Pérez-Benito, FJ.; Signol, F.; Perez-Cortes, J.; Pollán, M.; Perez-Gómez, B.; Salas-Trejo, D.; Casals, M.... (2019). Global parenchymal texture features based on histograms of oriented gradients improve cancer development risk estimation from healthy breasts. Computer Methods and Programs in Biomedicine. 177:123-132. https://doi.org/10.1016/j.cmpb.2019.05.022

Global parenchymal texture features based on histograms of oriented gradients improve cancer development risk estimation from healthy breasts

Francisco Javier Pérez-Benito[a,*], Francois Signol[a], Juan-Carlos Pérez-Cortés[a], Marina Pollán[b,c], Beatriz Pérez-Gómez[b,c], Dolores Salas-Trejo[d,e], María Casals[d,e], Inmaculada Martínez[d,e], Rafael LLobet[a]

[a] *Institute of Computer Technology, Universitat Politècnica de València, Camino de Vera, s/n, 46022 València, Spain*
[b] *National Center for Epidemiology, Carlos III Institute of Health, Monforte de lemos, 5, 28029 Madrid, Spain*
[c] *Consortium for Biomedical Research in Epidemiology and Public Health (CIBER en Epidemiología y Salud Pública - CIBERESP), Carlos III Institute of Health, Monforte de Lemos, 5, 28029 Madrid, Spain*
[d] *Valencian Breast Cancer Screening Program, General Directorate of Public Health, València, Spain*
[e] *Centro Superior de Investigación en Salud Pública CSISP, FISABIO, València, Spain*

Abstract

Background: The breast dense tissue percentage on digital mammographies is one of the most commonly used markers for breast cancer risk estimation.

Abbreviations: AUC: Area under Receiver Operating Characteristic curve; BI-RADS: Breast Imaging Reporting and Data System; BMI: Body Mass index; CC: Cranio-Caudal view; DBT: Digital Breast Tomosynthesis; FGT: Fibroglandular Tissue; G-HOGH Global HOG-based Histogram; GLCM: Gray-level co-ocurrence matrix; HOG: Histogram of Oriented Gradients; MD: Mammographic Density; MLO: Medio Lateral Oblique view; MTR: Mammographic Texture Resemblance; NRI: Net Reclassification Improvement; PD: Percent Density; RF: Random Forest; ROC: Receiver Operating Characteristic

*Corresponding author

Email addresses: frapebe@doctor.upv.es (Francisco Javier Pérez-Benito), fsignol@iti.es (Francois Signol), fcperez@iti.upv.es (Juan-Carlos Pérez-Cortés), mpollan@isciii.es (Marina Pollán), bperez@isciii.es (Beatriz Pérez-Gómez), salas_dol@gva.es (Dolores Salas-Trejo), casals_mar@gva.es (María Casals), martinez_inm@gva.es (Inmaculada Martínez), rllobet@iti.upv.es (Rafael LLobet)

Geometric features of dense tissue over the breast and the presence of texture structures contained in sliding windows that scan the mammographies may improve the predictive ability when combined with the breast dense tissue percentage.

Methods: A case/control study nested within a screening program covering 1563 women with craniocaudal and mediolateral-oblique mammographies (755 controls and the contralateral breast mammographies at the closest checkpoint before cancer diagnostic for 808 cases) aging 45 to 70 from Comunitat Valenciana (Spain) was used to extract geometric and texture features. The dense tissue segmentation was performed using DMScan and validated by two experienced radiologists. A model based on Random Forests was trained several times varying the set of variables. The dataset of 1172 patients was evaluated with a 10-stratified-fold cross-validation scheme. The area under the Receiver Operating Characteristic curve (AUC) was the metric for the predictive ability. The results were assessed by only considering the output after applying the model to the test set, which was composed of the remaining 391 patients.

Results: The AUC score obtained by the dense tissue percentage (0.55) was compared to a machine learning-based classifier results. The classifier, apart from the percentage of dense tissue of both views, firstly included global geometric features such as the distance of dense tissue to the pectoral muscle, dense tissue eccentricity or the dense tissue perimeter, obtaining an accuracy of 0.56. By the inclusion of a global feature based on local histograms of oriented gradients, the accuracy of the classifier was significantly improved (0.61). The number of well-classified patients was improved up to 236 when

it was 208.

Conclusion: Relative geometric features of dense tissue over the breast and histograms of standardized local texture features based on sliding windows scanning the whole breast improve risk prediction beyond the dense tissue percentage adjusted by geometrical variables. Other classifiers could improve the results obtained by the conventional Random Forests used in this study.
Keywords:
Breast density, Texture features, Cancer development risk, Breast Cancer

---

1. Background

Since the last quarter of the 20th century, thanks to Wolfe's work [1], the study of parenchymal breast patterns has become increasingly important in the pursuit predicting the risk of developing breast cancer. Wolfe was one of the main contributors in Mammographic Density (MD) based research. This term is used to quantify the fibroglandular as opposed to fatty tissue in the breast. It is considered to be among the strongest risk factors of breast cancer, [2, 3]. Furthermore, the authors of a recent study [4] have demonstrated that MD is an age-related feature which is highly consistent across diverse groups of women worldwide, suggesting that it results from an intrinsic mechanism common to women.

One way to measure MD is the well-studied percentage density (PD) which is the fibroglandular tissue area FGT_{area} over breast area B_{area} ratio $\frac{FGT_{area}}{B_{area}}$. Boyd et al. in their work [5] concluded that PD is a factor to which a substantial fraction of breast cancers can be attributed. Actually, women with a percentage of dense tissue higher than 75% have four to six times

higher risk of developing the disease than similar aged women with lower density [5–8].

The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) has reported a breast classification based on density, disposition, shape, and granularity of the dense tissue [9] suggesting that geometric and texture patterns may influence in a better prediction of the risk of developing the disease [10, 11]. This classification was probably motivated by the influence of some authors that had already studied percentages of dense tissue in different breast areas [12, 13], assuming that the distance of this tissue to the pectoral muscle could influence the breast cancer development. Previously to the aforementioned classification by BI-RADS, some features based on texture approaches were explored, like the study of gray-level co-occurrence matrices (GLCM) of neighboring pixels [14] or the analysis of the relationship between PD and a range of texture features based on digital mammography and Digital Breast Tomosynthesis (DBT) [15].

