Spatial Distribution Analysis of Novel Texture Feature Descriptors for Accurate Breast Density Classification

Haipeng Li [1,*] , Ramakrishnan Mukundan [1] and Shelley Boyd [2]

1   Department of Computer Science and Software Engineering, University of Canterbury,
    Christchurch 8140, New Zealand; mukundan@canterbury.ac.nz
2   Canterbury Breastcare, St. George's Medical Centre, Christchurch 8014, New Zealand;
    shelley.boyd@pacificradiology.com
*   Correspondence: haipeng.li@canterbury.ac.nz

Abstract: Breast density has been recognised as an important biomarker that indicates the risk of developing breast cancer. Accurate classification of breast density plays a crucial role in developing a computer-aided detection (CADe) system for mammogram interpretation. This paper proposes a novel texture descriptor, namely, rotation invariant uniform local quinary patterns (RIU4-LQP), to describe texture patterns in mammograms and to improve the robustness of image features. In conventional processing schemes, image features are obtained by computing histograms from texture patterns. However, such processes ignore very important spatial information related to the texture features. This study designs a new feature vector, namely, K-spectrum, by using Baddeley's K-inhom function to characterise the spatial distribution information of feature point sets. Texture features extracted by RIU4-LQP and K-spectrum are utilised to classify mammograms into BI-RADS density categories. Three feature selection methods are employed to optimise the feature set. In our experiment, two mammogram datasets, INbreast and MIAS, are used to test the proposed methods, and comparative analyses and statistical tests between different schemes are conducted. Experimental results show that our proposed method outperforms other approaches described in the literature, with the best classification accuracy of 92.76% (INbreast) and 86.96% (MIAS).

Keywords: breast density classification; mammography; local quinary patterns; spatial distribution analysis; texture features

1. Introduction

Mammography screening is the most reliable and widely used method for observing breast lesions and evaluating breast health, and has been demonstrated to be effective for preventing breast cancer in its early detection stage. Breast density classification relates to measuring the amount of radiodense tissue (i.e., fibroglandular tissue) in mammograms and producing an important biomarker for indicating the risk of developing breast cancer in the future. Related clinical work shows that women with dense breast could face four- to six-fold higher risk than other women with low density [1]. Different metrics have been proposed and used to classify breast density, including six-class-categories (SCC) [2], Wolfe's four categories [3], and breast imaging-reporting and data system (BI-RADS) [4]. BI-RADS criterion proposed by the American College of Radiology (ACR) has been widely used in clinical applications and includes four density categories: fatty (i.e., BI-RADS I), scattered density (i.e., BI-RADS II), heterogeneously dense (i.e., BI-RADS III), and extremely dense (i.e., BI-RADS IV). However, current clinical workflow heavily depends on radiologists' subjective visual assessment, which is known to cause both inter- and intraobserver disagreements [5].

AI-driven classification algorithms have the potential to contribute to radiologists' workflow by offering a secondary perspective of evaluating breast density. An automated

breast density estimation method usually contains the following key steps: (1) extracting discriminative image features related to target density labels, (2) feature selection and parameter optimisation, (3) training a machine-learning-based classification model, and (4) predicting density labels for test images. As such, extracting effective features in mammograms plays an important role in obtaining accurate classification results. For breast density classification, relevant features can be broadly grouped into three types: heterogeneous features, deep-learning-based features, and texture descriptors.

Methods using heterogeneous features try to collect multiple image features, including image intensity, morphological features, entropy, texture features, etc., and combine them into one feature vector for classifying mammograms. Mario et al. [6] extracted mammogram image features based on image intensity, histograms, and grey-level co-occurrence matrix (GLCM). They also used Wrappers to select relevant features for improving the classification accuracy. Li et al. [7] extracted 137 pixel-level features containing intensity, GLCM, and morphological features, to group pixels into fatty or dense classes. In [8], intensity, texture, and morphological features were combined to develop a binary classification framework similar to that given in [7]. Statistical features including mean, standard deviation, smoothness, third moment, uniformity, and entropy were extracted in [9] to classify mammographic density into three groups. Qu et al. [10] proposed a fuzzy–rough refined image processing method to enhance local image regions and extracted GLCM-based statistical features for classifying breast density. Tzikopoulos et al. [11] extracted image features based on intensity and fractal texture, and SVM was used to classify mammograms into three categories. A multiresolution histogram method was used to analyse texture features in [12], and mammograms were classified into three density categories using a directed acyclic graph (DAG)–SVM classifier. In [13], different texture feature sets were investigated separately, including LBP, local grey-level appearance (LGA), textons (MR8 texton and image-patch texton), and basic image features (BIF). Their experimental results indicated that image-patch texton features performed better for four BI-RADS density categories classification. A lattice-based approach [14] was proposed to extract statistical and structural features for analysing parenchymal texture in mammograms.

Deep-learning methods have been used recently to analyse medical images for both classification and segmentation tasks with promising results. For evaluating breast density, convolutional neural networks (CNN) were applied to manage binary classification tasks. An AlexNet-based CNN model was proposed in [15] to distinguish two BI-RADS categories ('scattered density' and 'heterogeneously dense'). Ahn et al. [16] designed a CNN architecture to extract features from mammogram patches and to classify them into dense and fatty groups. In [17], a fully convolutional network (FCN) was used to segment breast region and fibroglandular areas, based on which a percentage density was estimated. A deep convolutional neural network (DCNN) was used to classify mammographic pixels into fatty or dense class in [7]. Li et al. [18] proposed a CNN-based radiomics method combined with dilated convolutions and attention mechanisms to extract high-throughput features that were used to classify mammograms into four BI-RADS density categories. One limitation of applying deep-learning methods for mammogram classification task is that they require a huge number of training images with accurate annotations by clinicians [19].

Recent study using texture descriptors such as local binary patterns (LBP) and their variants have shown promising classification performance with over 80% accuracy. The local binary patterns (LBP) method was first proposed in [20] to describe image texture patterns. Owing to its simplicity and efficiency, LBP has been studied widely and new variants were proposed for extracting texture features and classifying medical images. In [21], LBP was extended to elliptical LBP (ELBP) and mean-elliptical LBP (M-ELBP) by considering various neighbourhood topologies and different local region scales. M-ELBP presented a desirable classification result (77.38 ± 1.06) on an MIAS dataset. Tan et al. [22] proposed local ternary patterns (LTP) using a three-value encoding approach compared to two-value encoding of LBP. Then, Rampun et al. [23] extracted LTP-based texture features to classify MIAS mammograms into four BI-RADS categories. Nanni et al. [24] proposed

local quinary patterns (LQP) by extending LBP from a binary encoding system to a five-value encoding system, and used three different medical image classification tasks to test this new texture descriptor. Subsequently LQP was investigated and extended with multiresolution and multiorientation schemes in [25] to classify mammographic density. Their experimental results have shown that the use of LQP-based texture features helps improve the overall classification accuracy. In a recent study, Rampun et al. [26] tried the local septenary patterns (LSP) method by using seven-value encoding approach to further improve the classification performance. Their experimental results demonstrated that classification accuracy was slightly improved by using LSP compared to LQP (80.5 ± 9.2 vs. 80.1 ± 10.5 on INBreast), but this improvement was not statistically significant ($p = 0.45$).

Through reviewing relevant feature extraction methods above, we can find that multilabel (three or four) classification is challenging work with only a few reported results surpassing 80% accuracy. The work in [11,12,21] reported accuracy around 77% for classifying MIAS mammograms into three density categories; experimental results in [6,13] showed 79% and 75% accuracy for four categories classification on the same dataset. A recent study based on texture analysis using LBP variants has shown promising classification performance with over 80% accuracy. LTP, LQP and LST were used in [23,25,26] and obtained the accuracy of 82%, 86%, and 83%, respectively, with four density categories classification on MIAS. When testing the INbreast dataset, classification accuracy of 82.02% and 80.10% were seen by using LQP-based texture features [25,26].

