

Mammogram Quantitative Features Associated with Histological High-Grade Breast Cancer

Bonou Malomon Aime^{1,*}, Topanou Roland Guy Boniface¹, Hounsossou Cocou Hubert¹, Gbossa Eddy Hans¹, Dossou Julien¹, Biaou Olivier²

¹Non-Communicable Diseases and Cancer Research Unit, Laboratory of Applied Biology Research, Polytechnic School of Abomey-Calavi, University of Abomey-Calavi, Abomey-Calavi, Benin

²Medical Imaging Unit of "National Hospital and University center H. K. Maga", Cotonou, Benin

Email address:

malombonou@yahoo.fr (Bonou M. A.), topanourol@yahoo.fr (Topanou R. G. B.), hchube1963@gmail.com (Hounsossou C. H.), juju_dos@yahoo.fr (Gbossa E. H.), mahoutondji1@gmail.com (Dossou J.), biaoulivier@gmail.com (Biaou O.)

*Corresponding author

To cite this article:

Bonou Malomon Aime, Topanou Roland Guy Boniface, Hounsossou Cocou Hubert, Gbossa Eddy Hans, Dossou Julien, Biaou Olivier. Mammogram Quantitative Features Associated with Histological High-Grade Breast Cancer. *International Journal of Medical Imaging*. Vol. 8, No. 3, 2020, pp. 39-44. doi: 10.11648/j.ijmi.20200803.11

Received: June 15, 2020; **Accepted:** July 8, 2020; **Published:** July 17, 2020

Abstract: High grade breast cancer is recognized as more aggressive cancer type and is the worst survival prognostic. To explore the association of quantitative features extracted from mammograms with histological high-grade breast cancer. We conducted a retrospective study using an open source data got from *figshare repository*. These anonymized data were collected and used for a study approved by the institutional review board. Cranio-Caudal (CC) and Medio-lateral (MLO) mammograms and their tumor segmented images from 66 patients subdivided in two groups high histological grade (n=23) low-grade (low and intermediate, n=41). From breast cancer image segmentation, we extracted 480 features using *python* software radiomics package *Pyradiomics* 2.2. With the features extracted from CC and MLO images, we used them separately for histological high-grade breast, relevant feature selection. We performed univariate feature selection based on ANOVA test using machine learning *python* package: *sklearn*. A feature was considered relevant when *P* value is at least 0.05. At the end we represented the boxplot of the distribution of the low-and high-grade subject using each relevant feature selected. Twenty (20) CC images features were selected, seventeen (17) were based on wavelets and three (3) were from original image. Their *p* values were ranged between 0.017 and 0.046. In the case of MLO features, four (04) relevant features were exclusively based on wavelets with 0.046 as the maximum *p*-value and 0.006 as minimum. These results suggested mammogram quantitative feature based on wavelets will be useful for high-grade breast cancer identification on mammographic image. In this study we explored the association between IBSI 2D quantitative features from mammogram with the histological high-grade breast cancer. Finally, we recorded twenty (20) relevant features from CC projection and four for MLO mammogram projection. Wavelets based features were more represented in relevant quantitative feature.

Keywords: Quantitative Feature, Mammography, High Grade Cancer, Breast

1. Introduction

Breast cancer is the most common cancer in women and a leading cause of cancer death worldwide [1]. Management of breast cancer relies on the availability of robust clinical and pathological prognostic and predictive factors to guide patient decision making and the selection of treatment. Histological grade is one of important prognostic factor. It is

based on the degree of differentiation of the tumor tissue and based on the evaluation of three morphological features: (a) degree of tubule or gland formation, (b) nuclear pleomorphism, and (c) mitotic count. It is used to categorize breast cancer patient in three clinical groups grade I (low), grade II (intermediate) and grade III (high) [2]. High grade breast cancer is recognized as more aggressive cancer type and is the worst survival prognostic [3, 4].

Motivated by the high-grade breast cancer characteristic

description on medical image, Lamb *et al* found in their study that its classical appearance is a mass with round shape on mammography [5]. SHIN *et al.* had also attempted to described morphological aspect on mammogram because mammography is one of the primary breast imaging modalities used in breast cancer diagnosis. They found that having Fairly slow developing, grade I tumors (low grade) and grade II tumors (intermediate grade) present a stroma reaction resulting in imaging by spicules while high grade with rapid evolution, do not develop a stroma reaction and have a round shape [6]. These finding suggested that the high-grade breast cancer presents a particular characteristic on mammogram.

