

Research Article

Predicting the pathological status of mammographic microcalcifications through a radiomics approach

Min Li¹, Liyu Zhu^{2,3}, Guangquan Zhou⁴, Jianan He⁵, Yanni Jiang^{2,*}, Yang Chen^{5,6,*}¹ School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu 210096, China² Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China³ Department of Radiology, the Second Affiliated Hospital of Nantong University, Nantong, Jiangsu 226001, China⁴ The State Key Laboratory of Bioelectronics, Southeast University, Nanjing, Jiangsu 210096, China⁵ Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China⁶ The Laboratory of Image Science and Technology, Key Laboratory of Ministry of Education, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China

ARTICLE INFO

Keywords:

Nonpalpable microcalcifications

Radiomics

Machine learning

ABSTRACT

Objective The study aimed to develop a machine learning (ML)-coupled interpretable radiomics signature to predict the pathological status of non-palpable suspicious breast microcalcifications (MCs).**Methods** We enrolled 463 digital mammographical view images from 260 consecutive patients detected with non-palpable MCs and BI-RADS scored at 4 (training cohort, $n = 428$; independent testing cohort, $n = 35$) in the First Affiliated Hospital of Nanjing Medical University between September 2010 and January 2019. Subsequently, 837 textures and 9 shape features were subsequently extracted from each view and finally selected by an XGBoost-embedded recursive feature elimination technique (RFE), followed by four machine learning-based classifiers to build the radiomics signature.**Results** Ten radiomic features constituted a malignancy-related signature for breast MCs as logistic regression (LR) and support vector machine (SVM) yielded better positive predictive value (PPV)/sensitivity (SE), 0.904 (95% CI, 0.865–0.949)/0.946 (95% CI, 0.929–0.977) and 0.891 (95% CI, 0.822–0.939)/0.939 (95% CI, 0.907–0.973) respectively, outperforming their negative predictive value (NPV)/specificity (SP) from 10-fold cross-validation (10FCV) of the training cohort. The optimal prognostic model was obtained by SVM with an area under the curve (AUC) of 0.906 (95% CI, 0.834–0.969) and accuracy (ACC) 0.787 (95% CI, 0.680–0.855) from 10FCV against AUC 0.810 (95% CI, 0.760–0.960) and ACC 0.800 from the testing cohort.**Conclusion** The proposed radiomics signature depends on a set of ML-based advanced computational algorithms and is expected to identify pathologically cancerous cases from mammographically undecipherable MCs and thus offer prospective clinical diagnostic guidance.

1. Introduction

Breast cancer screening is widely recognized as the most effective way to improve the rate of early diagnosis, survival rate and quality of life of breast cancer patients [1]. Digital Mammography (DM) proves superior to other breast screening methods, e.g., breast magnetic resonance imaging (MRI) and ultrasound, for the detection of breast MCs, further allowing radiologists to scrutinize possible invasive ductal carcinoma in situ (DCIS) [2–5]. Mammographic abnormalities radiologically scored at 4 according to the Breast Imaging Reporting and Data System (BI-RADS) Atlas [6] are considered highly suspicious for malignancy, and therefore, should undergo pathological examination which confirms the ultimate results as noncancerous or cancerous for further surgical or noninvasive strategy making. However, real clinical statistics reveal only approximately 35% cancerous MCs calling for mastectomy of all the biopsied samples [7]. To raise diagnostic accuracy rate in radiology and to reduce unnecessary biopsies, radiomics for interpretable feature analysis concerning breast MCs is thus in the ascendant.

Radiomics provides prospect for precise diagnosis in clinical oncology studies through the quantitative enhancement of medical imaging repositories that depend on advanced mathematical analysis and artificial intelligence techniques [8–9]. Although the present radiomics

* Corresponding authors: Yang Chen, The Laboratory of Image Science and Technology, Key Laboratory of Ministry of Education, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China (Email: chenyang.list@seu.edu.cn); Yanni Jiang, Department of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China (Email: jyn_njmu@163.com).

scheme is in the paradigm shift to a symbiosis of handcrafted features based on image intensity, shape, or texture and deep features generated by various deep learning networks, the latter presents a surplus of computer science automation and thus undermines its decipherability in clinical diagnosis [10–14]. As noted by Afshar et al. [15], four main processing tasks, i.e., image acquisition, segmentation, feature extraction and statistical analysis with model building constitute the conventional radiomics pipeline for handcrafted features. Specifically, handcrafted features are able to represent local statistical properties established on pixel intensity, e.g., energy, entropy and gray-level co-occurrence matrix (GLCM), from the transformative image derived by wavelet decomposition or by multiscale operators such as the Laplacian of Gaussian filter (LoG) and the Local Binary Pattern technique (LBP) [16–18]. Generally, all the studies reported in medical literature employed a broad range of machine learning techniques such as KNN [19], AdaBoost [20], and Gaussian Naive Bayes (GaussianNB) [21].

In this study, we generated a radiomics discriminative signature using machine learning techniques as the data miner to analyze its determinability between noncancerous and cancerous cases measured by NPV/SP and PPV/SE to reduce dispensable biopsy and supplement further analysis of malignant calcification with radiological imaging knowledge as reference for clinical use.

2. Methods

2.1. Patients

This retrospective study was approved by the Institutional Ethics Committee of the First Affiliated Hospital of Nanjing Medical University. Informed consent was signed and obtained from patients before biopsy or surgery. Furthermore, our study was also in compliance with the Helsinki declaration.

We considered patients who presented to the First Affiliated Hospital of Nanjing Medical University between September 2010 and January 2019. In the study, 260 patients with MCs (mean age 47 years; range 30–85 years) were recruited for investigation, who were diagnosed by core needle biopsy. Among them, 130 (50%) cases were diagnosed as malignant and others (50%) as benign by definitive surgery pathology.

All patients met the following criteria: (1) with MCs detected and BI-RADS scored at 4; (2) undergoing surgical biopsy with mammographic guidance; (3) microcalcifications were confirmed to be successfully excised by X-Ray photography; and (4) pathological result as gold standard. Patients who had the history of any therapy for breast lesions before mammography were excluded.

