventive services task force screening mammography recommendations: Science ignored. *Am. J. Roentgenol.* **2011**, *196*, W112–W116. [[CrossRef](#)] [[PubMed](#)]
5. Newman, L. US Preventive Services Task Force breast cancer recommendation statement on risk assessment, genetic counseling, and genetic testing for BRCA-related cancer. *JAMA Surg.* **2019**, *154*, 895–896. [[CrossRef](#)] [[PubMed](#)]
6. McCormack, V.A.; dos Santos Silva, I. Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. *Cancer Epidemiol. Prev. Biomark.* **2006**, *15*, 1159–1169. [[CrossRef](#)]
7. Boyd, N.F.; Martin, L.J.; Yaffe, M.J.; Minkin, S. Mammographic density and breast cancer risk: Current understanding and future prospects. *Breast Cancer Res.* **2011**, *13*, 223. [[CrossRef](#)]
8. Almalki, Y.E.; Soomro, T.A.; Irfan, M.; Alduraibi, S.K.; Ali, A. Impact of image enhancement module for analysis of mammogram images for diagnostics of breast cancer. *Sensors* **2022**, *22*, 1868. [[CrossRef](#)]
9. Fausto, A.M.; Lopes, M.; de Sousa, M.; Furquim, T.A.; Mol, A.W.; Velasco, F.G. Optimization of image quality and dose in digital mammography. *J. Digit. Imaging* **2017**, *30*, 185–196. [[CrossRef](#)]
10. Angelone, F.; Ponsiglione, A.M.; Grassi, R.; Amato, F.; Sansone, M. A general framework for the assessment of scatter correction techniques in digital mammography. *Biomed. Signal Process. Control* **2024**, *89*, 105802. [[CrossRef](#)]
11. Kerlikowske, K. The mammogram that cried Wolfe. *N. Engl. J. Med.* **2007**, *356*, 297–300. [[CrossRef](#)]
12. Carney, P.A.; Miglioretti, D.L.; Yankaskas, B.C.; Kerlikowske, K.; Rosenberg, R.; Rutter, C.M.; Geller, B.M.; Abraham, L.A.; Taplin, S.H.; Dignan, M.; et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann. Intern. Med.* **2003**, *138*, 168–175. [[CrossRef](#)] [[PubMed](#)]
13. Mann, R.M.; Athanasiou, A.; Baltzer, P.A.; Camps-Herrero, J.; Clauser, P.; Fallenberg, E.M.; Forrai, G.; Fuchsjäger, M.H.; Helbich, T.H.; Killburn-Toppin, F.; et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). *Eur. Radiol.* **2022**, *32*, 4036–4045. [[CrossRef](#)] [[PubMed](#)]
14. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.; 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](#)]
15. van Timmeren, J.E.; Cester, D.; Tanadini-Lang, S.; Alkadhi, H.; Baessler, B. Radiomics in medical imaging—“How-to” guide and critical reflection. *Insights Imaging* **2020**, *11*, 1–16. [[CrossRef](#)] [[PubMed](#)]
16. Angelone, F.; Ciliberti, F.K.; Tobia, G.P.; Jónsson Jr, H.; Ponsiglione, A.M.; Gislason, M.K.; Tortorella, F.; Amato, F.; Gargiulo, P. Innovative Diagnostic Approaches for Predicting Knee Cartilage Degeneration in Osteoarthritis Patients: A Radiomics-Based Study. *Inf. Syst. Front.* **2024**, 1–23. [[CrossRef](#)]

17. Zhu, H.B.; Zhu, H.T.; Jiang, L.; Nie, P.; Hu, J.; Tang, W.; Zhang, X.Y.; Li, X.T.; Yao, Q.; Sun, Y.S. Radiomics analysis from magnetic resonance imaging in predicting the grade of nonfunctioning pancreatic neuroendocrine tumors: A multicenter study. *Eur. Radiol.* **2024**, *34*, 90–102. [[CrossRef](#)]
18. Jin, J.; Jiang, Y.; Zhao, Y.L.; Huang, P.T. Radiomics-based Machine Learning to Predict the Recurrence of Hepatocellular Carcinoma: A systematic review and Meta-analysis. *Acad. Radiol.* **2024**, *31*, 467–479. [[CrossRef](#)]
19. Ye, J.Y.; Fang, P.; Peng, Z.P.; Huang, X.T.; Xie, J.Z.; Yin, X.Y. A radiomics-based interpretable model to predict the pathological grade of pancreatic neuroendocrine tumors. *Eur. Radiol.* **2024**, *34*, 1994–2005. [[CrossRef](#)]