With the radiomic advent, quantitative features were used to describe the breast cancer characteristic. Those features are created from computer analysis of images, either alone or with the guidance of a reader to assist in segmentation. These imaging variables must be well-defined to limit inter- or intra-observer variability [7]. In this context several features and different software were proposed for using in tumor biological phenotyping [8]. Among all its many proposals, the Image Biomarker Standardization Initiative (IBSI) [9] had performed a selection in order to contribute to the reproducibility research. Some software such as *Lifex*, *Pyradiomics* had been developed based on IBSI.

In this study, we explored the association of IBSI quantitative features extracted from mammograms with histological high-grade breast cancer. We used *pyradiomics* for feature extraction and univariate feature selection method for relevant feature identification.

2. Material and Methods

2.1. Patients Data

We conducted a retrospective study using an open source

data downloaded from figshare repository [10]. These anonymized data were collected and used withing a study that was approved by the institutional review board. It aimed to establish an association between digital mammography radiomic and breast cancer OncotypeDX and PAM50 (Prosigna Breast Cancer Prognostic Gene Signature Assay) recurrence scores. The study englobes a total of 71 breast cancer cases with clinicopathologic information (age, Tumor size, regional lymph Node status, and distant Metastases staging, Estrogen Receptor, Progesterone Receptor, and Human epidermal growth factor receptor 2 status), digital mammograms (Cranio-caudal CC and Medio-Lateral Oblique MLO), microarray data and tumor segmentation on mammograms images. A digital mammography system (Selenia, Hologic, Bedford, MA), with an automatic intensity adjustment was used to acquire mammogram of 70 microns per pixel and 12-bits grayscale for codification. Manuel segmentation of tumors were performed by an experienced breast radiologist [11]. Five (05) patients were excluded because their histological grading status is missing. Amongst the sixty-six (66) patients of our cohort, twenty-three (23) were high grade, three (03) were intermediate grade, and six (06) had low histological grading status with respective mean age of 50, 50.5 and 54 years.

2.2. Radiomic Morphological Features Extraction

Dicom Mammograms and tumor segmentation images were decompressed with the open source Dicom viewer software *MicroDicom 2.7.9* Tumor segmented images were rescaled between 0 and 1 grayscale with the python package *ITKsimple* [12]. The tumor region segmented from each mammogram view was used to extracted 480 features using *python* software radiomics package *Pyradiomics 2.2* [13] (Figure 1).

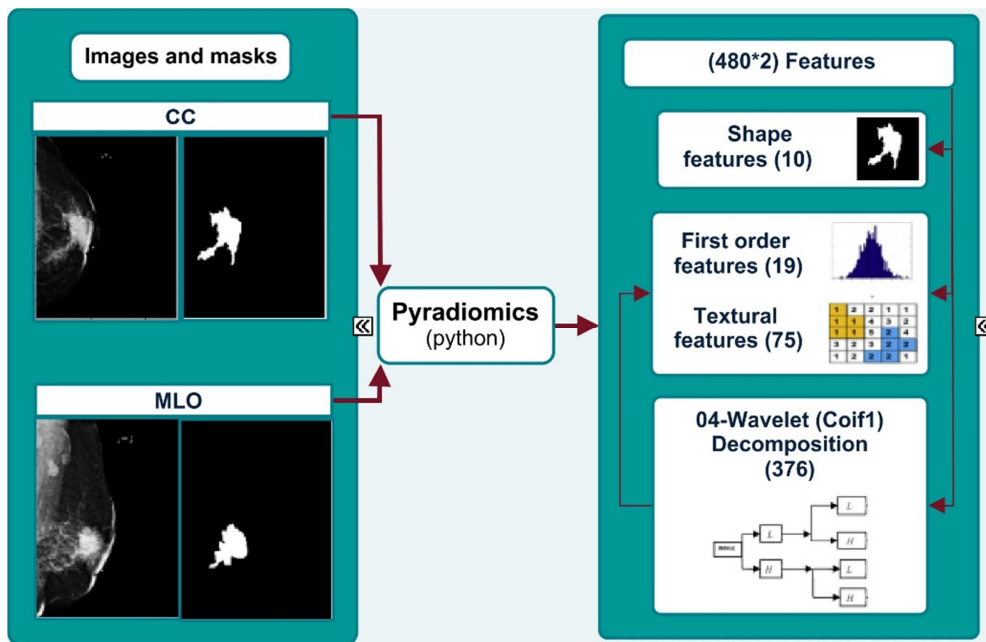


Figure 1. Features extract flowchart using the pyradiomics package.