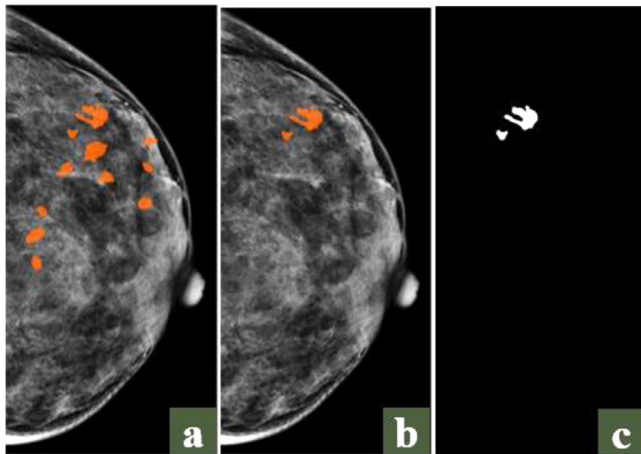


Figure 1. Multiple lesions of MCs presented in a left craniocaudal (LCC) view image (a) and its ROI (b) pathologically diagnosed as malignant and its manually annotated mask (c).

2.2. Data acquisition and annotation

All patients involved underwent digital mammographic examination of the breast on a full-field digital mammography unit (Selenia, Hologic, USA). Two expert radiologists who were blinded to the pathology results independently examined mammograms and categorized MCs according to morphology and distribution descriptors at screening. The tissue blocks were reviewed and discussed by radiologists and pathologists during the consultation after biopsies. They would discuss and make consensus in case of disagreement. Following that, the results were sent to locate and select regions of interest (ROIs) in the Digital Imaging and Communications in Medicine (DICOM) images with bounding boxes.

To analyze the corresponding computational shape features of ROIs, regional masks representing MCs distributional information had to be manually segmented by a radiologist with 10 years of experience and is displayed in Figure 1c, owing to its relevance to the possible malignancy as defined in the latest BI-RADS Atlas Fifth Edition [6].

2.3. Radiomics feature extraction

Based on the original DM and its corresponding ROI mask integrated with another three categories of graphical transform operators at multiple scales as derivative image types, shape features, intensity-based statistics and interpixel relationships were accordingly calculated. All image analysis development work was performed within Python 3.7.3 using the PyRadiomics package.

2.3.1. Shape indicators

Nine of the two-dimensional shape features (Shape2D) exhibited in the last row of Table 1, i.e., elongation, major axis length, maximum diameter, mesh surface, minor axis length, perimeter, perimeter surface ratio, pixel surface and sphericity, were served as distributional descriptors of breast MCs and were only calculated on the nonderived image and mask for the sake of its independency from graylevel intensity distribution. Definitions of each aforementioned shape metrics e.g., elongation, can be found in [22].

2.3.2. Image transformation

To combine intensity properties and textures on various hierarchical scales, three universally used texture operators, i.e., Laplacian of Gaussian (LoG) spatial filters ($\sigma = 1, 2, 3$), single level coiflet decomposition and Local Binary Pattern (LBP) algorithm in its uniform pattern were applied to the original DM image, respectively, before following the texture extraction.

2.3.3. Intensity-based and interpixel relationships

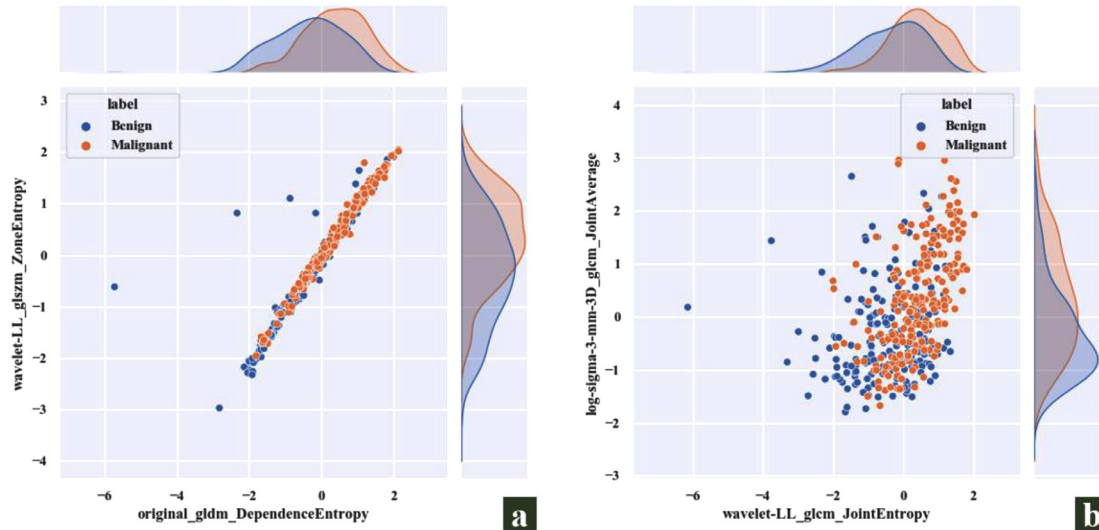
With regard to texture analysis, we investigated 18 first-order statistical features representing intensity properties, 24 gray-level co-occurrence matrix (GLCM) features, 16 gray-level size zone matrix (GLSZM) features, 16 gray-level run length matrix (GLRLM) features, 5 neighboring gray tone difference matrix (NGTDM) features and 14 gray-level dependence matrix (GLDM) features representing interpixel relationships, as defined in [22] and also stated in Table 1. A total of 93 metrics were employed on the aforementioned three derivative image types and the original one. Ultimately, all the six classes of radiomics textures amounting to 837 for each of the view images were extracted in four different image types, including an original image and derived images of the other three types in question, i.e., LoG filtered, wavelet transformed, and LBP operated.