However, the use of LBP and its variants for feature extraction have some limitations that can affect the accuracy of results: (1) For capturing more local texture information, increasing the number of neighbouring pixels and a multiscale scheme are usually used, but they lead to an exponential increase in the number of features (i.e., high feature dimensionality). (2) The conventional rotation invariant uniform method (i.e., 'RIU2') can reduce the high feature dimensionality, but it also suppresses some important/distinguishing features and lowers the accuracy of classification results. (3) Most approaches in the literature extract texture information by computing histograms. Histograms offer quantity information of features but ignore their spatial distribution characteristics, which may contain other important and complementary information related to mammographic density. (4) Related work in this field did not give much attention to feature optimisation after collecting the initial feature set, while a proper feature selection method could optimise the feature vector and improve the accuracy of results.

Motivated by relevant work and a comparison of their performance characteristics, this study chooses LQP as the base feature extraction method, and two improvement schemes called RIU4-LQP and K-spectrum are proposed to develop a more robust and efficient texture descriptor for mammogram analysis. Focusing on the breast density classification task, the main contributions of this paper include:

1. A novel texture descriptor, namely, rotation invariant uniform local quinary patterns (RIU4-LQP), is proposed by developing a uniform rotation invariant version of LQP and considering a higher number of bit transitions while computing the invariant descriptors. To the best of our knowledge, this is the first study that extends the uniform encoding technique by using the transition number of four.
2. In addition to LQP feature histograms, we address a spatial feature extraction framework using Baddeley's K-inhom function, which outputs a new feature vector called K-spectrum.
3. A new feature space is proposed by concatenating RIU4-LQP histograms and K-spectra and is used in the mammographic density classification model.
4. Machine-learning-based feature selection methods are employed to optimise the initial texture feature set, which does not only reduce the high feature dimensionality but also lead to a better classification performance.
5. We empirically proved the effectiveness of the proposed classification model on two publicly available mammogram datasets. Experimental results indicate that the extra

spatial distribution features considered in this work are beneficial to the improvement of the classification accuracy.

The remaining part of this paper is organised as follows. Section 2 introduces the datasets used in this study. Research methods including the development of pre-processing methods, feature extraction schemes, feature selection algorithms, and the classification model are described in Section 3. Parameter optimisation is introduced in Section 4. Experimental results and discussion are given separately in Sections 5 and 6. Section 7 summarises the paper.

2. Materials

To test the proposed methods, two mammogram datasets INbreast [27] and MIAS [28] containing different image types and using different density classification criteria, are used in our experiments.

INbreast [27] is a full field digital mammogram (FFDM) dataset and consists of 410 images from 115 cases, including bilateral mediolateral oblique (MLO) and craniocaudal (CC) views. Each image is saved in the DICOM format and in size of 3328×4084 or 2560×3328 pixels, depending on the compression plate used in the acquisition. This dataset offers carefully associated ground truth (GT) annotations made by a specialist in the field and validated by a second specialist. For each mammogram, its density is labelled as one out of four BI-RADS categories (BI-RADS I-IV). Images distribution with their density labels are as follows: 136 (BI-RADS I), 146 (BIRADS II), 99 (BI-RADS III), and 28 (BI-RADS IV).

MIAS [28] is a scan field mammograms (SFM) dataset, containing 322 images taken from 161 women, with only MLO views on both sides from the UK National Breast Screening Programme. Each mammogram is at 50 micron resolution in portable grey map (PGM) format. All images are associated with density ground-truth labels of three classes: fatty (F), fatty-glandular (G), and dense-glandular (D). There are 106 images belonging to fatty group, 104 and 112 images to the fatty-glandular and dense-glandular classes. Figure 1 shows sample images belonging to each density category from the two datasets.
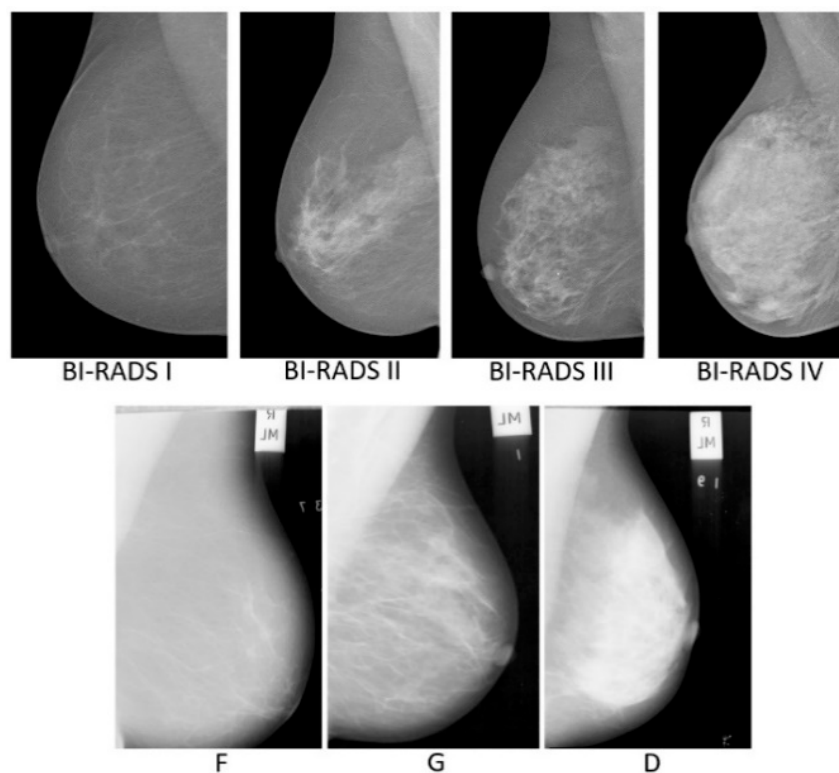
Figure 1. First row: INbreast mammograms in four BI-RADS density classes. Second row: MIAS mammograms in three tissue density classes.