Numerous articles have emerged from this new approach, demonstrating the importance and interest from the community to address this challenge by adding new feature extraction methods based on parenchymal textures. Nielsen et al. [16] defined a mammographic texture resemblance (MTR) marker independent on PD which improved the prediction ability by the combination of both features. In the case of Wang et al. [8], the baseline was the estimation ability of a volume-based measure of fibroglandular dense tissue (FGT), which was improved by the addition of gray-level co-occurrence matrix (GLCM) texture features at low mammography resolutions. The lattice-based method proposed by Zheng et al. [17] obtained texture features

after splitting the images in a regular grid; the final features were a summary (mean) of statistical and structural-based measures at different-sized windows. Novel approaches based on deep learning have also been assessed. Among them, we could highlight the work carried out by Kallenberg et al. [18] which included an unsupervised autoencoder method to extract texture features not only for risk scoring estimation but also for automatic breast segmentation. The study of Gastounioti et al. [19] extended Zheng work [17] using convolutional neural networks to lattice-based feature extraction.

The objective of this work is to improve the rate of risk breast cancer estimation. To this end, we attempted to define global texture features from both cranio-caudal (CC) and medio-lateral oblique (MLO) views which summarize local texture structures from healthy breasts images according to the hypothesis that they could improve the estimation of breast cancer risk provided by PD only.

The paper is organized as follows: data description, preprocessing, feature extraction and model development are presented in Section 2. Section 3 shows the experimental results on real images. Impact, contributions, limitations and future work are presented in Section 4. Finally, a short conclusion is drawn in Section 5.

2. Methods

2.1. Setting and study design

A 1:1 case/control study was designed using women recruited from 10 breast cancer medical centers, in and out of the screening program.The acquisition devices for each center are shown in Table 1. Cases without

pathological anatomy report, bilateral tumors or without mammography were excluded from this study. A total of 1563 patients aging from 45 to 70 are paired in case/control couples so that each control was attempted to be associated with a case by age and screening center. This pairing strategy resulted in 808 cases of breast and 755 controls which is slightly unbalanced. Nevertheless, the evaluation was designed to compensate for this imperfect 1-to-1 matching using stratified fold of case/control pairs (see section 3.1). Each patient has both CC and MLO DICOM-formatted mammographies. If available, for each case the contralateral mammography was taken from the checkpoint previous to diagnostic. Otherwise, the contralateral mammography to the one diagnosed with cancer from the closer previous checkpoint was selected; Finally, if no previous mammography exists, then the contralateral mammography at the diagnostic time was extracted. Please, note that although they are called cases, the images come from breasts that are still healthy.

Unit	**Mammography device**
Castellón	FUJIFILM
Fuente de San Luis	FUJIFILM
Alcoi	IMS s.r.l. / Giotto IRE (*)
Xàtiva	FUJIFILM
Requena	HOLOGIC / Giotto IRE (*)
Elda	SIEMENS / Giotto IRE (*)
Elche	FUJIFILM
Orihuela	FUJIFILM
Denia	IMS s.r.l. / Giotto IRE(*)
Serrería	(**)

Table 1: Screening units and their mammography devices. (*) implies the use of a new device [Gioto IRE] since 2015. (**) the device is not known

Mammographies were analyzed using DMScan [20, 21], the semi-automatic thresholding segmentation produced by this tool was validated by two experienced radiologists. The PD's was obtained by the validation of the breast segmentations provided by the two radiologists were highly correlated (0.807 and 0.894 Pearson's correlation coefficients for CC and MLO views respectively). The segmentation applied to each mammography is randomly taken among the two available.

2.2. Data description

The variables which were progressively introduced into the model can be classified as geometrical and textural.

2.2.1. Geometrical variables

In addition to the percentage of dense tissue, 22 geometrical features were computed for each of the two available views of the mammographies. The resolution variability produced by the use of different devices was normalized by converting pixel (px) to millimeters (mm) with the DICOM metadata.

Breast and FGT area were obtained by counting the pixels contained in the breast and FGT binary masks after image segmentation; these masks were also used to compute the breast perimeter. The relation of the FGT area over the breast area defined the PD. FGT morphology was introduced by the calculation of the FGT perimeter, compactness, aspect ratio, and eccentricity after reducing noise by the morphological opening operation [22]. The position in millimeters of the FGT with respect to the breast was obtained by the coordinates of the FGT center of mass taking as origin the breast center of mass. We also characterized FGT distribution among

the breast by the calculation of the PD into the five regions proposed by Blot et al. [12] and Oliver et al. [13]. Lastly, we calculated the seven FGT translation, scale and, rotation invariant moments defined by Hu [23] as geometric features to be included in the study. The imaging preprocessing steps carried out to extract geometrical variables are shown in figure 1.

2.2.2. Texture variables

We have built global texture features based on Histograms of Oriented Gradients (HOG) [24]. They were extracted from the whole breast as a single region by using squared different-sized windows to scan the breast and to calculate the proposed textural features (G-HOGH) as a combination of the local window-based HOG's.

The DICOM images were resized so that each pixel covered 0.05×0.05 mm since this resolution was the most frequent in the dataset. Windows sizes were selected taking into account the image area wanted to be covered, 24×24, 32×32, 64×64 and 128×128 pixel-sized windows were built to analyze areas of 1.44, 2.56, 5.12 and 10.24 mm^2 respectively. Only windows with more than 40% of pixels belonging to the breast were considered.

HOG descriptors represent images taking into account the distribution of the neighboring pixel gradients. The region of interest is divided into blocks, which can overlap, and each block into cells, which are composed of pixels. Each pixel is represented by a magnitude and an angle (m, α) computed from the gradient using Sobel's Operator [25]. Each cell produces a histogram with a predefined number of bins covering the range of possible orientations ($[0, \pi]$ for unsigned gradients or $[0, 2\pi]$ for signed gradients). For example, let's imagine 9 bins covered unsigned gradients.