2.4. Radiomics signature building

The yielded textural and shape radiomic features were scaled by Z-score standardization for preprocessing to render into the standard normal distribution. Next, three feature selection methods were applied successively on the preprocessed features for signature construction. Firstly,

Table 1 Statistical measures of radiomics features

Feature categories	Feature names (optional)	Dimensions
First-order statistics	10Percentile, 90Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity, Variance	18
GLCM	Autocorrelation, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Inverse Difference (ID), ID Normalized (IDN), Inverse Difference Moment (IDM), IDM Normalized (IDMN), Informational Measure of Correlation (IMC) 1, IMC2, Inverse Variance, Joint Average, Joint Energy, Joint Entropy, Maximal Correlation Coefficient (MCC), Maximum Probability, Sum Average, Sum Entropy, Sum Squares	24
GLSZM	Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity, Size Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Gray Level Emphasis, Small Area Low Gray Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance	16
GLRLM	Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis	16
NGTDM	Busyness, Coarseness, Complexity, Contrast, Strength	5
GLDM	Dependence Entropy, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Dependence Variance, Gray Level Non-Uniformity, Gray Level Variance, High Gray Level Emphasis, Large Dependence Emphasis, Large Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Low Gray Level Emphasis, Small Dependence Emphasis, Small Dependence High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis	14
Shape2D	Elongation, Major Axis Length, Maximum Diameter, Mesh Surface, Minor Axis Length, Perimeter, Perimeter Surface Ratio, Pixel Surface, Sphericity	9

**Figure 2.** Plots demonstrate features with (a) and without (b) significant linear dependency.

we employed the one-way Analysis of Variance (ANOVA) F-test to sift discriminant features for their pathological nature. However, [Figure 2](#) revealed that ANOVA-selected candidates still intertwined in certain linearity. Therefore, Pearson redundancy-based filter (PRBF) [23] was performed on the rest of the features to statistically drop out linearly dependent features from the feature space ([Figure 2](#)). Subsequently, ordinary least squares (OLS) [24], a backward elimination algorithm that established a linear regression model for the estimation of each feature variable's weight further crossed off those less important candidates. Following that, the final feature importance ranking was generated through the retrospective algorithm, i.e., XGBoost-embedded RFE [25], to gain insights into clinical interpretability. The importance metric was set as 'weight', namely the percentage representing the relative number of times a particular feature occurs in the boosted trees of the model.

After obtaining the 10 top-ranking features, four machine learning-based algorithms, i.e., LR, SVM, Gaussian Naive Bayes (GaussianNB) and K-nearest neighbors (KNN), were performed to classify our clinical DM dataset. As for SVM, Gaussian Radial Basis Function (RBF) was set

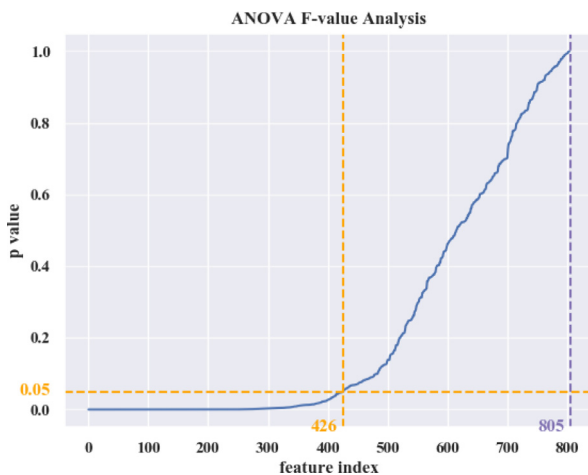
as the kernel function. The other key parameters such as C/gamma in SVM and solver/tolerance in LR were cherry-picked by grid search.

2.5. Validation of the radiomics signature model

During training, validation and testing, we considered views, i.e., craniocaudal (CC) and mediolateral oblique (MLO), as individuals for pure imaging research of DMs. However, predictive outputs derived from multiple views for the same patient were supposed to be combined through voting for the final diagnostic result in clinical practice. Thus, the prognostic performance of the above signature models was evaluated using 10FCV method and also the independent testing set. The so-called independent indicates that different views of the same breast were not involved in both training set and testing set simultaneously. Instead, this situation could possibly happen during 10FCV in light of regarding each view as a single study object.

The overall quality of classifiers in question was quantified by AUC scores and ACC, with its diagnostic ability for noncancerous MCs assessed by NPV/SP and cancerous PPV/SE.

OLS Regression Results						
Dep. Variable:	label	R-squared:	0.167			
Model:	OLS	Adj. R-squared:	0.155			
Method:	Least Squares	F-statistic:	14.06			
Date:	Thu, 25 Mar 2021	Prob (F-statistic):	1.32e-14			
Time:	12:56:57	Log-Likelihood:	-271.48			
No. Observations:	428	AIC:	557.0			
Df Residuals:	421	BIC:	585.4			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.5093	0.022	22.904	0.000	0.466	0.553
original_shape2D_PerimeterSurfaceRatio	0.1008	0.092	1.093	0.275	-0.080	0.282
original_shape2D_MaximumDiameter	-0.0500	0.037	-1.368	0.172	-0.122	0.022
original_shape2D_Sphericity	-0.0062	0.042	-0.147	0.883	-0.089	0.076
wavelet-LL_gldm_JointEntropy	0.2752	0.118	2.326	0.020	0.043	0.508
wavelet-LL_gldm_DependenceEntropy	0.2200	0.155	1.419	0.157	-0.085	0.525
wavelet-LL_gldm_SumEntropy	-0.1846	0.175	-1.054	0.292	-0.529	0.160
Omnibus:	4960.619	Durbin-Watson:	0.196			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	38.898			
Skew:	-0.078	Prob(JB):	3.58e-09			
Kurtosis:	1.531	Cond. No.	24.5			
Warnings:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Figure 3. Intermediate results of *t*-test in ordinary least squares.Figure 4. *P* values derived from ANOVA analysis.

2.6. Statistical analysis

The feature-label correlation and the inter-feature redundancy were evaluated using ANOVA F-statistical and Pearson correlation coefficients successively, and their nonparametric version as appropriate. T-test was performed in OLS, as exhibited in Figure 3, to examine the significance of the parameters fitted by its linear regression model.