20. Bahl, M. Combining AI and Radiomics to Improve the Accuracy of Breast US. *Radiology* **2024**, *312*, e241795. [[CrossRef](#)]
21. Jiang, W.; Meng, R.; Cheng, Y.; Wang, H.; Han, T.; Qu, N.; Yu, T.; Hou, Y.; Xu, S. Intra-and peritumoral based radiomics for assessment of Lymphovascular invasion in invasive breast cancer. *J. Magn. Reson. Imaging* **2024**, *59*, 613–625. [[CrossRef](#)]
22. Liu, X.; Xiao, W.; Qiao, J.; Luo, Q.; Gao, X.; He, F.; Qin, X. Prediction of lymph node metastasis in endometrial cancer based on color doppler ultrasound radiomics. *Acad. Radiol.* **2024**, *31*, 4499–4508. [[CrossRef](#)] [[PubMed](#)]
23. Angelone, F.; Ricciardi, C.; Gatta, G.; Sansone, M.; Ponsiglione, A.M.; Belfiore, M.P.; Amato, F.; Grassi, R. Breast Density Analysis on Mammograms: Application of Machine Learning with Textural Features. In Proceedings of the 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRaine), Rome, Italy, 26–28 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 295–300.
24. Sansone, M.; Fusco, R.; Grassi, F.; Gatta, G.; Belfiore, M.P.; Angelone, F.; Ricciardi, C.; Ponsiglione, A.M.; Amato, F.; Galdiero, R.; et al. Machine Learning Approaches with Textural Features to Calculate Breast Density on Mammography. *Curr. Oncol.* **2023**, *30*, 839–853. [[CrossRef](#)] [[PubMed](#)]
25. Angelone, F.; Ponsiglione, A.; Belfiore, M.; Gatta, G.; Grassi, R.; Amato, F.; Sansone, M. Evaluation of breast density variability between right and left breasts. In Proceedings of the Convegno Nazionale di Bioingegneria, Padua, Italy, 21–23 June 2023.
26. Zhu, S.; Wang, S.; Guo, S.; Wu, R.; Zhang, J.; Kong, M.; Pan, L.; Gu, Y.; Yu, S. Contrast-enhanced mammography radiomics analysis for preoperative prediction of breast cancer molecular subtypes. *Acad. Radiol.* **2024**, *31*, 2228–2238. [[CrossRef](#)] [[PubMed](#)]
27. Wei, P. Radiomics, deep learning and early diagnosis in oncology. *Emerg. Top. Life Sci.* **2021**, *5*, 829–835. [[CrossRef](#)] [[PubMed](#)]
28. Petrillo, A.; Fusco, R.; Di Bernardo, E.; Petrosino, T.; Barretta, M.L.; Porto, A.; Granata, V.; Di Bonito, M.; Fanizzi, A.; Massafra, R.; et al. Prediction of breast cancer histological outcome by radiomics and artificial intelligence analysis in contrast-enhanced mammography. *Cancers* **2022**, *14*, 2132. [[CrossRef](#)]
29. Acciavatti, R.J.; Lee, S.H.; Reig, B.; Moy, L.; Conant, E.F.; Kontos, D.; Moon, W.K. Beyond breast density: Risk measures for breast cancer in multiple imaging modalities. *Radiology* **2023**, *306*, e222575. [[CrossRef](#)]
30. Jing, X.; Wielema, M.; Monroy-Gonzalez, A.G.; Stams, T.R.; Mahesh, S.V.; Oudkerk, M.; Sijens, P.E.; Dorrius, M.D.; van Ooijen, P.M. Automated breast density assessment in MRI using deep learning and radiomics: Strategies for reducing inter-observer variability. *J. Magn. Reson. Imaging* **2024**, *60*, 80–91. [[CrossRef](#)]
31. Landsmann, A.; Ruppert, C.; Wieler, J.; Hejduk, P.; Ciritsis, A.; Borkowski, K.; Wurnig, M.C.; Rossi, C.; Boss, A. Radiomics in photon-counting dedicated breast CT: Potential of texture analysis for breast density classification. *Eur. Radiol. Exp.* **2022**, *6*, 30. [[CrossRef](#)]
32. Cai, L.; Sidey-Gibbons, C.; Nees, J.; Riedel, F.; Schaeffgen, B.; Togawa, R.; Killinger, K.; Heil, J.; Pfob, A.; Golatta, M. Ultrasound radiomics features to identify patients with triple-negative breast cancer: A retrospective, single-center study. *J. Ultrasound Med.* **2024**, *43*, 467–478. [[CrossRef](#)]