With the features extracted from CC and MLO images, we used them separately for histological high-grade breast, relevant feature selection.

2.3. Statistical Analysis

We performed univariate feature selection based on ANOVA test using machine learning python package: sklearn [14]. This method had been used by some authors for relevant features selection [15, 16]. A feature was considered relevant when P value is at least 0.05. At the end we represented the boxplot of the distribution of the low and high-grade subject using each relevant feature selected.

3. Results

Table 1 presents CC and MLO mammograms quantitative features which were relevant with high-grade breast cancer. Among twenty (20) CC images features selected, 17 based on wavelets and three (3) were from original image. Their p values were ranged between 0.017 and 0.046. In the case of MLO features, four (04) relevant features were exclusively based on wavelets with 0.046 as the maximum p-value and 0.006 as minimum.

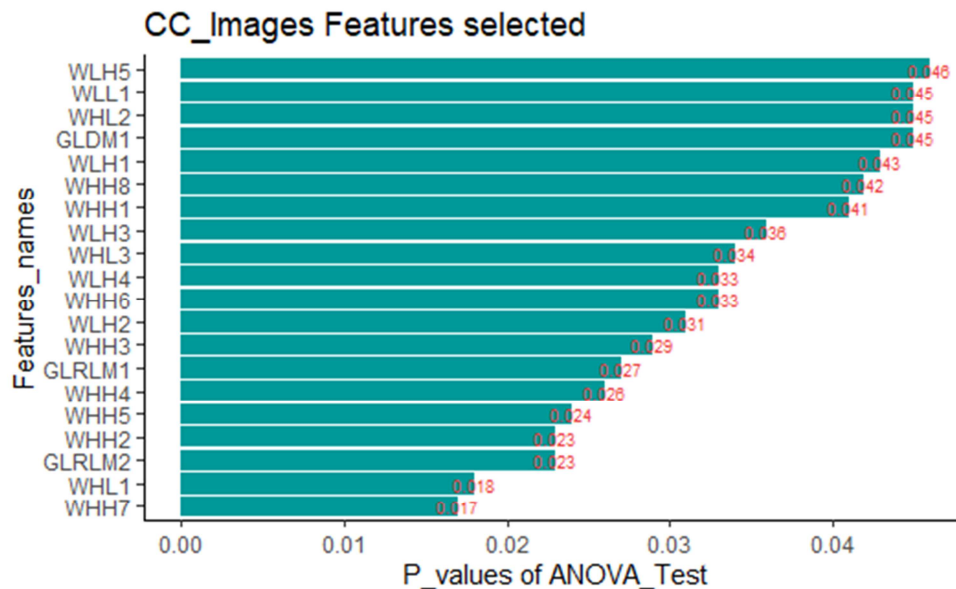


Figure 2. Twenty (20) CC features selected and their p_values.