With respect to AUC analysis, MannWhitney statistics was used as an estimator of AUC, around which the 95% confidence interval (CI) was obtained by the Delong test. Youden's index was thereby calculated to find the optimal cut-point, so was the confusion matrix to produce ACC, PPV/SE and NPV/SP.

The mentioned statistical techniques were all practiced based on the Python 3.7.3 packages concerned, e.g., sklearn. metrics and scipy. stats. As a routine, the significance level was set as 5%.

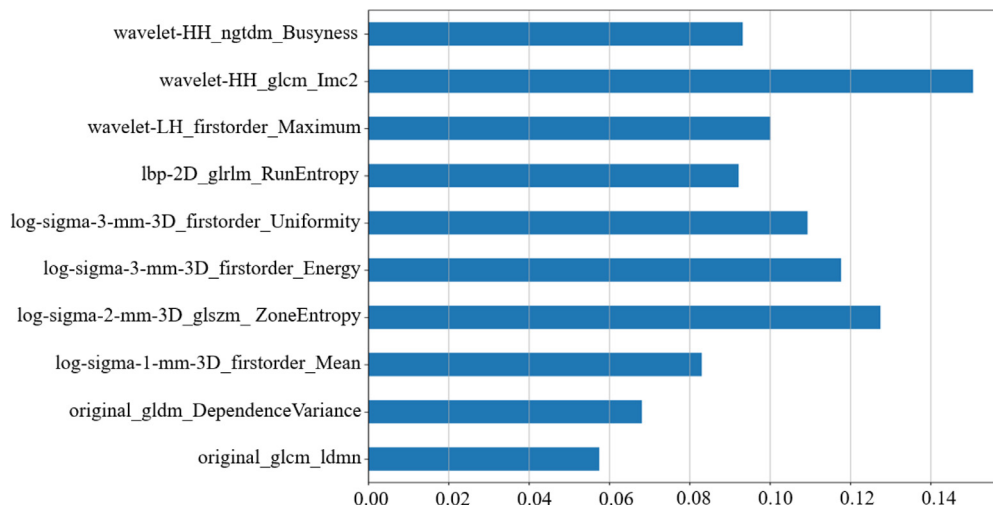


Figure 5. Feature importance ranking by XGBoost in RFE.

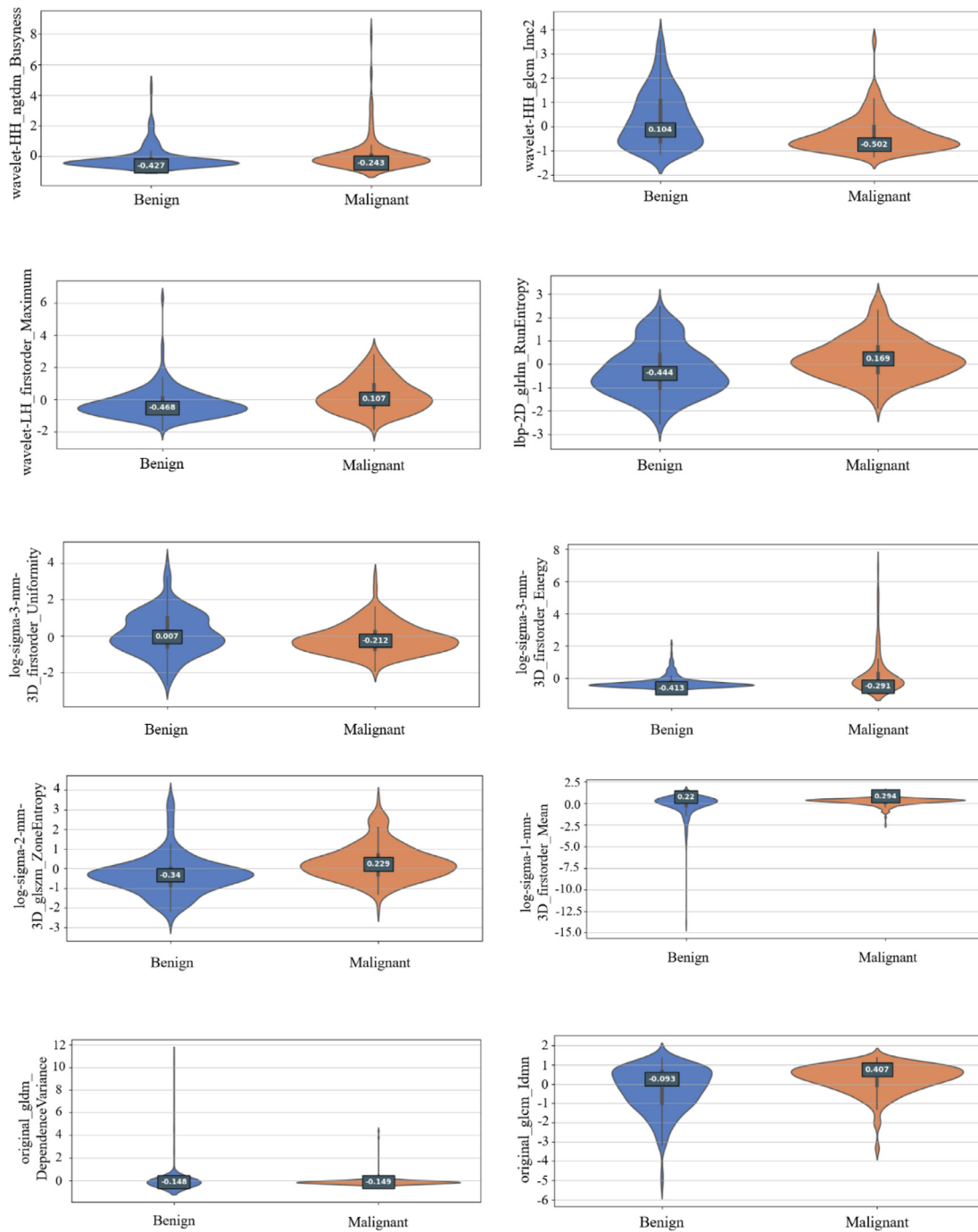


Figure 6. The violinplots of the 10 textural features.