3. Methodology

This paper proposes a novel texture descriptor called rotation invariant uniform local quinary patterns (RIU4-LQP) to describe texture patterns in mammograms. In addition to histogram information, this study further explores richer statistical information extracted from the RIU4-LQP feature set by using Baddeley's K-inhom method [29]. A new feature vector called K-spectrum is developed based on the extracted spatial information, which offers supplementary and important image features. Histograms and K-spectra are first extracted and tested separately and then are combined together to create a new texture feature space. A novel feature selection step is considered carefully in this study for removing redundant image features. An overview of the workflow using our proposed methods is shown in Figure 2, and more details are given in the following sections.
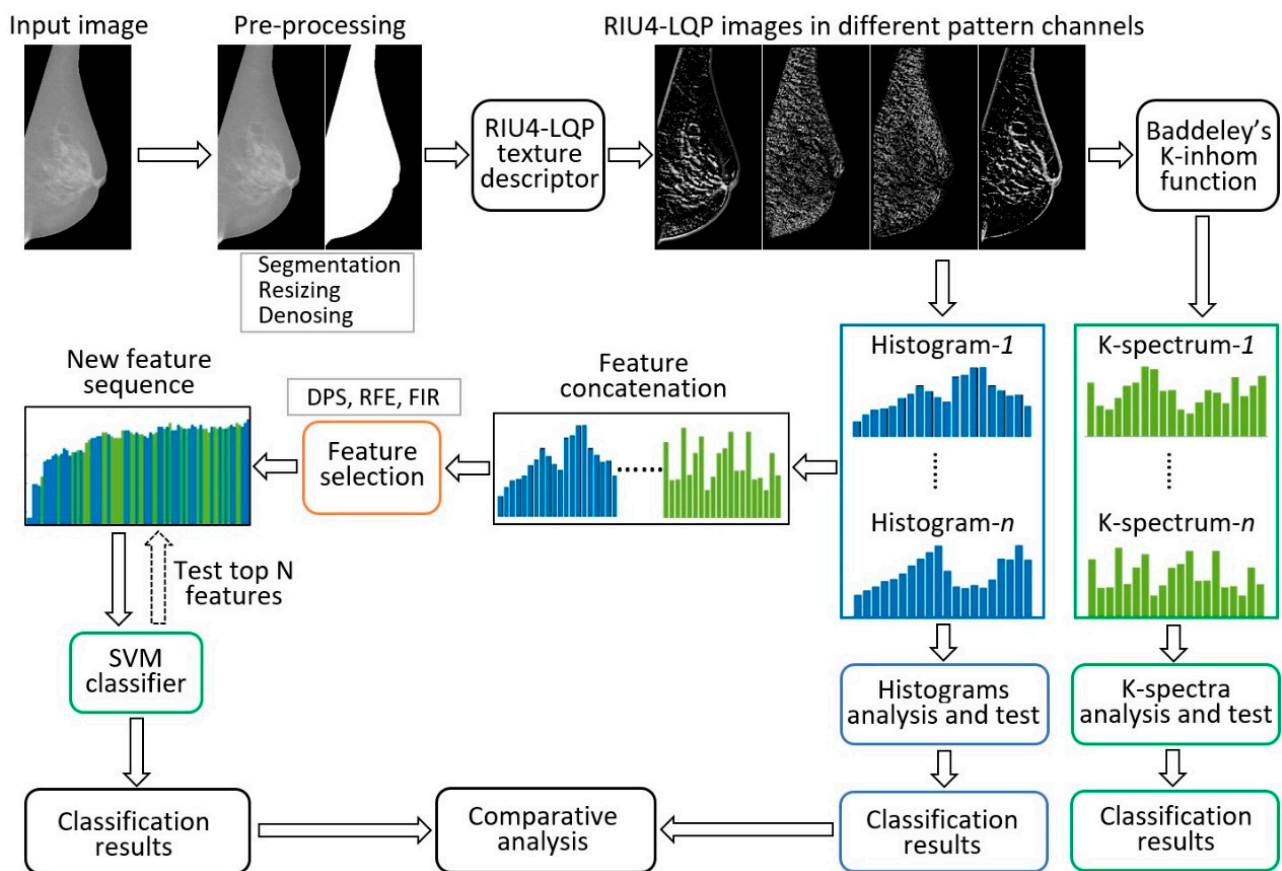

Figure 2. An overview of the breast density classification process using the proposed methods.

3.1. Pre-Processing

This pre-processing step contains breast area segmentation, denoising operation, and resizing the image. Breast segmentation is first applied to remove nonbreast areas such as background region, pectoral muscle, and label artefacts. A multifractal-method-based feature-enhanced image [30] is used to highlight the contrast between image background and the breast tissue region. The intensity thresholding method and morphological operations [31] are used to separate breast region and artefacts from the background. The label artefacts can be removed by keeping only the largest connected area (breast region). K-means algorithm and polynomial fitting approach [32] are employed to eliminate pectoral muscle from the breast region in MLO view mammograms. A median filter of 3×3 size is used to reduce noise. Finally, mask images are obtained, which are used to extract image features only from the region of interest (breast area) in the following steps. As relative work [16,17] reported promising classification results using resized mammogram images, this study applies the bicubic interpolation method to resize processed images with a

scale factor s [33], resulting in a resized image that is s times the size of original image. Figures 3 and 4 show some examples of segmenting breast region from image background using INbreast and MIAS mammograms, respectively. For the MIAS dataset that only contains MLO view images, relative work [21,34] demonstrated that using a cropped square region of interest (ROI) produces a better classification result. This study therefore uses a similar method as that in [34] to obtain the ROIs in MIAS. Figure 5 illustrates ROI extraction in MIAS mammograms. A detailed introduction of this pre-processing stage can be found in Supplementary Materials.
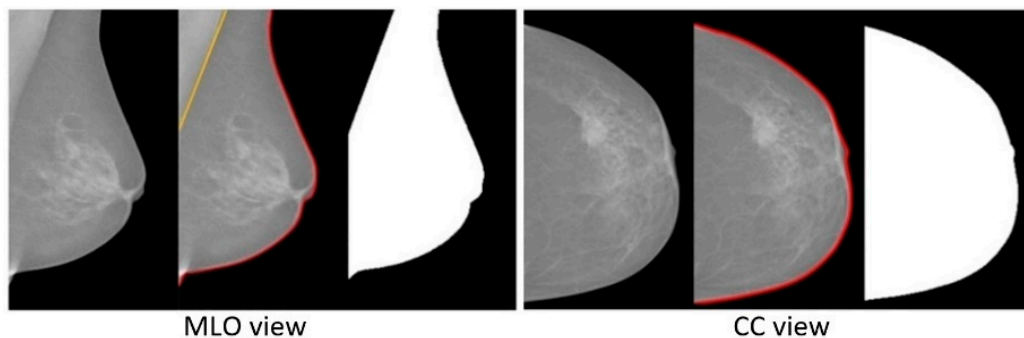
Figure 3. Examples of segmenting breast region and removing pectoral muscle area in INbreast mammograms.
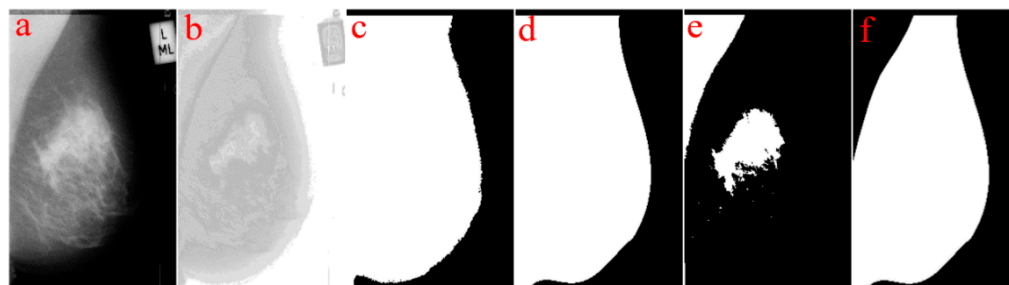
Figure 4. Segmentation of breast region in MIAS mammograms: (**a**) input mammogram, (**b**) enhancing breast region and artefact areas, (**c**) rough contour of breast region, (**d**) smoothing breast contour, (**e**) finding pectoral muscle area, (**f**) breast mask image without pectoral muscle region.
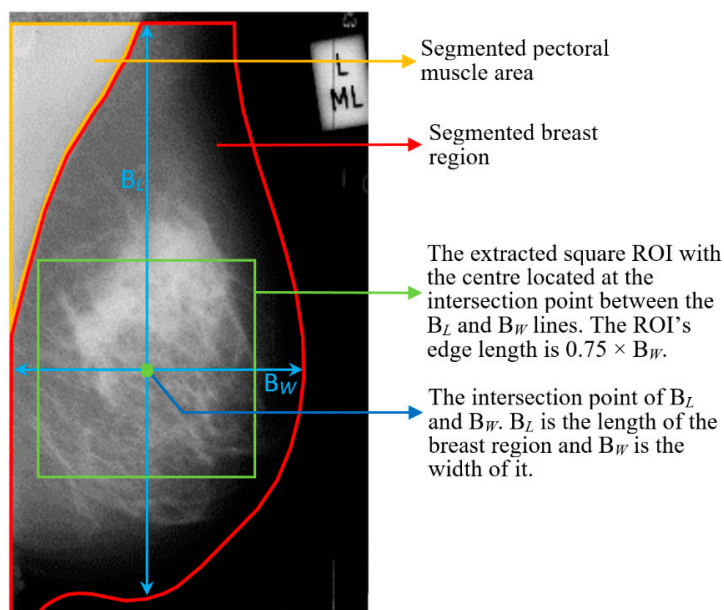
Figure 5. ROI extraction from the central area of MLO-view mammogram in MIAS dataset.