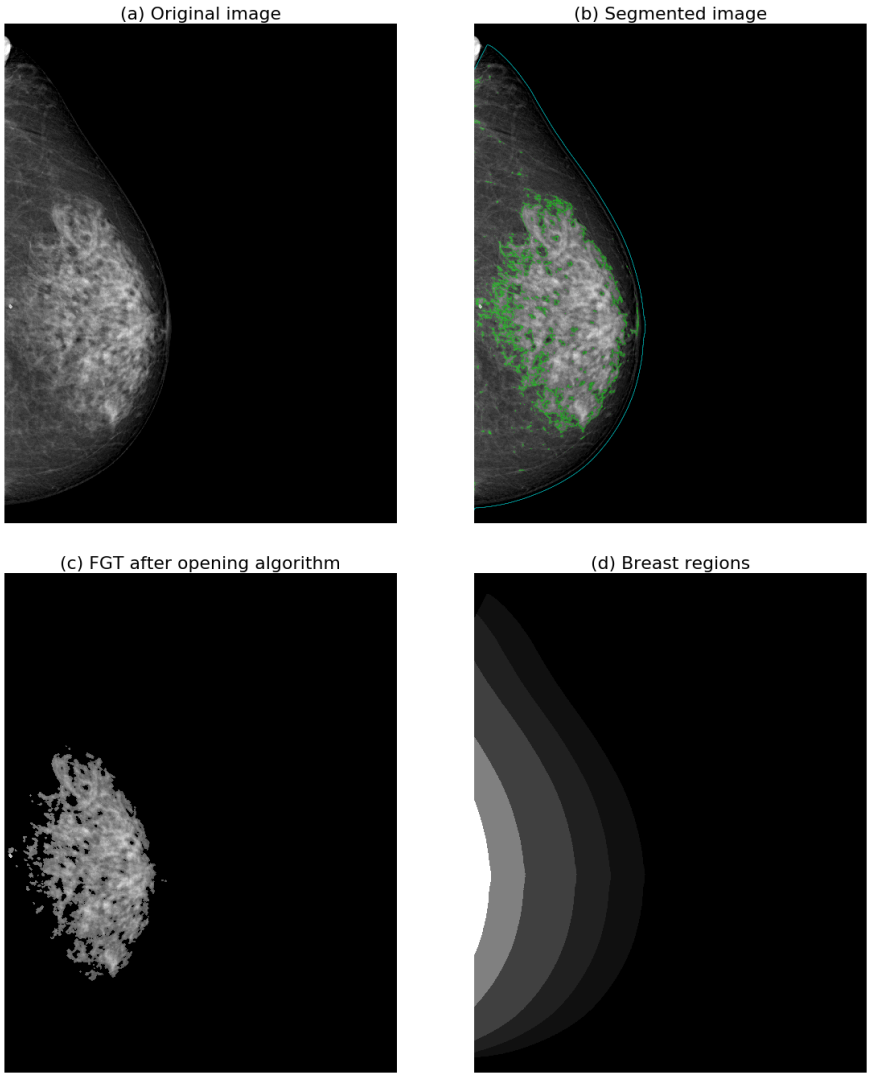

Figure 1: Imaging preprocessing steps for extraction of geometrical variables. (a) Original image. (b) The breast and the FGT thresholding segmentation (blue and green lines respectively); the breast and FGT area, the breast perimeter, the FGT center of mass, and the distance to pectoral muscle is calculated from this information. (c) The FGT region once opening algorithm was applied five times; FGT morphology features and Hu moments were obtained from this image. (d) Regions at different distances to the breast skin line were used to calculate PD in each region.

The value of each bin is calculated as follows:

1. Each bin is represented by the first angle that it covers. For instance, bin 0 represents angles between $[0, 20)$.
2. Considering the pixel defined by (m, α), if α is equal to the name of a bin then m is added to this bin, else m is balanced between this bin and the following. For example, a magnitude 5 with angle 0 will be added to the bin 0, while a magnitude 5 with angle 12 will add 2 to the bin represented by 0 and 3 to the bin represented by 20.

An example of 32×32 Global-HOG based Histogram (G-HOGH) calculation may be followed in Figure 2 for the purpose of making the descriptor definition clearer.

Once the histograms of the whole set of cells belonging to a block have been extracted, a vector concatenating the magnitude of all bins is built and normalized by its euclidean norm. This step is known as *block normalization*. The vector which is composed of the normalized vectors of all blocks covering the window is the HOG descriptor for the window defining a local texture feature of the region of interest.

The novelty of our approach resides in the construction of a global descriptor using the local texture features obtained by the HOGs computation in each window. We have tested the aforementioned window sizes (without overlapping, 50% overlap, and 75% overlap) using 8×8 pixels per cell, 2×2 cells per block, 50% overlap blocks and signed gradients (orientations $\in [0, 2\pi]$) for windows HOG calculation. When all HOG features describing windows had been extracted, a matrix in which each row is the one-window HOG-feature was built. With the purpose of reducing the columns dimension