3. Results

3.1. Feature selection

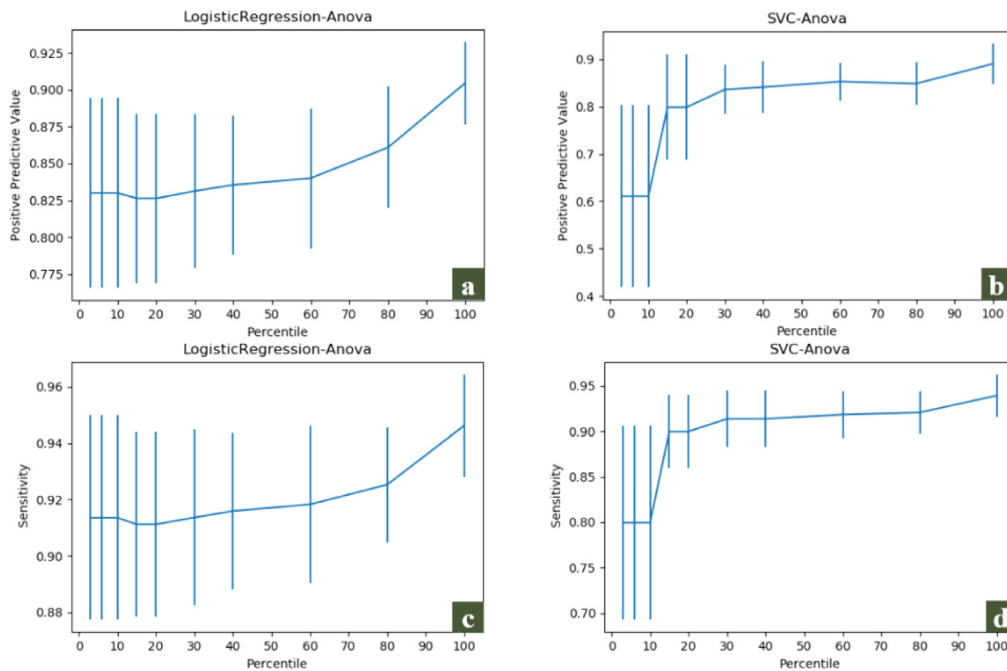
Preliminarily, 837 textures and 9 shapes were extracted from each view in our clinical non-public training cohort. Posterior to Z-score and missing/invalid data handling, 805 features were trimmed by the one-way Analysis of Variance (ANOVA) F-test, yielding 426 candidates with their P values less than 0.05 as in Figure 4. PRBF was then implemented within the above candidates to shrink linearity in the feature set. The optimal set containing 10 textures, as given in Figure 5 displaying their importance ranking, was ultimately derived by RFE. Moreover, it was also expected for all the selected features to significantly differ between the two categories of benign and malignant ($P < 0.05$).

3.2. Radiomics signature

As presented in Figure 6, larger “wavelet-HH_ngtdm_Busyness” and “wavelet-LH_firstorder_Maximum”, indicating more rapid changes between pixels on the coiflet transformed image, were found in the malignant cases than those in the benign cases, so were “lbp-2D_glrlm_RunEntropy”, “log-sigma-3-mm-3D_firstorder_Energy”, “log-sigma-2-mm-3D_glszm_ZoneEntropy” and “log-sigma-1-mm-3D_firstorder_Mean” which implied higher heterogeneity or disorder in the texture patterns for the malignant. However, “wavelet-HH_glcmlm_Imc2” and “log-sigma-3-mm-3D_firstorder_Uniformity” were higher in the benign group, lending support to the assumption that the noncancerous MCs exhibited higher textural uniformity under the corresponding distributional mask of ROIs.

Table 2 Performance summary of the radiomics signature

10FCV Description	Metrics (95%CI)	LR	SVM	KNN	GaussianNB
Malignancy-focused	PPV	0.904 (0.865–0.949)	0.891 (0.822–0.939)	0.768 (0.557–0.985)	0.688 (0.519–0.890)
	SE	0.946 (0.929–0.977)	0.939 (0.907–0.973)	0.512 (0.257–0.845)	0.773 (0.482–0.920)
Benignity-focused	NPV	0.763 (0.592–0.907)	0.782 (0.606–0.878)	0.592 (0.436–0.688)	0.767 (0.591–0.922)
	SP	0.807 (0.566–0.980)	0.763 (0.433–0.971)	0.848 (0.639–0.989)	0.710 (0.467–0.913)
Overall Performance	AUC	0.900 (0.836–0.964)	0.906 (0.834–0.969)	0.831 (0.749–0.913)	0.776 (0.674–0.878)
	ACC	0.787 (0.685–0.860)	0.787 (0.680–0.855)	0.654 (0.545–0.757)	0.705 (0.516–0.890)
Testing					
Overall Performance	AUC	0.790 (0.723–0.956)	0.810 (0.760–0.960)	0.783 (0.623–0.950)	0.703 (0.516–0.890)
	ACC	0.713	0.800		

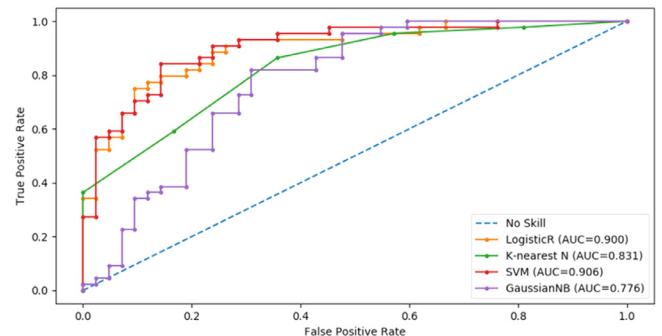
**Figure 7.** PPV/SE values of the LR (a, c) and of the SVM (b, d).