3.2. Local Quinary Patterns (LQP) and RIU4-LQP

Ojala et al. [20] proposed LBP to describe local image structure and to extract texture features. Nanni et al. [24] proposed LQP by extending LBP from a binary value encoding scheme to a 5-value encoding algorithm. LQP uses values $-2, -1, 0, 1$, and 2 to describe the relations between the intensity values of a central point and its neighbours. Each LQP code can be split to 4 LBP patterns, capturing more detailed texture information. Thus, for analysing medical images such as mammograms that contain regions of subtle texture differences, texture features extracted by LQP are more informative to be used to improve the classification accuracy. To implement the LQP operator, two threshold values $\{\tau_1, \tau_2\}$ are required for generating a 5-value encoding pattern. The calculation of LQP code can be described as follows:

$$LQP_i(c, R, P) = \sum_{p=0}^{P-1} s_i(I_p - I_c)2^p, \; i = 1, 2, 3, 4 \tag{1}$$

$$s_1(x) = \begin{cases} 1, & if \; d(x) = 2 \\ 0, & otherwise \end{cases} \tag{2}$$

$$s_2(x) = \begin{cases} 1, & if \; d(x) = 1 \\ 0, & otherwise \end{cases} \tag{3}$$

$$s_3(x) = \begin{cases} 1, & if \; d(x) = -1 \\ 0, & otherwise \end{cases} \tag{4}$$

$$s_4(x) = \begin{cases} 1, & if \; d(x) = -2 \\ 0, & otherwise \end{cases} \tag{5}$$

$$d(x) = \begin{cases} 2, & \tau_2 \le x \\ 1, & \tau_1 \le x < \tau_2 \\ 0, & -\tau_1 \le x < \tau_1 \\ -1, & -\tau_2 \le x < -\tau_1 \\ -2, & x < -\tau_2 \end{cases} \tag{6}$$

where R denotes the radius of a circular neighbourhood of the centre pixel c and P is the number of neighbourhood pixels used to calculate LQP code. I_p and I_c are intensity values of the pth neighbour pixel and the centre pixel c, respectively. From a specific position (usually the top-left corner) the binary values given by $s_i(x)$ are gathered in a specific sequence (usually in clockwise order) to obtain the LQP_i codes. Figure 6 illustrates how the LQP operator works in a local region with $R = 2$ and $P = 16$. After obtaining the 5-value pattern, it is split into 4 binary patterns by $s_i(x)$ in Equations (2)–(5).

LQP produces a high feature dimensionality given by 2^{P+2}, which increases exponentially with P. A large feature space cannot be utilised efficiently to train a classification model, and information redundancy in the feature set can have a negative impact in the final classification performance. To resolve the high dimensionality problem, one option is to use a rotation invariant strategy to extend LQP to rotation invariant uniform LQP (RIU2-LQP) [35]. However, RIU2-LQP suppresses too much texture information (from 2^{P+2} to $4 \times (P + 2)$) which may in turn reduce the final classification performance (Figure 6).
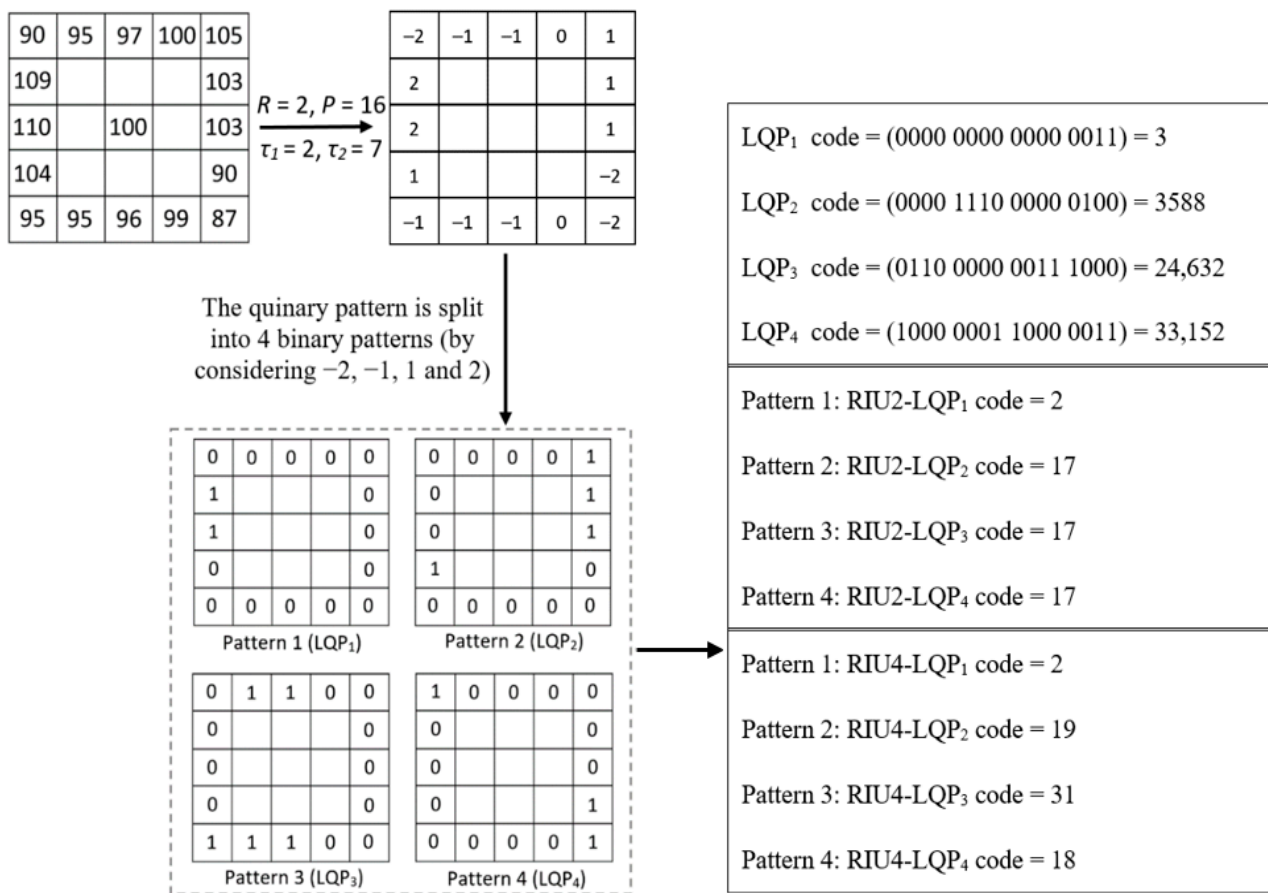
Figure 6. LQP encoding example, one 5-value pattern is split into 4 LBP patterns.