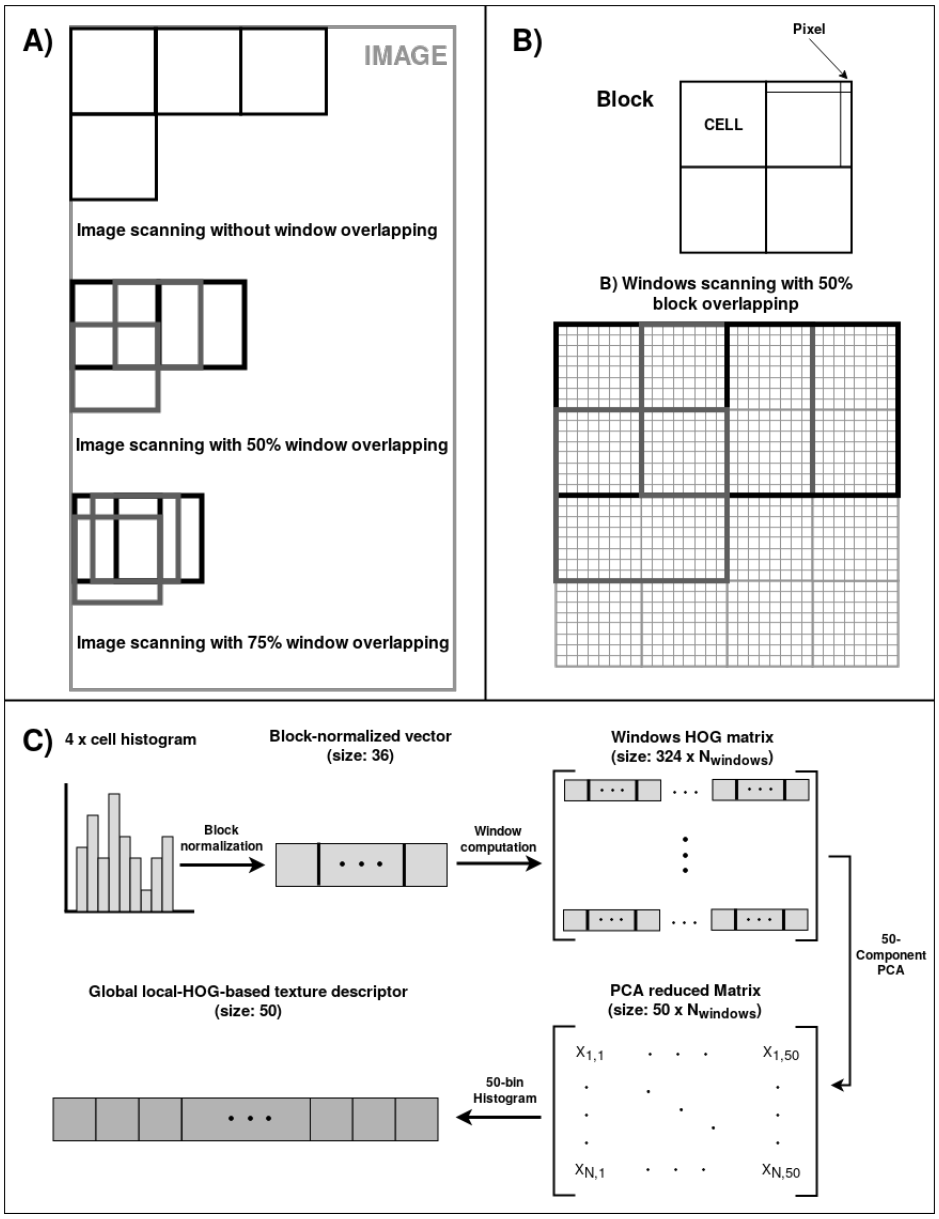

Figure 2: G-HOGH calculation A) describes the schemes used to cover the image with the windows. B) shows how blocks of cells run through windows. C) is the flow-chart to obtain the G-HOGH. Each cell produces a 9-bin histogram then each block produces a 36-bin normalized histogram. The number of blocks needed to scan a 32×32 window is 9, so the HOG window vector has 324 components. Building the matrix in which each row is one HOG window vector allows the dimension reduction by PCA. Computing a 50-bin histograms, the G-HOGH is finally obtained.

of the matrix, we applied principal component analysis (PCA) [26]. By varying the number of components we observed that when 50 components were chosen, more than 60% of the cumulative variance was explained. Since the new coordinates of the window HOG-features described in some way the most variable window gradients, by building a histogram from the new descriptors of all the windows, we were able to define a Global HOG-based Histogram (G-HOGH) which characterized the difference between window gradients through the entire breast.

2.2.3. Predictive model: Random Forest Classifier

In this study, the predictive model is trained to differentiate between women being or becoming cases and healthy controls.

A family of techniques that is receiving a lot of attention in machine learning research is the paradigm of ensemble methods that combine their results, as the case of Random Forests (RF) which construct many decision trees that are used to classify by the majority vote [27, 28].

Many recent works [29–31] have shown that in some tasks RF classification can outperforms other conventional machine learning algorithms, such as back propagation neural networks or support vector machines, and with the advantage of dealing easily with unbalanced or multiclass classification problems [30].

These reasons have motivated the use of RF in our experiments. The parameters were fixed to 512 decision trees composing the forests, the maximum number of decision variables in each tree are set to the $log_2 N$ where N is the number of model inputs and the remaining parameters were fixed to the default proposed by the python implementation of scikit-learn [32]. For

example, our model takes $N = 146$ when more variables are used, then the maximum number of decision variables in each tree is $log_2 146 = 7$.

2.2.4. Evaluation of model performance

Although case/control studies are efficient for diseases with long latency periods, such as cancer [33], there is a lot of controversy about how to measure the model performance because of the potential vulnerability to bias and loss of generalization power [34]. Several studies have proposed different ways to quantify the discriminatory power of classifiers [35, 36] but the area under the Receiver Operating Characteristic (ROC) curve [37] is still the most widely used metrics to assess the model performance. Due to the design of our study, besides the AUC, we have used the Net Reclassification Improvement (NRI) score proposed by Pencina et al. [36] to assess the added predictive ability provided by the inclusion of new features to the model. This score was defined as follows:

$$\widehat{NRI} = (\hat{p}_{up,events} - \hat{p}_{down,events}) - (\hat{p}_{up,nonevents} - \hat{p}_{down,nonevents}) \qquad (1)$$

Where $\hat{p}_{up,events}$ denotes the percentage of misclassified cases in the old model that are well-classified in the new model, $\hat{p}_{down,events}$ the percentage of well-classified cases in the old model that are misclassified in the new model, $\hat{p}_{down,nonevents}$ the percentage of well-classified controls in the old model that are misclassified in the new model and $\hat{p}_{up,nonevents}$ the percentage of misclassified controls in the old model that are well-classified in the new model. A positive value represents that the new model outperforms the number of well-classified patients while a negative value represents that the old-model had more well-classified patients, finally, a value of 0 indicates that

both models have the same number of well-classified patients.

3. Results

3.1. Study characteristics

The original dataset contained a total of 808 case/control pairs, 51.69% of 1563 patients had developed a breast cancer and the remaining percentage had not. It was randomly stratified to shape training and test sets preserving the paired couples and the a priori case/control proportion. The training case/control set had a total of 606 cases and 566 controls summing up to 1172 patients (51.71% cases and 48.29% controls) and was 10-fold stratified to cross-validate the models. The test set was composed of 202 (51.66%) cases and 189 (48.34%) controls.

The original dataset included risk factors such as the number of pregnancies, the age at first pregnancy, the number of breastfeeding months and the age at the onset of menopause. These risk factors were completed by adding three preprocessed variables. The patient age at mammography acquisition was computed as the difference between the mammography date and their birth date. The Body Mass Index (BMI) was calculated as the weight over the height squared. The number of cigarettes per day was ponderated by the weight defined in Equation 2, where f is a logical variable that denotes whether a patient is currently a smoker or not and y denotes the years without smoking. The statistics of risk factor features of the patients at the groups defined by cases and controls in the training and test sets are summarized in

Table 2.

$$w(f,y) = \begin{cases} 1 & \text{if } f \\ 1 - \frac{y}{10} & \text{if } \bar{f} \text{ and } y < 10 \\ 0 & \text{if } \bar{f} \text{ and } y \geq 10 \end{cases} \quad (2)$$

	Training set				**Test set**			
	Cases		Controls		Cases		Controls	
	Mean	*std*	*Mean*	*std*	*Mean*	*std*	*Mean*	*std*
Age	57.20	6.46	57.39	6.61	57.03	6.82	56.66	6.80
BMI	27.18	4.81	27.32	4.88	27.05	4.54	27.23	5.31
Cigarettes per day ratio	3.03	6.34	2.68	6.12	2.97	6.76	3.14	6.69
Number of pregnancies	2.40	1.04	2.58	1.20	2.34	0.95	2.56	1.22
First pregnancy	24.69	5.20	24.30	5.07	24.58	5.27	24.67	4.82
Breastfeeding months	9.02	9.89	9.85	10.42	8.94	10.48	9.57	10.18
Menopause age	48.43	5.47	48.44	4.85	47.85	5.59	48.38	4.30

Table 2: Descriptive statistics for cases/controls in the training and test sets. BMI is the body mass index and std is the standard deviation.

Taking as baseline the accuracy obtained after computing the AUC defined by the ROC of PD in CC and MLO views (0.559 and 0.551 respectively), we have followed a methodology in which each iteration added features sets as inputs for a Random Forest classifier. The features sets, which were added in this order, were:

- **Percentage of dense tissue.** In order to assess an estimator that mimics the volumetric-based PD (which is another extended-used marker for mammographic density), RF will firstly take as inputs the PD for CC and MLO views.

- **Global geometric variables** covering from layout features, such as relative FGT center of mass, FGT distance to pectoral muscle, breast area or FGT perimeter, to the rotation, translation, and scale invariant FGT moments.