3.3. Validation and prognosis

To acquire in-depth knowledge about the clinical implication of the selected 10 radiomics features, we investigated six metrics, i.e., PPV/SE, NPV/SP, ACC and AUC, derived from 10FCV and on the testing cohort, as summarized in Table 2. Performance of KNN and of GaussianNB is generally inferior to that of LR and of SVM from 10FCV. As a result, they were dismissed in the testing evaluation of ACC. Furthermore, the developed radiomics signature was considered to be more related to the prediction of cancerous lesions based on the higher and more stable PPV/SE scores from both LR and SVM, specifically achieving 0.904 (95% CI, 0.865–0.949)/0.946 (95% CI, 0.929–0.977) and 0.891 (95% CI, 0.822–0.939)/0.939 (95% CI, 0.907–0.973) respectively for validation, as observed in Figure 7. Alongside the confusion matrix-related metrics, the AUC scores of the proposed prognostic models were also analyzed as in Figure 8 for validation and in Figure 9a for testing.

4. Discussion

In this study, we developed an ML-aided radiomics signature to predict microcalcifications as cancerous or noncancerous and found 10 interpretable textures as malignancy-related. It demonstrated a moderate prognostic performance for non-palpable MCs status in our clinical dataset, presenting possibility to provide clinical guidance in pertinent therapy.

**Figure 8.** AUC scores from 10FCV within the training cohort of the 428 DM views.

For the signature construction, 837 textures and 9 shapes serving as feature candidates were trimmed to the 10 most contributive, among which GLCM-associated characteristics following the coiflet transformation and first-order statistics succeeding to the LoG spatial filtering, e.g., “wavelet-HH_glcmmc2”, “log-sigma-3-mm-3D_firstorder_Energy” and “log-sigma-1-mm-3D_firstorder_Mean”, dictated the discriminability of the classification model. More specifically, the selected radiomics features translated properties such as heterogeneity, uniformity and disorder of mammographic MCs into computable values. It is also worth noticing that all the shape indicators were excluded by the OLS - because

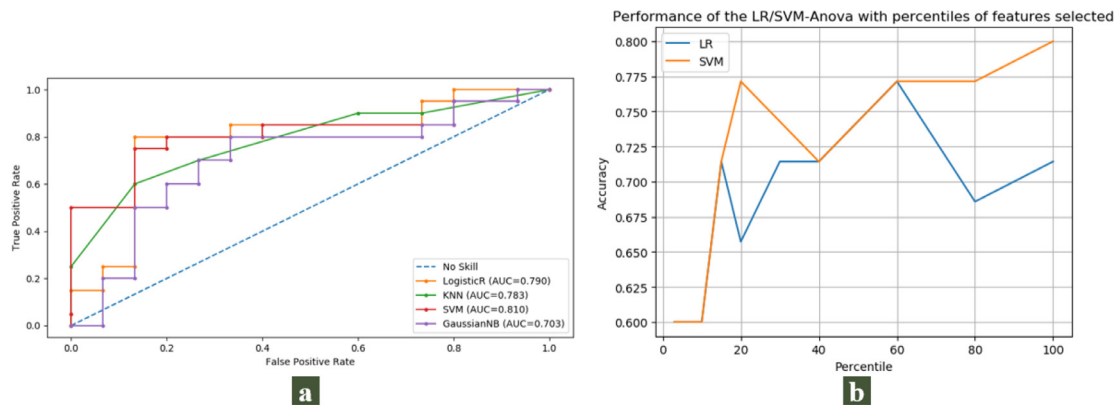


Figure 9. AUC (a) and ACC (b) results using 35 view images of 20 patients.

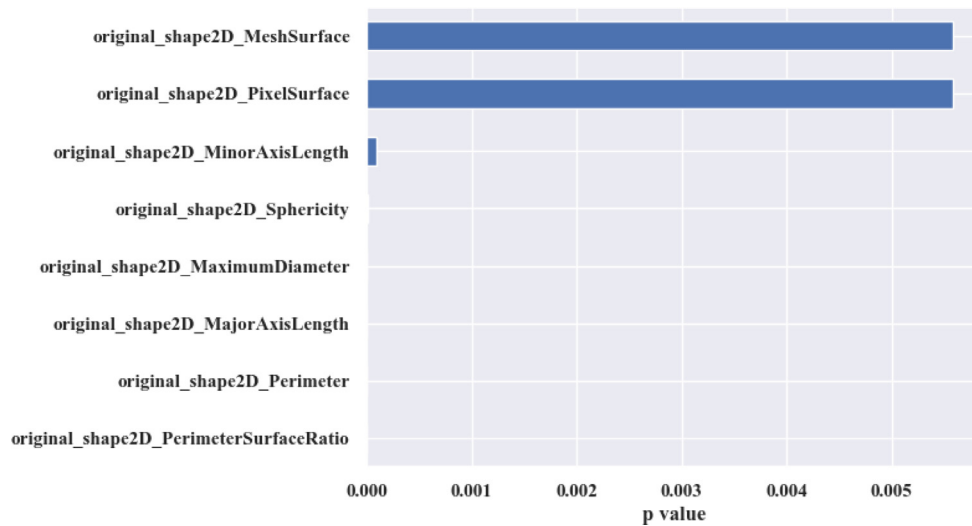


Figure 10. Significance of difference between benign/malignant shapes tested by ANOVA.

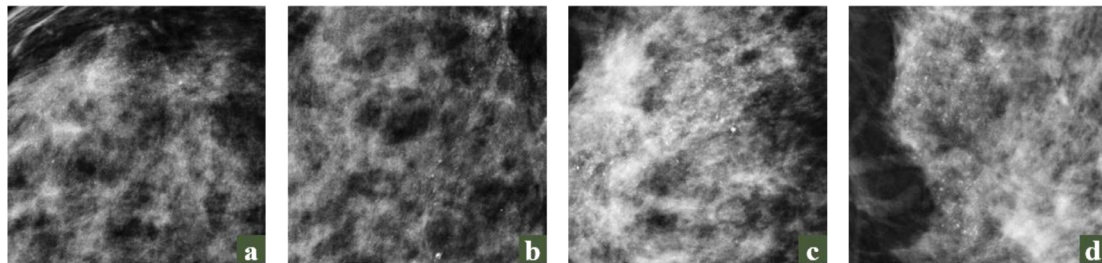


Figure 11. Misclassified ROIs from the benign cases (a-b) and the malignant (c-d).

of their obscurity in feature importance ranking against textural ones during backward elimination - despite their P value suggesting significant difference between their benign and malignant groups in ANOVA ($P < 0.05$) as presented by Figure 10. This obscurity is also demonstrated in Figure 3 as the first three shape2d-related features yielded P values in t -test greater than 0.05. As to the classification, varied machine learning-based methods also influenced the final result. In our present study, LR and SVM proved to be more robust with higher prognostic AUC, generally outclassing KNN and GaussianNB as reported in Table 2.