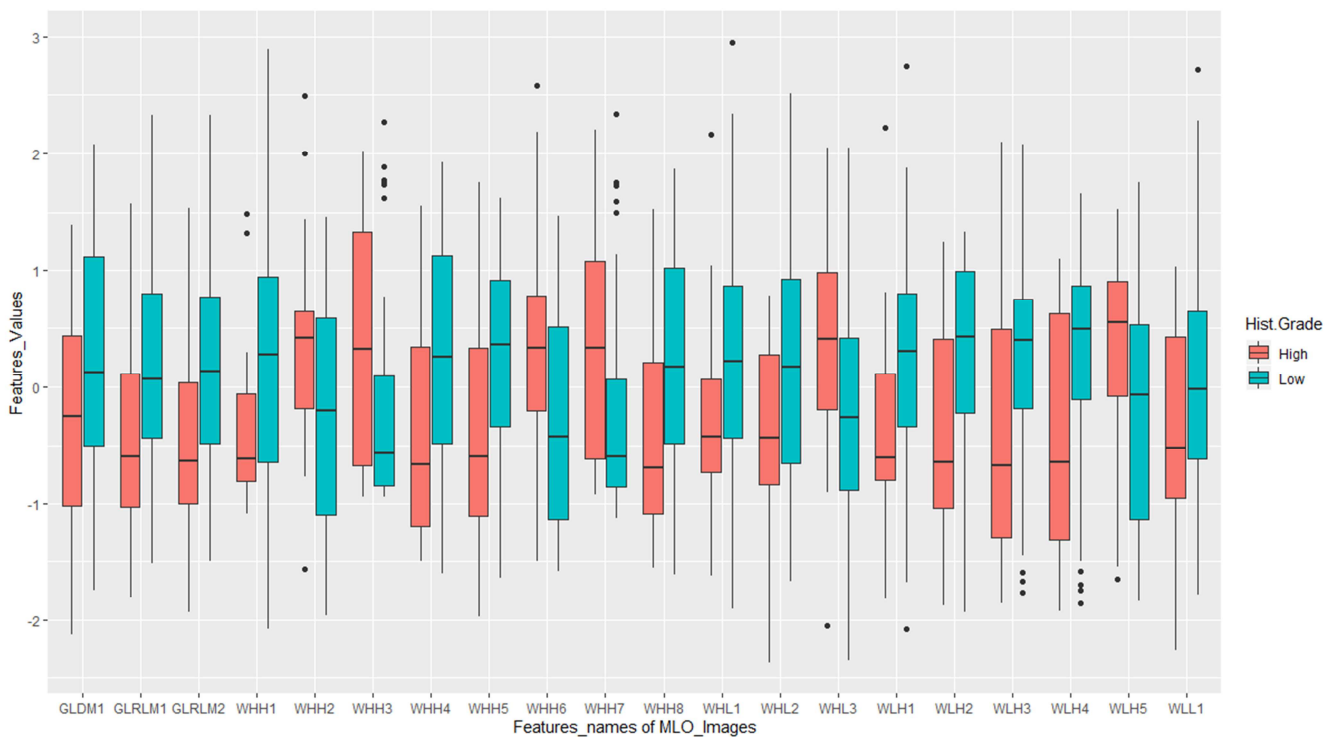


Figure 3. Four (04) MLO features selected and their p_values.

Table 1. CC and MLO mammograms features selected.

Features Selected	
CC Features Names	Abreviation
original_gldm_SmallDependenceHighGrayLevelEmphasis	GLDM1
original_glrlm_RunLengthNonUniformityNormalized	GLRLM1
original_glrlm_ShortRunEmphasis	GLRLM2
wavelet-HH_firstorder_Maximum	WHH1
wavelet-HH_firstorder_Median	WHH2
wavelet-HH_firstorder_Minimum	WHH3
wavelet-HH_firstorder_Range	WHH4
wavelet-HH_glszm_GrayLevelVariance	WHH5
wavelet-HH_glszm_LowGrayLevelZoneEmphasis	WHH6
wavelet-HH_glszm_SmallAreaLowGrayLevelEmphasis	WHH7
wavelet-HH_glszm_ZoneEntropy	WHH8
wavelet-HL_firstorder_Median	WHL1
wavelet-HL_glszm_GrayLevelNonUniformityNormalized	WHL2
wavelet-HL_glszm_GrayLevelVariance	WHL3
wavelet-LH_firstorder_Maximum	WLH1
wavelet-LH_firstorder_Median	WLH2
wavelet-LH_glszm_GrayLevelNonUniformityNormalized	WLH3
wavelet-LH_glszm_GrayLevelVariance	WLH3
wavelet-LH_glszm_SmallAreaEmphasis	WLH4
wavelet-LL_gldm_SmallDependenceHighGrayLevelEmphasis	WLL1

MLO Features Names	Abreviation
wavelet-LH_firstorder_Skewness	WLH1
wavelet-HH_firstorder_Median	WHH1
wavelet-HH_firstorder_Skewness	WHH2
wavelet-HH_glcmm_ClusterShade	WHH3