- **G-HOGH.** Global summaries of local HOG descriptors. We have tested the results with different window sizes and methods to cover the breast.

The training set was 10-fold stratified cross-validated in order to assure that overfitting did not occur. The following results show the performance evaluation using the test set which was not used as training corpus.

3.2. Experimental results

The PD is known to be a marker of breast cancer development risk. In this sense, to fix the PD-risk obtained from the test set of our data as a baseline seems reasonable. After calculating the PD ROC curve for both views in the test set (391 patients, 202 of which developed cancer and 189 were controls), we obtain AUCs of 0.559 and 0.551 for CC view and MLO view respectively.

Taking the aforementioned results as the baseline, new features were progressively added to test its discriminatory capacity between cases and controls and their possible contribution to a better estimate of cancer risk compared to PD.

Our first Random Forest-based model only included PD from both views and the results showed an AUC of 0.560. After adding the geometrical variables, the AUC improved up to 0.568. The number of well-classified cases

was 116 while the number of well-classified controls was 92 which meant a total of 208 (53%) patients well-classified, always taking as cohort point 0.50.

Finally, several G-HOGH configurations were tested, as mentioned in Section 2.2.2. All of them produced better results than those obtained by the previous models, with exception of 24-pixel overlap windows. It may indicate that the proposed texture feature strengthens the variables introduced before into the model. A short brief of the results are shown in Figure 3 and Table 3.

Window size	**Overlap (%)**	**AUC**	**NRI 1**	**NRI 2**	**TP**	**TN**
24×24	0	0.571	0.130	0.064	131/202	90/189
24×24	50	0.566	0.110	0.044	126/202	91/189
24×24	75	0.558	0.089	0.023	125/202	88/189
32×32	0	0.614	0.190	0.154	142/202	94/189
32×32	50	0.605	0.118	0.083	135/202	86/189
32×32	75	0.589	0.119	0.083	133/202	88/189
64×64	0	0.610	0.165	0.130	136/202	94/189
64×64	50	0.600	0.165	0.130	137/202	93/189
64×64	75	0.588	0.114	0.079	130/202	90/189
128×128	0	0.609	0.171	0.135	136/202	95/189
128×128	50	0.600	0.165	0.130	137/202	93/189
128×128	75	0.584	-0.005	0.110	133/202	93/189

Table 3: Indicators for the evaluation of models performance. Number of cases and controls well classified. NRI 1 denotes the NRI score from CC PD view taking as cohort point 0.5 and NR2 denotes the NRI score from PD + geometrical model. TP refers to true positives and TN to true negatives

It is worth to mention that the percentage of well-classified cases improves up to 70% when the percentage of well-classified cases in the model which

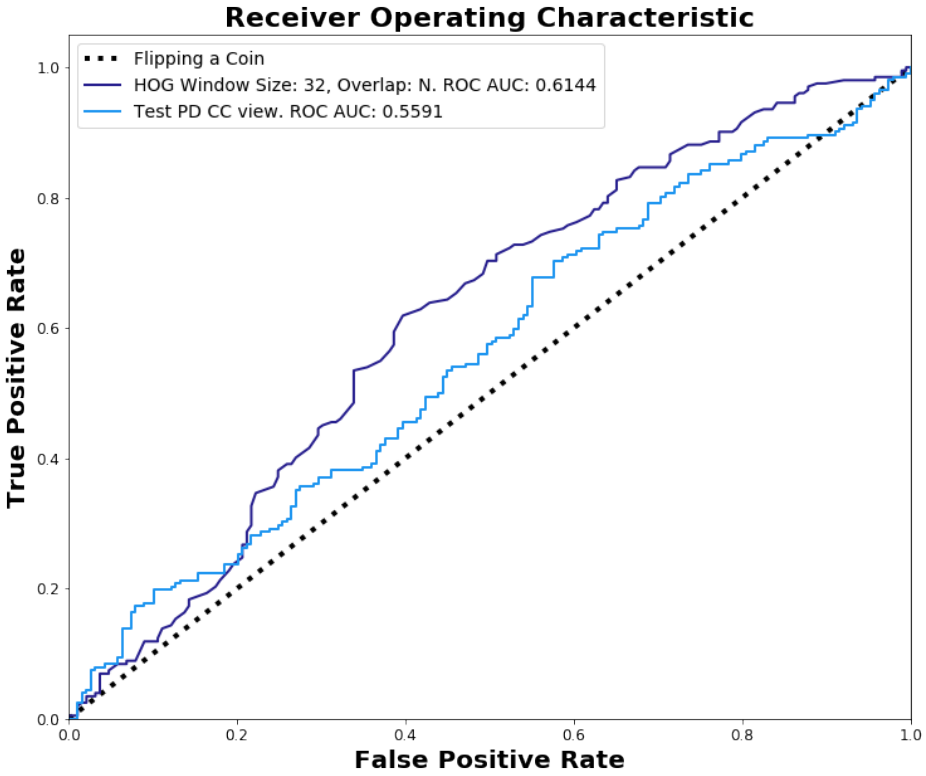

Figure 3: Geometric and Best G-HOGH configuration ROC curve against the baseline ROC curve. Different G-HOGH configurations produced ROC curves which outperform the results obtained by only including geometrical variables. The best AUC, which is shown in this figure, increased a 5% with respect to the previous model.

only considered both PD views and geometrical variables was 56%. After adding the different G-HOGH features, the best ROC curve produces a 5-point increment from the previous model. The best NRI score is obtained by the G-HOGH configuration with non-overlap 32-pixel window size. This configuration achieved 31 patients (28 cases and 3 controls) more well-classified from those obtained only considering PD, and 28 (26 cases and 2 controls) more well-classified patients from the model which took as entries PD and

geometric features.

These results arouse that the texture information extracted from the mammography increase in a higher extent the sensitivity of PD as a risk marker of developing breast cancer than the specificity, although this last is also improved.

In order to assess the performance indicators for the best model, the confusion matrix, AUC, and sensitivity and specificity scores are presented in Table 4.

	$\boldsymbol{CASES_{real}}$	$\boldsymbol{CONTROLS_{real}}$	$\boldsymbol{Total_{pred}}$
$\boldsymbol{CASES_{real}}$	142	96	**238**
$\boldsymbol{CONTROLS_{real}}$	60	93	**153**
$\boldsymbol{Total_{real}}$	**202**	**189**	**391**
		AUC ROC	0.614
		Sensitivity	$\approx 70\%$
		Specificity	$\approx 49\%$

Table 4: Confusion matrix for the predictions achieved by the best model. AUC score, sensitivity and specificity are also shown.