Table 3 compares the overall performance from the corresponding cross validation of previous benchmark radiomics studies on the breast cancer classification with that of our study. Notably, only Chen et al. [26] among the benchmarks in question had the BI-RADS category of

their image dataset clearly stated. The majority of existing radiomics studies performed their predictive model on medical image cohorts throughout BI-RADS categories 2–5, which decentralized the discriminative target to some extent owing to the visible difference between samples at BI-RADS 2–3 and those at BI-RADS 5. In contrast to previous studies, our radiomics signature model exhibits certain competence by tan AUC score of 0.906 and handled a more challenging task, particularly in regard to pure mammographic microcalcifications that are suspicious for or highly suggestive of malignancy, namely with BI-RADS equal to 4.

There are some drawbacks and limitations in our project that require modifications. Figure 11 displays four misclassified samples from ROIs pathologically confirmed as benign and malignant respectively. Despite their inappreciable distinction, possible causes and further suggestions

Table 3 Comparison of the validation performance with the previous benchmark studies

Sources	Dataset/Size (n)	Feature type	AUC (95% CI)	ACC
Li et al. (2018) [10]	Non-public/388	Radiomics (energy, entropy, etc.) + Clinical statistics (age, the status of lymph node metastasis, etc.)	0.72 (0.63–0.81)	-
Chen et al. (2019) [26]	Non-public/81	Mammographical features pixel intensity-based metrics) + MRI features (kinetic metrics)	0.903	0.815
Fanizzi et al. (2020) [27]	BCDR [29]/130	Multiscale wavelet textures	0.921	0.885
Chen et al. (2015) [28]	DDSM [30]/300	Topology (graph)	0.91±0.05	0.852±0.057
George et al. (2019) [31]	DDSM/289	Topology (chain)	0.892±0.008	0.865±0.013
Suhail et al. (2019) [32]	DDSM/288	Topology (tree)	-	0.79
This work	Non-public/463	Textures (first-order statistics, etc.) + Shapes (Elongation, etc.)	0.906 (0.834–0.969)	0.787 (0.680–0.855)

AUC: area under the curve; ACC: accuracy.

are concluded as follows. Firstly, it was labor-consuming and clinician-dependent to manually annotate the ROI areas as we carried out in the study, thereafter calling for the prospective auto-detection techniques on the fine morphology and distribution of MCs. Secondly, the absence of shape signatures was at variance with empirical knowledge from radiologists, implying that more morphological statistics should be supplementally scrutinized. Thirdly, PRBF performed in this study was not competent enough to take more correlation other than linearity into consideration during redundancy elimination requiring more elaborated study on other non-linear relationship. Finally, considering that the AUC was down from 0.906 (95% CI, 0.834–0.969) to 0.810 (95% CI, 0.760–0.960) with the participation of our independent dataset, large-scale future cases from more centers should be enrolled to further evaluate and adjust our model.

In conclusion, our study extracted radiomics features depicting the distributional characteristics of breast MCs followed by XGBoost-embedded RFE as feature selection to constitute a predictive radiomics signature and to achieve moderate performance. Besides, the studied interpretable textural signature more or less disclosed some underlying relevance to malignant lesions as reported by our probe into their performance of PPV/SE and NPV/SP.

Conflicts of interest statement

The authors declare that they have no competing interests.

Funding

This work was supported in part by the State's Key Project of Research and Development Plan (Grant Nos. 2017YFC0109202 and 2017YFA0104302), in part by the National Natural Science Foundation (Grant No. 61871117), in part by Science and Technology Program of Guangdong (Grant No. 2018B030333001).

Author contributions

Yanni Jiang and Yang Chen contributed to the conceptualization and funding acquisition. Liyu Zhu was responsible for the data curation. Formal analysis and Methodology were jointly performed by Min Li and Jianan He. Min Li wrote the manuscript. Guangquan Zhou, Yanni Jiang and Yang Chen edited and reviewed the manuscript.

Acknowledgements

The authors gratefully acknowledge the First Affiliated Hospital of Nanjing Medical University for their clinical digital mammography dataset support.