**Figure 4.** Boxplot of CC images features selected values in High- and Low-grade groups.

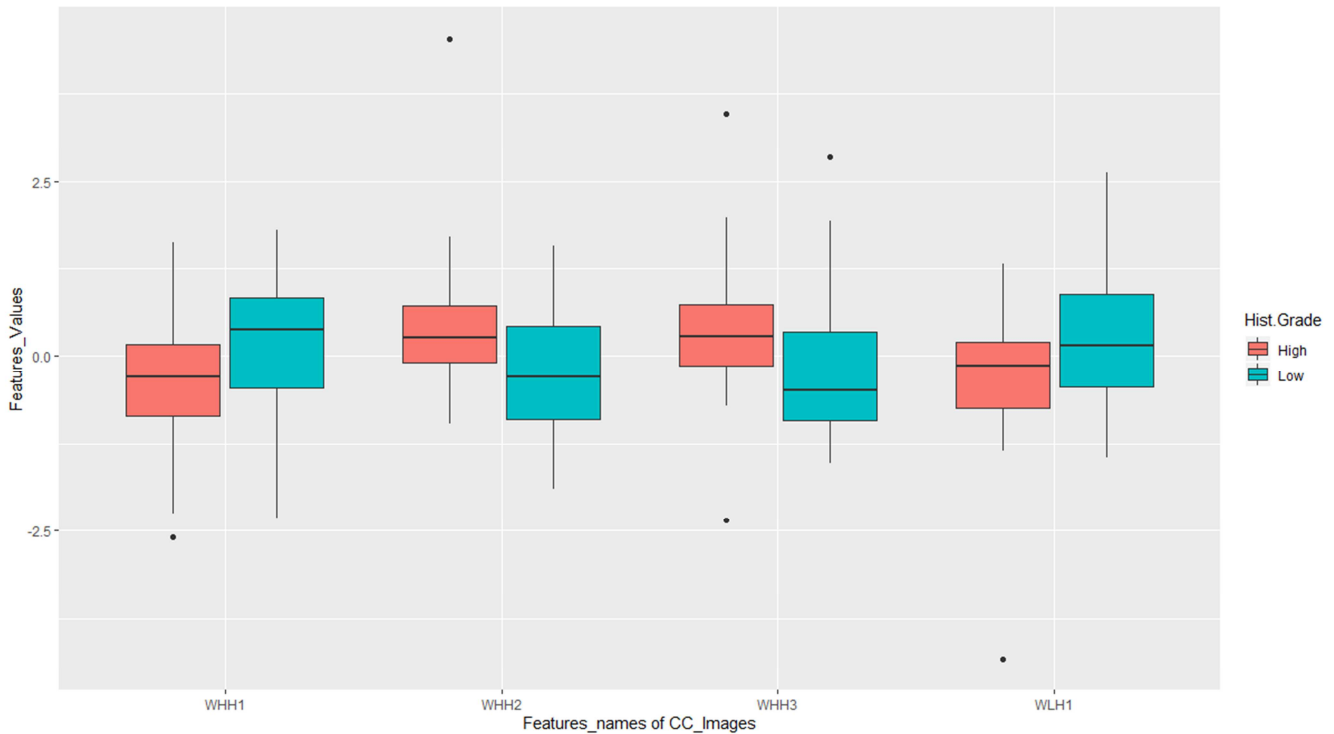


Figure 5. Boxplot of MLO images features selected values in High- and Low-grade groups.

CC mammogram features selected p-values were ranged between 0.017 and 0.046 (Figure 2). Regarding the case of MLO features, p-values were ranged between 0.006 and 0.046 (Figure 3). These results suggested mammogram quantitative feature based on wavelets will be useful for high-grade breast cancer identification on mammographic image.

The differentiation capacity of features selected were demonstrated with low and high-grade boxplot using each feature (figures 4&5). Means of low and high-grade groups in according to each feature were statistically different.

4. Discussion

In this study, we assessed the association of IBSI 2D quantitative imaging feature with high grade breast cancer. Our works represent a preliminary study of high-grade breast cancer identification using quantitative imaging features extracted from mammogram. We recorded that some features specially those based on wavelets were relevant to histological grade. These results suggested mammogram quantitative feature based on wavelets will be useful for high-grade breast cancer identification on mammographic image. This finding is agree with Huang *et al.* study which used quantitative features, extracted from PET and MRI, for breast cancer histological grading decoding [17]. They extracted 104 features from both modalities images for each patient. Ten (10) features were selected like the best relevant for the breast cancer histological grade prediction. Those relevant features were only from PET images and allowed to achieve the best predictive performance (AUC=0.76). Also

Fan *et al.* used in recent study used the same kind of quantitative feature extracted from dynamic contrast-enhanced magnetic resonance image (DCE-MRI) for breast cancer histological grade prediction [18]. They have selected 24 optimal features over the total of 130 image features extracted to achieve their optimal predictive performance (AUC=0.82). After those previous studies, our study showed that mammography which is morphological medical imaging modality could be used for high grade breast cancer decoding. This finding could be helpful in developing countries, because mammography is more accessible than MRI and PET.