Under the consideration discussed above, this paper extends LQP to RIU4-LQP by analysing the transition number in a wider range. RIU4-LQP allocates different codes to binary patterns with two '1'-contiguous segments (e.g., patterns 2–4 in Figure 6), which captures richer image representations of texture patterns compared to RIU2-LQP. The proposed RIU4-LQP encoding scheme can be formulated by following equations:

$$LQP_i^{riu4}(c,R,P) = \begin{cases} \sum_{p=0}^{P-1} s_i(I_p - I_c), & if\ T_i(LQP(c,R,P)) \in \{0,\ 2\} \\ P + index, & if\ T_i(LQP(c,R,P)) = 4 \\ \left\lceil \frac{P^2+11}{4} \right\rceil - 1, & otherwise \end{cases} \tag{7}$$

$$T_i(LQP(c,R,P)) = |s_i(I_{P-1} - I_c) - s_i(I_0 - I_c)| + \sum_{p=1}^{P-1} |s_i(I_p - I_c) - s_i(I_{p-1} - I_c)| \tag{8}$$

$$index = \begin{cases} \sum_{n=1}^{X-1}(P-3-2(n-1)) + (Y-X+1), & if\ X \geq 2 \\ Y-X+1, & if\ X = 1 \end{cases} \tag{9}$$

where $i \in \{1, 2, 3, 4\}$; $T_i(\cdot)$ is defined as the number of spatial transitions (bitwise 0/1 changes) in patterns. X and Y denote the number of occurrences of '1' (defined by $s_i(x)$) in two distinct contiguous segments when $T = 4$ (e.g., Figure 6, pattern 2–4). We impose a restriction with respect to the relationship between X and Y: X always denotes the shorter '1'-contiguous segment, i.e., $X \leq Y$, for example, the pattern-3 in Figure 6: $X = 2$, $Y = 3$. Equation (7) contains three parts to correspond to different bit transitions (i.e., the value of T) for recognising and encoding texture patterns. The first part (i.e., the first row of Equation (7)) is same with RIU2-LQP for distinguishing texture patterns that have none or only one '1'-contiguous segment (e.g., the pattern-1 in Figure 6). The second part of Equation (7)

aims to encode texture patterns with two '1'-contiguous segments (e.g., patterns 2–4 in Figure 6). As the encoding values from 0 to P have been allocated with the condition of $T \in \{0, 2\}$, the output code starts from P and adds the value of *index*, which counts from 1 using Equation (9). The value of *index* is used to recognise and label those texture patterns when $T = 4$, for example, *index* = 1 with the pattern of $X = 1$, $Y = 1$; *index* = 2 with the pattern of $X = 1$, $Y = 2$; and so on. The third part of Equation (7) uses a unified code to denote all the other texture patterns. By using the proposed RIU4-LQP encoding method, RIU4-LQP codes are allocated to different texture patterns shown in Figure 6.

3.3. Rotation Invariant Property of RIU4-LQP and Its Utilisation

The basic LQP method is not rotation invariant. When the processed image is rotated, the neighbour pixels I_p will correspondingly move along the perimeter of the circle around I_c (Figure 7). Since the binary coding sequence of LQP always begins from a pre-designated and fixed position, e.g., to the top-left of I_c, rotating a particular binary pattern naturally results in a different LQP code.
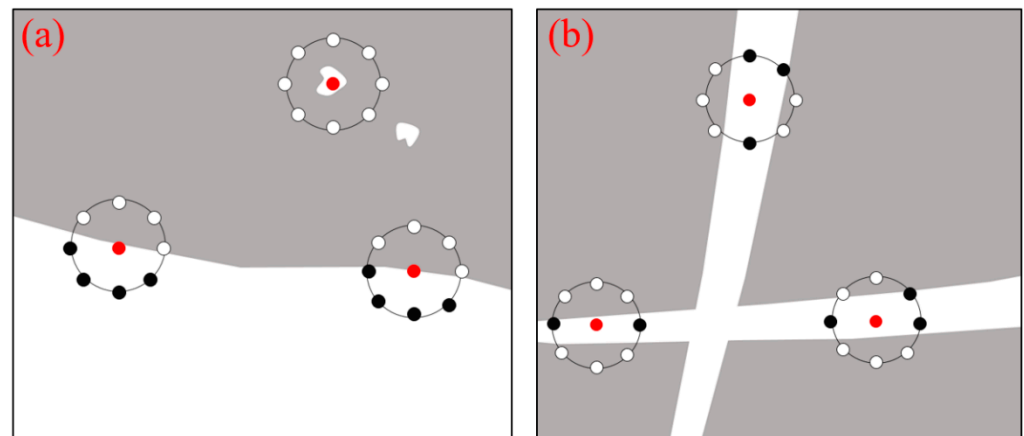
Figure 7. Different local texture structures are described by LQP variants using 8 neighbour pixels I_p with the central pixel I_c marked in red. Black points represent '0' in the circular bit patterns and white points represent '1'. (**a**) Microstructures of spots and edges can be captured by both RIU2-LQP and RIU4-LQP. (**b**) Microstructures of thin stripes are effectively represented by RIU4-LQP but not by RIU2-LQP.

To remove the effect of rotation, the function $T_i(\cdot)$ that measures bit transitions is used in Equation (8) for defining uniform patterns. In our proposed variant, the uniform patterns refer to circular structures that contain extended spatial transition conditions (i.e., $T = 0$, 2, 4, or others), such as patterns shown in Figure 6 with $T = 2$ or 4. By using $T_i(\cdot)$ and Equation (7), the encoding system works by recognising different '1'-sequence patterns without considering a fixed binary order, thus achieving rotation invariance. For example, when $T = 2$, the binary sequences '00111100' and '00001111' share the same RIU4-LQP code of 4, presenting a microstructure of an edge, while in the basic LQP, they have two different code values of 60 and 15. Similarly, when $T = 4$, the sequences of '11001110' and '11101100' share the same RIU4-LQP code of 15, as the related '1'- sequence patterns are same (with $X = 2$, $Y = 3$ using Equation (9)). In addition, under the condition of $T = 4$, some microstructures presenting thin stripe shapes with two-sided edges are captured by RIU4-LQP, but failed to be represented by LQP or RIU2-LQP. In mammograms, such local structures are commonly observed in regions with fibroglandular tissues that relate to breast density. Figure 7 illustrates the typical texture structures captured by RIU4-LQP. Based on this design, we assume that the proposed RIU4-LQP possibly captures more structural details for mammographic images, and this argument is tested and demonstrated in the following experiments.

The proposed RIU4-LQP reduces the feature dimensionality from 2^{P+2} to $P^2 + 11$. Meanwhile, compared to RIU2-LQP, more texture patterns are included in the extracted features by considering a higher number of bit transitions. Figure 6 gives the comparison between RIU2-LQP and RIU4-LQP codes on same patterns. Figure 8 shows RIU4-LQP images in different texture pattern channels by substituting corresponding RIU4-LQP codes for pixel intensities and their histogram-based feature vectors.
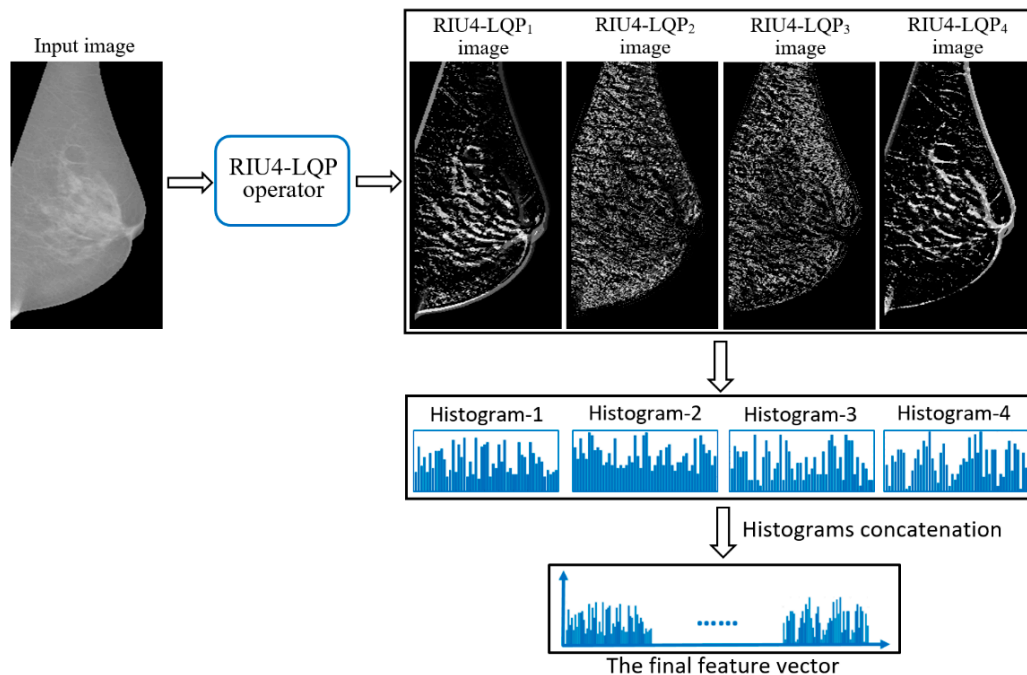
Figure 8. Example of RIU4-LQP images in different texture pattern channels and corresponding feature vectors.