References

- [1] Breast Cancer Early Detection and Diagnosis. American Cancer Society; 2021. Available from <https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html> (Accessed 16 Mar 2020).
- [2] Bonfiglio R, Scimeca M, Toschi N, et al. Radiological, histological and chemical analysis of breast microcalcifications: diagnostic value and biological significance. *J Mammary Gland Biol Neoplasia* 2018;23(1-2):89–99. doi:10.1007/s10911-018-9396-0.
- [3] Ali MA, Czene K, Hall P, et al. Association of microcalcification clusters with short-term invasive breast cancer risk and breast cancer risk factors. *Scientific reports* 2019;9(1):1–8. doi:10.1038/s41598-019-51186-w.
- [4] Agide FD, Sadeghi R, Garmaroudi G, et al. A systematic review of health promotion interventions to increase breast cancer screening uptake: from the last 12 years. *Eur J Public Health* 2018;28(6):1149–55. doi:10.1093/eurpub/ckx231.
- [5] The american society of clinical oncology (ASCO) Breast Cancer: Screening; 2021. Available from <https://www.cancer.net/cancer-types/breast-cancer/screening> (Accessed 9 Mar 2021).
- [6] Hernández PA, Estrada TT, Pizarro AL, et al. Breast calcifications: description and classification according to bi-rads 5th edition. *Rev Chil Radiol* 2016;22:80–91. doi:10.1016/j.rchira.2016.06.004.
- [7] Jiang Y, Lou J, Wang S, et al. Evaluation of the role of dynamic contrast-enhanced MR imaging for patients with BI-RADS 3–4 microcalcifications. *PLoS One* 2014;9(6):e99669. doi:10.1371/journal.pone.0099669.
- [8] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278(2):563–77. doi:10.1148/radiol.2015151169.
- [9] Li H, Mendel KR, Lan L, et al. Digital mammography in breast cancer: additive value of radiomics of breast parenchyma. *Radiology* 2019;291(1):15–20. doi:10.1148/radiol.2019181113.
- [10] Li J, Song Y, Xu S, et al. Predicting underestimation of ductal carcinoma in situ: a comparison between radiomics and conventional approaches. *International journal of computer assisted radiology and surgery* 2019;14(4):709–21. doi:10.1007/s11548-018-1900-x.
- [11] Parekh VS, Jacobs MA. Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI. *NPJ breast cancer* 2017;3(1):1–9. doi:10.1038/s41523-017-0045-3.
- [12] Parekh VS, Macura KJ, Harvey SC, et al. Multiparametric deep learning tissue signatures for a radiological biomarker of breast cancer: Preliminary results. *Medical physics*. 2020;47(1):75–88. doi:10.1002/mp.13849.
- [13] Hjelm RD, Plis SM, Calhoun VC. Variational autoencoders for feature detection of magnetic resonance imaging data. *arXiv preprint arXiv:1603.06624*. 2016.
- [14] Parekh VS, Jacobs MA. Deep learning and radiomics in precision medicine. Expert review of precision medicine and drug development. 2019;4(2):59–72. doi:10.1080/23808993.2019.1585805.
- [15] Afshar P, Mohammadi A, Plataniotis KN, et al. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. *IEEE Signal Processing Magazine* 2019;36(4):132–60. doi:10.1109/MSP.2019.2900993.
- [16] Fanizzi A, Basile TM, Losurdo L, et al. Ensemble discrete wavelet transform and gray-level co-occurrence matrix for microcalcification cluster classification in digital mammography. *Applied Sciences* 2019;9(24):5388. doi:10.3390/app9245388.
- [17] Jian W, Sun X, Luo S. Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform. *Biomedical engineering online* 2012;11(1):1–2. doi:10.1186/1475-925X-11-96.
- [18] Saad G, Khadour A, Kanafani Q. ANN and AdaBoost application for automatic detection of microcalcifications in breast cancer. *The Egyptian Journal of Radiology and Nuclear Medicine* 2016;47(4):1803–14. doi:10.1016/j.ejrm.2016.08.020.
- [19] Guo G, Wang H, Bell D, et al. KNN model-based approach in classification. In: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", 2888. Berlin, Heidelberg: Springer; 2003. p. 986–96. doi:10.1007/978-3-540-39964-3_62.
- [20] Collins M, Schapire RE, Singer Y. Logistic regression, AdaBoost and Bregman distances. *Machine Learning* 2002;48(1):253–85. doi:10.1023/A:1013912006537.
- [21] Jahromi AH, Taheri M. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In: 2017 Artificial Intelligence and Signal Processing Conference (AISP). IEEE; 2017. p. 209–12. doi:10.1109/AISP.2017.8324083.
- [22] Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer research* 2017;77(21):e104–7. doi:10.1158/0008-5472.CAN-17-0339.
- [23] Biesiada J, Duch W. Feature selection for high-dimensional data: A kolmogorov-smirnov correlation-based filter. In: Computer Recognition Systems. Berlin, Heidelberg: Springer; 2005. p. 95–103. doi:10.1007/3-540-32390-2_9.

- [24] Erişti H, Yıldırım Ö, Erişti B, et al. Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. *International Journal of Electrical Power & Energy Systems* 2013;49:95–103. doi:[10.1016/j.ijepes.2012.12.018](https://doi.org/10.1016/j.ijepes.2012.12.018).
- [25] Zhang Q, Liu P, Wang X, et al. StackPDB: predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier. *Applied Soft Computing* 2021;99:106921. doi:[10.1016/j.asoc.2020.106921](https://doi.org/10.1016/j.asoc.2020.106921).
- [26] Chen S, Guan X, Shu Z, et al. A New Application of Multimodality Radiomics Improves Diagnostic Accuracy of Nonpalpable Breast Lesions in Patients with Microcalcifications-Only in Mammography. *Medical science monitor: international medical journal of experimental and clinical research* 2019;25:9786. doi:[10.12659/MSM.918721](https://doi.org/10.12659/MSM.918721).
- [27] Fanizzi A, Basile TM, Losurdo L, et al. A machine learning approach on multi-scale texture analysis for breast microcalcification diagnosis. *BMC bioinformatics* 2020;21(2):1–1. doi:[10.1186/s12859-020-3358-4](https://doi.org/10.1186/s12859-020-3358-4).
- [28] Chen Z, Strange H, Oliver A, et al. Topological modeling and classification of mammographic microcalcification clusters. *IEEE transactions on biomedical engineering* 2014;62(4):1203–14. doi:[10.1109/TBME.2014.2385102](https://doi.org/10.1109/TBME.2014.2385102).
- [29] Lopez MG, Posada N, Moura DC, et al. BCDR: a breast cancer digital repository. In: *15th International conference on experimental mechanics*; 2012. p. 1215.
- [30] Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data* 2017;4(1):1–9. doi:[10.1038/sdata.2017.177](https://doi.org/10.1038/sdata.2017.177).
- [31] George M, Chen Z, Zwiggelaar R. Multiscale connected chain topological modelling for microcalcification classification. *Computers in biology and medicine* 2019;114:103422. doi:[10.1016/j.compbiomed.2019.103422](https://doi.org/10.1016/j.compbiomed.2019.103422).
- [32] Suhail Z, Zwiggelaar R. Multi-scale Tree-Based Topological Modelling and Classification of Micro-calcifications. In: *Annual Conference on Medical Image Understanding and Analysis*. Cham: Springer; 2019. p. 48–58. doi:[10.1007/978-3-030-39343-4_5](https://doi.org/10.1007/978-3-030-39343-4_5).