33. Hussain, S.; Lafarga-Osuna, Y.; Ali, M.; Naseem, U.; Ahmed, M.; Tamez-Peña, J.G. Deep learning, radiomics and radiogenomics applications in the digital breast tomosynthesis: A systematic review. *BMC Bioinform.* **2023**, *24*, 401. [[CrossRef](#)]
34. Sansone, M.; Marrone, S.; Di Salvio, G.; Belfiore, M.P.; Gatta, G.; Fusco, R.; Vanore, L.; Zuiani, C.; Grassi, F.; Vietri, M.T.; et al. Comparison between two packages for pectoral muscle removal on mammographic images. *Radiol. Med.* **2022**, *127*, 848–856. [[CrossRef](#)] [[PubMed](#)]
35. Pertuz, S.; Torres, G.F.; Tamimi, R.; Kamarainen, J. Open framework for mammography-based breast cancer risk assessment. In Proceedings of the 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, 19–22 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
36. Zheng, Y.; Keller, B.M.; Ray, S.; Wang, Y.; Conant, E.F.; Gee, J.C.; Kontos, D. Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment. *Med. Phys.* **2015**, *42*, 4149–4160. [[CrossRef](#)] [[PubMed](#)]
37. Woods, J. Two-dimensional discrete Markovian fields. *IEEE Trans. Inf. Theory* **1972**, *18*, 232–240. [[CrossRef](#)]
38. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
39. Unser, M. Sum and difference histograms for texture classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 118–125. [[CrossRef](#)]
40. Sánchez-Yáñez, R.E.; Kurmyshev, E.V.; Cuevas, F.J. A framework for texture classification using the coordinated clusters representation. *Pattern Recognit. Lett.* **2003**, *24*, 21–31. [[CrossRef](#)]
41. Kontos, D.; Winham, S.J.; Oustimov, A.; Pantalone, L.; Hsieh, M.K.; Gastouniotti, A.; Whaley, D.H.; Hruska, C.B.; Kerlikowske, K.; Brandt, K.; et al. Radiomic phenotypes of mammographic parenchymal complexity: Toward augmenting breast density in breast cancer risk assessment. *Radiology* **2019**, *290*, 41–49. [[CrossRef](#)]

42. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]
43. Rokach, L.; Maimon, O. Decision trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2005; pp. 165–192.
44. Therneau, T.M.; Atkinson, E.J. *An Introduction to Recursive Partitioning Using the RPART Routines*; Technical Report; Mayo Foundation: Rochester, MN, USA, 1997.
45. Mehlig, B. *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers*; Cambridge University Press: Cambridge, UK, 2021.
46. Izenman, A.J. Linear discriminant analysis. In *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 237–280.
47. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
48. Liashchynskiy, P.; Liashchynskiy, P. Grid search, random search, genetic algorithm: A big comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.
49. Sacchetto, D.; Morra, L.; Agliozzo, S.; Bernardi, D.; Björklund, T.; Brancato, B.; Bravetti, P.; Carbonaro, L.A.; Correale, L.; Fantò, C.; et al. Mammographic density: Comparison of visual assessment with fully automatic calculation on a multivendor dataset. *Eur. Radiol.* **2016**, *26*, 175–183. [[CrossRef](#)] [[PubMed](#)]
50. Kriti; Virmani, J. Breast density classification using Laws' mask texture features. *Int. J. Biomed. Eng. Technol.* **2015**, *19*, 279–302. [[CrossRef](#)]
51. Carneiro, P.C.; Franco, M.L.N.; Thomaz, R.d.L.; Patrocinio, A.C. Breast density pattern characterization by histogram features and texture descriptors. *Res. Biomed. Eng.* **2017**, *33*, 069–077. [[CrossRef](#)]
52. Valencia-Hernandez, I.; Peregrina-Barreto, H.; Reyes-Garcia, C.; Lopez-Armas, G. Density map and fuzzy classification for breast density by using BI-RADS. *Comput. Methods Programs Biomed.* **2021**, *200*, 105825. [[CrossRef](#)]

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.