3.4. Baddeley's K-Inhom Function

In conventional applications of texture descriptors [21,25,26,34], image texture features are represented by histogram information that counts the quantity of feature points but ignores their spatial distribution characteristics (Figure 9). This study employs Baddeley's K-inhom method [29] to extract spatial information based on RIU4-LQP feature points. The K-inhom method is a variant of Ripley's K function [36], and both are statistical analysis schemes used for studying qualitative or quantitative characteristics of spatialised data. Generally, K-inhom function for one-dimensional data can be described as follows:

$$K_{\text{inhom}}(r) = \frac{1}{D}\sum_{i}\sum_{j \neq i}\frac{1\{\|x_i - x_j\| \leq r\}}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)}c(x_i, x_j; r) \tag{10}$$

$$D = \frac{1}{|W|}\sum_{i}\frac{1}{\hat{\lambda}(x_i)} \tag{11}$$

$$c(x_i, x_j; r) = \frac{1\{b_i > r\}}{\sum_{j}(1\{b_j > r\}/\hat{\lambda}(x_j))} \tag{12}$$

where $1\{\|x_i - x_j\| \leq r\}$ denotes an indicator that takes a value 1 if the distance between point x_i and point x_j is less than or equal to r or 0; otherwise, r is a distance measure; $c(x_i; x_j; r)$ corresponds to the correction of edge effects and W to the region of interest; and b_i is the distance from x_i to the boundary of W. The function $c(x_i; x_j; r)$ is implemented as in [37] for border corrections. $\lambda(x_i)$ denotes an intensity function around point x_i, which is defined by the number of neighbour points (x_j) expected in a small area (using the

radius of *r*) with x_i in the centre, but $\lambda(x_i)$ is not known in practice. Instead, '$\hat{\lambda}(x_i)$' is used in Equations (10) and (11) as the estimation of $\lambda(x_i)$, which is implemented by a nonparametric method [29]. More details concerning Baddeley's K-inhom function and its implementation can be found in [29,38]. The 'spatstat' package [38] in R is used to implement Baddeley's K-inhom method. Values of $K_{inhom}(r)$ are calculated with different distance measurements (*r*), which result in a K_{inhom} curve by connecting all the observed values of $K_{inhom}(r)$ (Figure 10). In [37,38], an expected reference value $K_{pois}(r) = \pi r^2$ (the red dotted line in Figure 10) obtained based on an inhomogeneous Poisson process, is used to compare with the observed $K_{inhom}(r)$ value. If the K_{inhom} curve is located below the reference curve (i.e., $K_{inhom}(r) < K_{pois}(r)$), it indicates that corresponding points scatter regularly in the region of interest. By contrast, if the K_{inhom} curve is located above the reference curve (i.e., $K_{inhom}(r) > K_{pois}(r)$), the distribution of points tends to be more aggregated. Therefore, the K_{inhom} curve can be used to describe the spatial distribution characteristics of a point set. As this paper focuses on mammogram image analysis, we use the segmented breast region as the region of interest *W* (Figure 10). The pixels decomposed by RIU4-LQP code values (Figure 9) constitute feature point sets in different code channels, which produce their K_{inhom} curves separately and show corresponding spatial characteristics.
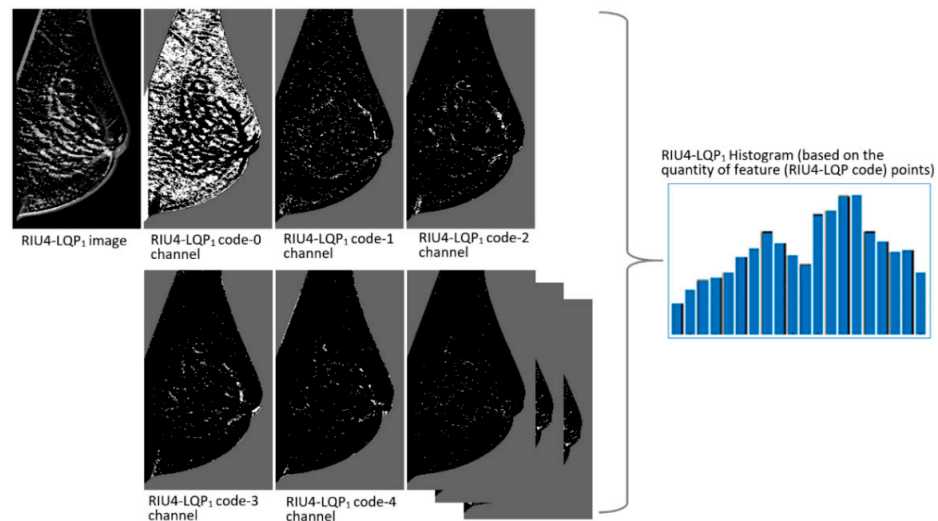
Figure 9. Decomposition of the RIU4-LQP$_1$ image in Figure 8 based on each encoding channel. The positions of pixels in each RIU4-LQP$_1$ code channel considered are marked by white points in the black breast region background.
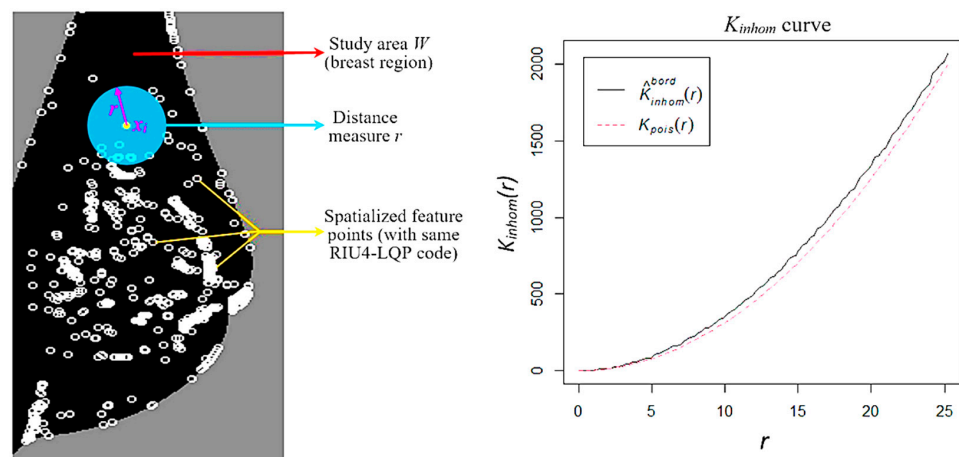
Figure 10. One feature pattern image (**left**) in a RIU4-LQP code channel with all feature points emphasised by white circles; the corresponding Kinhom curve (**right**) is generated by Baddeley's K-inhom function.