Besides, the experiment was also replicated by introducing the G-HOGH features calculated only scanning the FGT in addition to PD and geometric variables. The results (AUCs ranging from 0.557 to 0.586) did not improve those obtained by the G-HOGH across the entire breast, which indicates that texture information extracted from the whole breast is more discriminative than the extracted only from the FGT.

A model was also trained taking as inputs the G-HOGH features only. The results were lower than those obtained using only PD, which indicates that the information provided by the proposed texture descriptor strengthens

PD as cancer risk marker, but it can not be considered alone as a marker.

4. Discussion

4.1. Impact and contributions

The study aimed to predict breast cancer risk from healthy breast mammographies using texture information. The clinical importance of PD is not only the reduced mammographic sensitivity but also an increased breast cancer risk [38]. In this sense, the search of malignant patterns blurred by the dense tissue in digital mammographies is a growing research line [14–17, 19, 39] which is known as parenchymal texture descriptors.

In line with this, the first contribution of our work is the definition of a global texture descriptor based on local HOG covering the entire breast. The results showed evidence that the texture of a healthy breast increases the specificity of PD as a marker of breast cancer development risk. More than 30 of 391 ($\approx 8\%$) patients were well-classified by adding the information provided by the G-HOGH. Window sizes of 32×32 px covering areas of 2.56 mm^2 ($1.6 \times 1.6mm$) provided the best results both in terms of ROC AUC and number of well-classified patients.

The second noteworthy contribution is the use of Random Forests. The common model used in tasks similar to those carried out in this work is the logistic regression [8, 40–42]. Even though the principal argument of the present paper is the contribution of the texture features role in the breast cancer development risk, we have developed a logistic regressor using as inputs (PD, geometric variables, and G-HOGH) the same used in the Random Forest model for comparison purposes (see Figure 4).

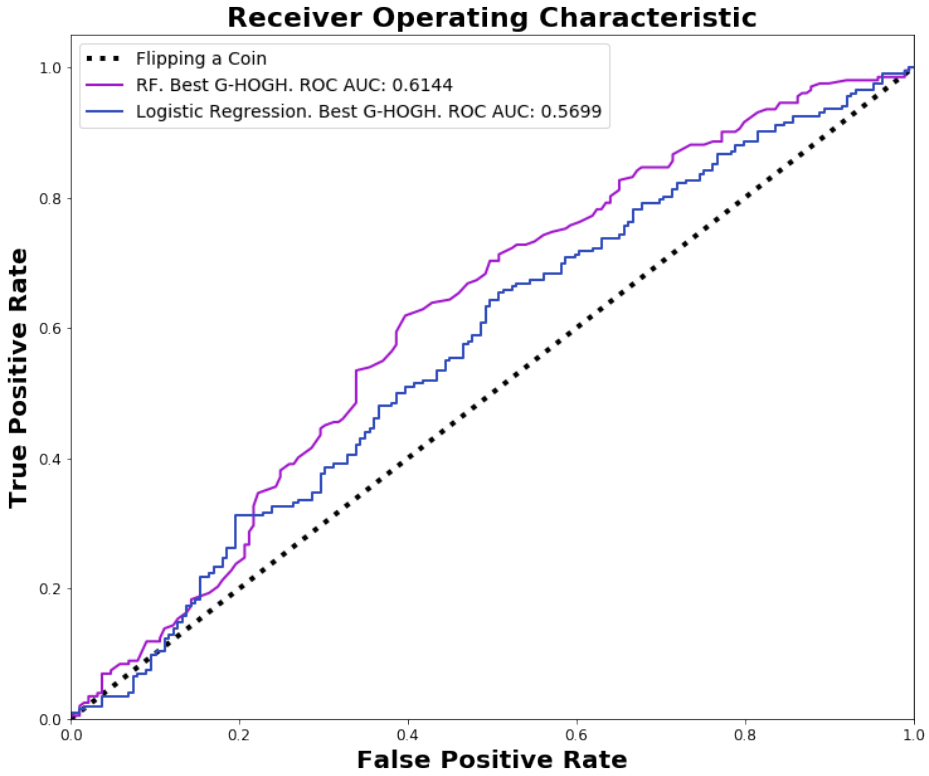

Figure 4: Comparison of the best ROC curve using logistic regression and the best ROC curve using RF. The ROC curves for testing samples after developing a logistic regression model with training samples obtained worse results than RF.

The AUC results of Random Forest ($[0.584, 0.614]$) for all the configurations of G-HOGH provided better results than logistic regression ($[0.551, 0.570]$) and the NRI scores comparing Random Forest with logistic regression are between $[0.005, 0.171]$ (all the values fall within the positive real numbers) demonstrating the better performance of that model.

4.2. Limitations an Future work

The use of a global texture feature summarizing local HOG descriptors strives to characterize the overall texture distribution in an affordable way, but it may become a drawback due to the possible local texture smoothing which could hide pattern structures that would be helpful for the task at hand. New encoded descriptors based on HOG are proposed as a line for future work, changing the use of dimension reduction linear methodologies, such as PCA which has been used in the present study, by non-linear methodologies such as ISOMAP [43] or autoencoders [44] what could improve the predictive ability.

As previously mentioned, the scope of this paper is the presentation of new texture features which strengthens PD as a cancer risk marker. In this sense, a machine learning based model was fixed and trained with incremental sets of variables, with the purpose of searching the combinations which produced the best estimation. Once the task was achieved, the next step consists of the performance optimization. Among the future works which are contemplated, we want to highlight the use of feature selection algorithms, such as Sequential Forward Floating Searching [45]. These methods not only search sets of variables but also particular variables influencing cancer development risk with the purpose of avoiding possible redundancies and consequently improving the performance. We also consider the use of other state-of-the-art classifiers, such as algorithms based on visual bag of words [46], which may improve the decision space representability and therefore provide better results.

5. Conclusion

A global texture feature (G-HOGH) based on HOG of the entire breast has been proposed as a characteristic which increases the specificity (up to 70%) of breast cancer development risk. The G-HOGH is a sliding-window-based algorithm; in our study, the $32 \times 32px$ windows covering the breast have provided the best results.

The G-HOGH bolstered the information which can be extracted from geometrical variables improving the predictive ability. The results of our experiments inform that in our dataset the proposed G-HOGH features contribute to a higher extent when non-overlapping windows are used. The model which obtained the best results increased in 31 the number of well-classified patients with respect to PD. Our model not only outperformed the specificity of PD as a marker of developing breast cancer risk, but also the sensitivity. It highlights this last with an increment of approximately 14%.

The use of Random Forests instead of other models commonly used in the field, such as logistic regression, improved the results thanks to its adaptability to the decision space. In any case, the use of more powerful machine learning-based classifiers may even increase the results shown in this paper.

Acknowledgements

The authors of this work like to thank to Guillermo García Colomina, Carlos Barata Ferrando and Empar Giner Ferrando for their support in recruitment and data collection.