There are limitations to this study. Manual segmentation does not allow to find with more accuracy breast cancer margin on all mammograms mainly of young subject who has more dense breast. Mammography is planar medical imaging modality leading the superposition of several glandular structures with breast tumor. These two realities could contribute sometimes to the inaccessibility of the real breast cancer margin. The small size of our cohort is also a limitation.

5. Conclusion

In this study we explored the association between IBSI 2D quantitative features from mammogram with the histological high-grade breast cancer. Finally, we recorded twenty (20) relevant features from CC projection and four for MLO mammogram projection. Wavelets based features were more represented in relevant quantitative feature.

References

- [1] Bray F, Ferlay J, Soerjomataram I, et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68: 394-424. <https://doi.org/10.3322/caac.21492>
- [2] Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19: 403-410. <https://doi.org/10.1111/j.1365-2559.1991.tb00229.x>
- [3] Rakha EA, El-Sayed ME, Powe DG, et al (2008) Invasive lobular carcinoma of the breast: response to hormonal therapy and outcomes. *Eur J Cancer* 44: 73-83. <https://doi.org/10.1016/j.ejca.2007.10.009>
- [4] Rakha EA, El-Sayed ME, Lee AHS, et al (2008) Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol* 26: 3153-3158. <https://doi.org/10.1200/JCO.2007.15.5986>
- [5] Lamb PM, Perry NM, Vinnicombe SJ, Wells CA (2000) Correlation Between Ultrasound Characteristics, Mammographic Findings and Histological Grade in Patients with Invasive Ductal Carcinoma of the Breast. *Clinical Radiology* 55: 40-44. <https://doi.org/10.1053/crad.1999.0333>
- [6] Shin HJ, Kim HH, Huh MO, et al (2011) Correlation between mammographic and sonographic findings and prognostic factors in patients with node-negative invasive breast cancer. *BJR* 84: 19-30. <https://doi.org/10.1259/bjr/92960562>
- [7] Grimm LJ, Mazurowski MA (2020) Breast Cancer Radiogenomics: Current Status and Future Directions. *Academic Radiology* 27: 39-46. <https://doi.org/10.1016/j.acra.2019.09.012>
- [8] Court LE, Fave X, Mackin D, et al (2016) Computational resources for radiomics. *Translational Cancer Research* 5: 340-348. <https://doi.org/10.21037/8409>
- [9] Zwanenburg A, Leger S, Vallières M, Löck S (2019) Image biomarker standardisation initiative. *arXiv: 161207003 [cs, eess]*.
- [10] Trevino V (2018) Breast Cancer Images & Segmentation - Correlation of Gene Expression Subtypes and Image Features
- [11] Tamez-Peña J-G, Rodriguez-Rojas J-A, Gomez-Rueda H, et al (2018) Radiogenomics analysis identifies correlations of digital mammography with clinical molecular signatures in breast cancer. *PLOS ONE* 13: e0193871. <https://doi.org/10.1371/journal.pone.0193871>
- [12] Yaniv Z, Lowekamp BC, Johnson HJ, Beare R (2018) SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *Journal of Digital Imaging* 31: 290-303. <https://doi.org/10.1007/s10278-017-0037-8>
- [13] van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* 77: e104-e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- [14] Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12: 2825-2830.
- [15] Omer Fadl Elssied N, Ibrahim O, Hamza Osman A (2014) A Novel Feature Selection Based on One-Way ANOVA F-Test for E-Mail Spam Classification. *Research Journal of Applied Sciences, Engineering and Technology* 7: 625-638. <https://doi.org/10.19026/rjaset.7.299>
- [16] Olalolu AM (2016) A Feature Selection Based on One-Way-Anova for Microarray Data Classification. 30-35.
- [17] Huang S, Franc BL, Harnish RJ, et al (2018) Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer* 4: <https://doi.org/10.1038/s41523-018-0078-2>
- [18] Fan M, Liu Z, Xie S, et al (2019) Integration of dynamic contrast-enhanced magnetic resonance imaging and T2-weighted imaging radiomic features by a canonical correlation analysis-based feature fusion method to predict histological grade in ductal breast carcinoma. *Phys Med Biol* 64: 215001. <https://doi.org/10.1088/1361-6560/ab3fd3>