3.5. K-Spectrum of RIU4-LQP

This paper proposes a new texture feature vector called K-spectrum that is based on the K_{inhom} curve representing the spatial distributions of texture feature point sets in mammograms. The RIU4-LQP operator is used to produce a RIU4-LQP$_i$ code set {*code-1, code-2 , . . . , code-k*}. Subsequently, pixels in breast region W are decomposed into k different point sets {X_1, X_2, . . . , X_k} by corresponding RIU4-LQP$_i$ code values (Figure 9). Baddeley's K-inhom function is adopted to output a K_{inhom} curve for each point set X_i, which reflects how these points are scattered in the breast region with respect to a specific distance measure (r). As introduced in the last section, a reference curve $K_{pois}(r)$ is used for comparing with the observed K_{inhom} curve. Therefore, this study uses a deviation (d) between the observed $K_{inhom}(r)$ value and the reference value $K_{pois}(r)$ under the radius r and its mean ($\overline{d}$) on a valid r range to evaluate the spatial distribution information of point sets. The deviation d and the mean $\overline{d}$ are computed as follows:

$$d(r) = K_{inhom}(r) - K_{pois}(r) \tag{13}$$

$$K_{pois}(r) = \pi r^2 \tag{14}$$

$$\overline{d_i} = \frac{\sum_{r=1}^{s} d_i(r)}{s}, \; i = 1, 2, \ldots, k. \tag{15}$$

All the means ($\overline{d}_1$, $\overline{d}_2$, . . . , $\overline{d}_k$) are concatenated to form a new feature vector called 'K-spectrum'. Figure 11 shows how the K-inhom function works in mammogram images with the proposed procedures and generates the K-spectrum.
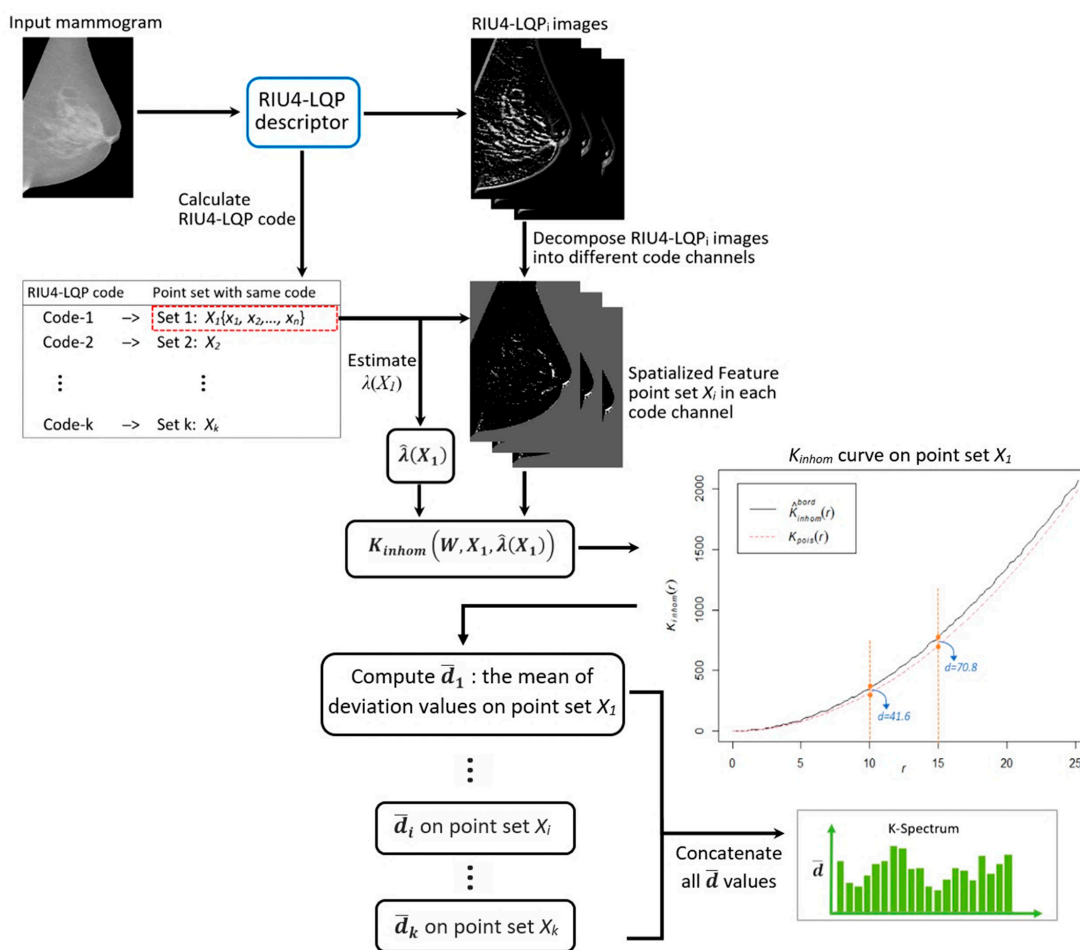

Figure 11. Illustration of using the K-inhom method and RIU4-LQP operator to generate the new feature vector (K-spectrum).

3.6. Feature Selection

As introduced in the first section, the initial feature space usually is too large to be used efficiently, and it is quite likely that some redundant features are contained within it. Therefore, a feature selection step is necessary to optimise the initial feature set. This study investigates three feature selection methods: dominant patterns set (DPS) [39], recursive feature elimination (RFE) [40], and feature importance ranking (FIR) [41].

Guo et al. [39] proposed a dominant patterns set (DPS) method to construct a subset of the initial feature set for filtering the most frequently occurring feature patterns. This feature selection method has been used in [25,26] for reducing feature dimensionality. In DPS, a set of dominant patterns of an image is defined as a minimum set that can cover $n\%$ ($0 < n < 100$) of all patterns. To produce a subset, a bin-wise summation for all the histograms of image features in a training set is performed first and this results in a histogram (H) with U bins. Then, all the bins of histogram (H) are sorted in descending order, and top M bins are selected using Equation (16):

$$\operatorname*{argmin}_{M} \frac{\sum\limits_{i=1}^{M} H(i)}{\sum\limits_{j=1}^{U} H(j)} \geq n\% \tag{16}$$

where U is the bins number in H and n is a threshold value by which M dominant patterns are selected from H.

The RFE [40] method recursively removes features and builds a model on remaining features. The model accuracy is used to identify which features contribute more than others for predicting the target classes. The estimated best feature is assigned a rank score '1', and the least related features have the highest rank sores.

The FIR method [41] uses ensembles of decision trees (e.g., random forest) to compute the relative importance of each attribute. An importance score is given for each feature to indicate that certain features play more important roles than others in the class prediction.

All feature selection methods produce a new sequence of features ranked according to their relevance/importance. The top (best) N features in this sequence can be selected and used to test the corresponding classification performance with the training set, and an N-feature set with the highest accuracy is the final feature vector used for the test set.

3.7. Classification

As related work [9,11,23,25,26] reported their best classification results by using SVM to predict target labels of mammographic density, SVM is used in this study for training the classification model and producing classification results on test images. Since this classification work aims to classify mammograms into multiple density categories (three or four), a multiclass SVM that is implemented by the one against all (OAA) method is used in this model. To obtain the optimum classification results, the three kernels used most often, RBF, Poly, and Sigmoid, are tested in this work. The grid-searching method is used to find the best combinations of parameters (*gamma*, *C*, and *degree*).

4. Parameter Optimisation

The methods used in the proposed classification model involve several parameters that affect the classification results differently. This section summarises the relevant parameters in different processing steps and addresses the test and optimisation methods.