Funding

This work was partially funded by Generalitat Valenciana through I+D IVACE (Valencian Institute of Business Competitiviness) and GVA (European Regional Development Fund) supports under the project IMAMCN/2018/1, and by Carlos III Institute of Health under the project DTS15/00080.

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of the Universitat Politècnica de València (project name: "DM-Scan Herramienta de lectura de densidad mamográfica como fenotipo marcador de riesgo de cáncer de mama") and consent was obtained from study participants at the time of screening.

References

[1] J. N. Wolfe, Breast patterns as an index of risk for developing breast cancer, AJR Am. J. Roentgenol. 126 (6) (1976) 1130–1137 (1976). doi:10.2214/ajr.126.6.1130.

[2] V. Assi, J. Warwick, J. Cuzick, S. W. Duffy, Clinical and epidemiological issues in mammographic density, Nat. Rev. Clin. Oncol. 9 (1) (2012) 33 (2012).

[3] N. F. Boyd, J. M. Rommens, K. Vogt, V. Lee, J. L. Hopper, M. J. Yaffe, A. D. Paterson, Mammographic breast density as an intermediate phenotype for breast cancer, Lancet Oncol. 6 (10) (2005) 798–808 (2005). doi:10.1016/S1470-2045(05)70390-9.

[4] A. Burton, G. Maskarinec, B. Perez-Gomez, C. Vachon, H. Miao, M. Lajous, R. López-Ridaura, M. Rice, A. Pereira, M. L. Garmendia, R. M. Tamimi, K. Bertrand, A. Kwong, G. Ursin, E. Lee, S. A. Qureshi, H. Ma, S. Vinnicombe, S. Moss, S. Allen, R. Ndumia, S. Vinayak, S.-H. Teo, S. Mariapun, F. Fadzli, B. Peplonska, A. Bukowska, C. Nagata, J. Stone, J. Hopper, G. Giles, V. Ozmen, M. E. Aribal, J. Schüz, C. H. Van Gils, J. O. P. Wanders, R. Sirous, M. Sirous, J. Hipwell, J. Kim, J. W. Lee, C. Dickens, M. Hartman, K.-S. Chia, C. Scott, A. M. Chiarelli, L. Linton, M. Pollan, A. A. Flugelman, D. Salem, R. Kamal, N. Boyd, I. Dos-Santos-Silva, V. McCormack, Mammographic density and ageing: A collaborative pooled analysis of cross-sectional data from 22 countries worldwide, PLoS Med. 14 (6) (2017) e1002335 (2017). doi:10.1371/journal.pmed.1002335.

[5] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, M. J. Yaffe, Mammographic Density and the Risk and Detection of Breast Cancer, N. Engl. J. Med. 356 (3) (2007) 227–236 (2007). doi:10.1056/NEJMoa062790.

[6] J. Li, L. Szekely, L. Eriksson, B. Heddson, A. Sundbom, K. Czene, P. Hall, K. Humphreys, High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer, Breast Cancer Res. 14 (4) (2012) R114 (2012). doi:10.1186/bcr3238.

[7] B. M. Keller, D. L. Nathan, Y. Wang, Y. Zheng, J. C. Gee, E. F. Conant, D. Kontos, Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-

means clustering and support vector machine segmentation, Med. Phys. 39 (8) (2012) 4903–4917 (2012). doi:10.1118/1.4736530.

[8] C. Wang, A. R. Brentnall, J. Cuzick, E. F. Harkness, D. G. Evans, S. Astley, A novel and fully automated mammographic texture analysis for risk prediction: results from two case-control studies, Breast Cancer Res. 19 (1) (2017) 114 (2017). doi:10.1186/s13058-017-0906-6.

[9] C. J. D'Orsi, E. Sickles, E. Mendelson, E. Morris, ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, Reston, VA, American College of Radiology, 2013 (2013).

[10] N. Vállez, G. Bueno, O. Déniz, J. Dorado, J. A. Seoane, A. Pazos, C. Pastor, Breast density classification to reduce false positives in CADe systems, Comput. Methods Programs Biomed. 113 (2) (2014) 569–584 (2014).

[11] S. Dhahbi, W. Barhoumi, J. Kurek, B. Swiderski, M. Kruk, E. Zagrouba, False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification, Comput. Methods Programs Biomed. 160 (2018) 75–83 (2018).

[12] L. Blot, R. Zwiggelaar, Background texture extraction for the classification of mammographic parenchymal patterns, in: Medical Image Understanding and Analysis, 2001, pp. 145–148 (2001).

[13] A. Oliver, J. Freixenet, R. Martí, J. Pont, E. Pérez, E. R. E. Denton, R. Zwiggelaar, A novel breast tissue density classification methodology,

IEEE Trans. Inf. Technol. Biomed. 12 (1) (2008) 55–65 (2008). doi:10.1109/TITB.2007.903514.

[14] A. Manduca, M. J. Carston, J. J. Heine, C. G. Scott, V. S. Pankratz, K. R. Brandt, T. A. Sellers, C. M. Vachon, J. R. Cerhan, Texture Features from Mammographic Images and Risk of Breast Cancer, Cancer Epidemiol. Biomarkers Prev. 18 (3) (2009) 837–845 (2009). doi:10.1158/1055-9965.EPI-08-0631.

[15] D. Kontos, L. C. Ikejimba, P. R. Bakic, A. B. Troxel, E. F. Conant, A. D. A. Maidment, Analysis of parenchymal texture with digital breast tomosynthesis: comparison with digital mammography and implications for cancer risk assessment, Radiology 261 (1) (2011) 80–91 (2011). doi:10.1148/radiol.11100966.

[16] M. Nielsen, C. M. Vachon, C. G. Scott, K. Chernoff, G. Karemore, N. Karssemeijer, M. Lillholm, M. A. Karsdal, Mammographic texture resemblance generalizes as an independent risk factor for breast cancer, Breast Cancer Res. 16 (2) (2014) R37 (2014). doi:10.1186/bcr3641.

[17] Y. Zheng, B. M. Keller, S. Ray, Y. Wang, E. F. Conant, J. C. Gee, D. Kontos, Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment, Med. Phys. 42 (7) (2015) 4149–4160 (2015). doi:10.1118/1.4921996.

[18] M. Kallenberg, K. Petersen, M. Nielsen, M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, M. Lillholm, Unsupervised Deep Learning

Applied to Breast Density Segmentation and Mammographic Risk Scoring, IEEE Trans. Med. Imaging 35 (5) (2016) 1322–1331 (2016). doi:10.1109/TMI.2016.2532122.

[19] A. Gastounioti, A. Oustimov, M.-K. Hsieh, L. Pantalone, E. F. Conant, D. Kontos, Using Convolutional Neural Networks for Enhanced Capture of Breast Parenchymal Complexity Patterns Associated with Breast Cancer Risk, Acad. Radiol. 25 (8) (2018) 977–984 (2018). doi:10.1016/j.acra.2017.12.025.

[20] M. Pollán, R. Llobet, J. Miranda-García, J. Antón, M. Casals, I. Martínez, C. Palop, F. Ruiz-Perales, C. Sánchez-Contador, C. Vidal, B. Pérez-Gómez, D. Salas-Trejo, Validation of DM-Scan, a computer-assisted tool to assess mammographic density in full-field digital mammograms, Springerplus 2 (1) (2013) 242 (2013). doi:10.1186/2193-1801-2-242.

[21] R. Llobet, M. Pollán, J. Antón, J. Miranda-García, M. Casals, I. Martínez, F. Ruiz-Perales, B. Pérez-Gómez, D. Salas-Trejo, J.-C. Pérez-Cortés, Semi-automated and fully automated mammographic density measurement and breast cancer risk prediction, Comput. Methods Programs Biomed. 116 (2) (2014) 105–115 (2014). doi:10.1016/j.cmpb.2014.01.021.

[22] R. M. Haralick, L. G. Shapiro, Computer and robot vision, Addison-wesley, 1992 (1992).

[23] M.-K. Hu, Visual pattern recognition by moment invariants, IRE transactions on information theory 8 (2) (1962) 179–187 (1962).

[24] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 Proc IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR'05), Vol. 1, 2005, pp. 886–893 vol. 1 (2005). doi:10.1109/CVPR.2005.177.

[25] I. Sobel, G. Feldman, A 3x3 isotropic gradient operator for image processing, a talk at the Stanford Artificial Project in (1968) 271–272 (1968).

[26] I. Jolliffe, Principal component analysis, Springer, 2011 (2011).

[27] L. Breiman, Random Forests, Mach. Learn. 45 (1) (2001) 5–32 (2001). doi:10.1023/A:1010933404324.

[28] T. M. Oshiro, P. S. Perez, J. A. Baranauskas, How Many Trees in a Random Forest?, in: International workshop on machine learning and data mining in pattern recognition, Springer, Berlin, Heidelberg, 2012, pp. 154–168 (2012). doi:10.1007/978-3-642-31537-4-13.

[29] B. Wu, T. Abbott, D. Fishman, W. McMurray, G. Mor, K. Stone, D. Ward, K. Williams, H. Zhao, Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data, Bioinformatics 19 (13) (2003) 1636–1643 (2003). doi:10.1093/bioinformatics/btg210.

[30] M. Liu, M. Wang, J. Wang, D. Li, Comparison of random forest, support vector machine and back propagation neural network for electronic

tongue data classification: Application to the recognition of orange beverage and Chinese vinegar, Sens. Actuators B Chem. 177 (2013) 970–980 (2013). doi:10.1016/j.snb.2012.11.071.

[31] T. PanduRanga Vital, M. Murali Krishna, G. V. L. Narayana, P. Suneel, P. Ramarao, Empirical analysis on cancer dataset with machine learning algorithms, Vol. 758, Springer Verlag, 2018 (2018). doi:10.1007/978-981-13-0514-6_75.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830 (2011).

[33] K. F. Schulz, D. A. Grimes, Case-control studies: research in reverse, Lancet 359 (9304) (2002) 431–434 (2002). doi:10.1016/S0140-6736(02)07605-5.

[34] K. Han, K. Song, B. W. Choi, How to Develop, Validate, and Compare Clinical Prediction Models Involving Radiological Parameters: Study Design and Statistical Methods, Korean J. Radiol. 17 (3) (2016) 339–350 (2016). doi:10.3348/kjr.2016.17.3.339.

[35] A. R. Brentnall, J. Cuzick, J. Field, S. W. Duffy, A concordance index for matched case–control studies with applications in cancer risk, Stat. Med. 34 (3) (2015) 396–405 (2015). doi:10.1002/sim.6335.

[36] M. J. Pencina, R. B. D. Agostino, R. B. D. Agostino, R. S. Vasan, Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond, Stat. Med. 27 (2) (2008) 157–172 (2008). doi:10.1002/sim.2929.

[37] S. H. Park, J. M. Goo, C.-H. Jo, Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists, Korean J. Radiol. 5 (1) (2004) 11–18 (2004). doi:10.3348/kjr.2004.5.1.11.

[38] S. J. Vinnicombe, Breast density: why all the fuss?, Clin. Radiol. 73 (4) (2018) 334–357 (2018). doi:10.1016/j.crad.2017.11.018.

[39] W. Sun, T.-L. B. Tseng, W. Qian, E. C. Saltzstein, B. Zheng, H. Yu, S. Zhou, A new near-term breast cancer risk prediction scheme based on the quantitative analysis of ipsilateral view mammograms, Comput. Methods Programs Biomed. 155 (2018) 29–38 (2018).

[40] C. Wang, A. R. Brentnall, J. Cuzick, E. F. Harkness, D. G. Evans, S. Astley, Exploring the prediction performance for breast cancer risk based on volumetric mammographic density at different thresholds, Breast Cancer Res. 20 (1) (2018) 49 (2018). doi:10.1186/s13058-018-0979-x.

[41] S. M. Wong, I. Prakash, N. Trabulsi, A. Parsyan, D. Moldoveanu, D. Zhang, B. Mesurolle, A. Omeroglu, A. Aldis, S. Meterissian, Evaluating the Impact of Breast Density on Preoperative MRI in Invasive Lobular Carcinoma, J. Am. Coll. Surg. 226 (5) (2018) 925–932 (2018). doi:10.1016/j.jamcollsurg.2018.01.045.

[42] S. W. Duffy, O. W. E. Morrish, P. C. Allgood, R. Black, M. G. C. Gillan, P. Willsher, J. Cooke, K. A. Duncan, M. J. Michell, H. M. Dobson, R. Maroni, Y. Y. Lim, H. N. Purushothaman, T. Suaris, S. M. Astley, K. C. Young, L. Tucker, F. J. Gilbert, Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer, Eur. J. Cancer 88 (2018) 48–56 (2018). doi:10.1016/j.ejca.2017.10.022.

[43] J. B. Tenenbaum, V. d. Silva, J. C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science 290 (5500) (2000) 2319–2323 (2000). doi:10.1126/science.290.5500.2319.

[44] G. E. Hinton, R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, Science 313 (5786) (2006) 504–507 (2006). doi:10.1126/science.1127647.

[45] A. Jain, D. Zongker, Feature selection: evaluation, application, and small sample performance, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 153–158 (1997). doi:10.1109/34.574797.

[46] P. Koniusz, F. Yan, K. Mikolajczyk, Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection, Comput. Vision Image Understanding 117 (5) (2013) 479–492 (May 2013). doi:10.1016/j.cviu.2012.10.010.