4.1. Relative Parameters

The parameters and the range of their initial values are listed in Table 1. In the pre-processing stage, a scaling parameter s is selected from a set of {1, 1/2, 1/4, 1/8, and 1/16} by testing different mammogram datasets. Comparative analysis is conducted in our previous work [33] and the same settings of s are used. When using RIU4-LQP to extract

the feature set, a multiscale strategy is used for capturing richer image representation. Referring to related work in [25], three pairs of (R, P) with corresponding settings shown in Table 1 are used. The LQP-based method needs two extra threshold values {τ_1, τ_2} in its encoding system, which requires manual determination. An automatic approach is proposed in [42] by considering central pixel's intensity (I_c). Similarly, we introduce the following empirical rules for adaptively deciding τ_1 and τ_2: $\tau_1 = I_c \times 2\%$ and $\tau_2 = I_c \times 7\%$. For calculating the K-spectrum feature vector, different values of the distance measure (r) are used to produce the K_{inhom} curve (i.e., the K_{inhom} curve in Figure 10). Valid r values rely on the region of interest (W) (i.e., the segmented breast region), which does not have a uniform size among different mammograms. After comparing the observed r-ranges on all images in the dataset, a maximum valid r-range 1~25 is found and used to generate K-spectra. In the feature selection step, as the concrete values of (R_i, P_i) of RIU4-LQP are designated in Table 1, the feature dimensionality (n_1 and n_2) of feature sets can be obtained by Equations (7)–(9).

Table 1. Parameters and their value ranges used in different steps.

Method	Related Parameters
Image pre-processing	1. Resizing scale $s = 0.125$ (on INbreast) and 0.5 (on MIAS) 2. Median filter size: 3×3
RIU4-LQP	In multiscale method, three pairs of (R, P) are used: 1. ($R_1 = 2$, $P_1 = 10$), ($R_2 = 4$, $P_2 = 14$), ($R_3 = 8$, $P_3 = 18$) 2. Threshold values {τ_1, τ_2} are decided by an adaptive method
K-inhom function	1. Valid r-range: 1~25
Features selection	Initial features number: 1. $n_1 = 656$ (histogram or K-spectrum features) 2. $n_2 = 1312$ (concatenated features)
SVM classifier	For grid-searching: 1. Kernel candidates: RBF/Poly/Sigmoid 2. γ searching range: $[10^{-4}, 10^3]$ 3. C searching range: $[10^{-3}, 10^4]$ 4. *Degree* (only Poly kernel) searching range: {1, 2, 3, 4, 5, 6}

4.2. Selection of r-Range in K-Spectrum

As shown in Table 1, the maximum valid r value range 1~25 is extracted when applying Baddeley's K-inhom method in mammograms. Since there is no guarantee of the maximum range that is most effective for K-spectrum features, we narrow this maximum r-range by five-unit intervals, and five subranges are generated and tested for obtaining the optimal K-spectrum. Figure 12 shows that the highest classification accuracy (CA) is 0.83 and the highest AUCROC is 0.95 by using the r-range of 1~10 on INbreast dataset. Therefore, we use this r value range as the optimum distance measurement for producing K-spectrum in the following experiments.
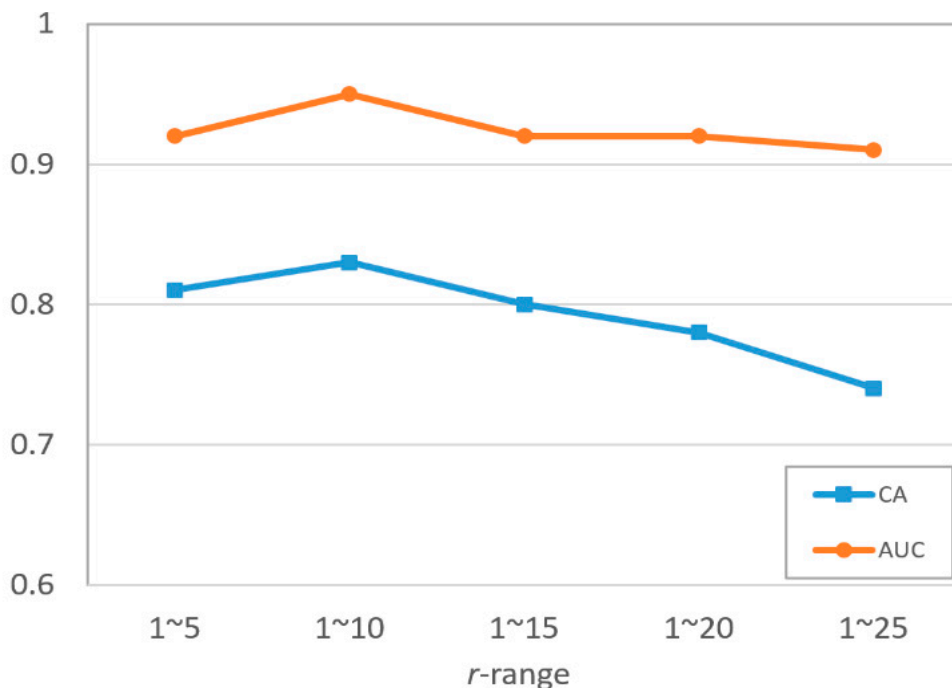
Figure 12. Comparison of classification performance on INbreast training set by using different value ranges of distance measurement *r*.

4.3. Grid-Searching Results for SVM Classifier

We consider three kernels (RBF, Sigmoid, and Poly) and different value ranges of other parameters (*gamma*, *C*, and *degree*) for the SVM classifier, as given in Table 1. The best combination of the kernel and parameters are found by grid-searching on two datasets (Figure 13).

Figure 13. Heatmaps of grid-searching results using training sets on two datasets.

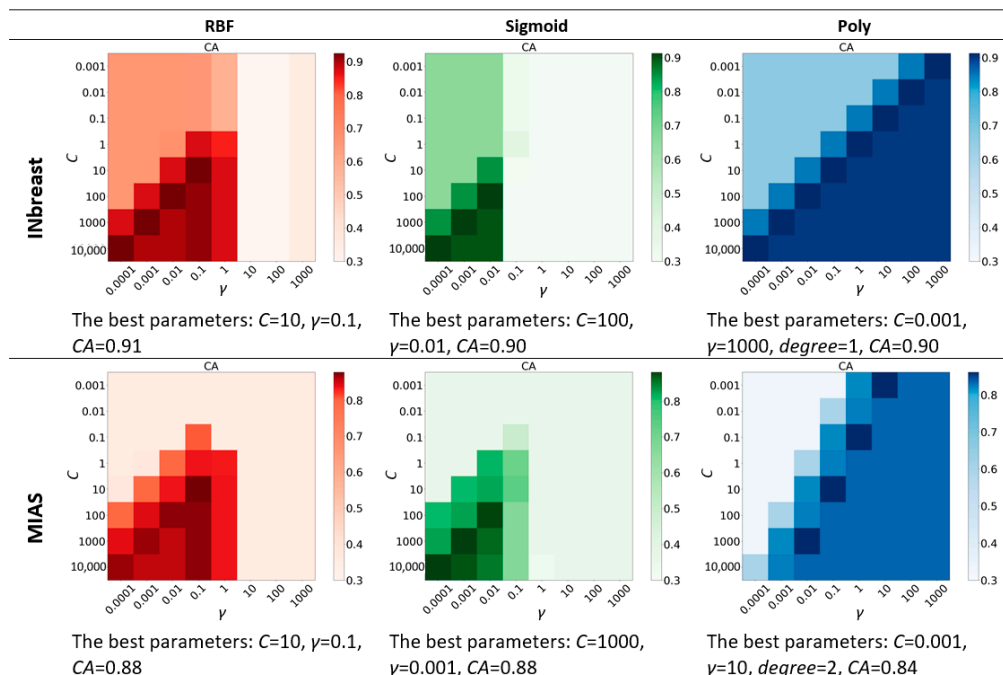## 5. Experiment and Results Analysis

As mentioned earlier, two mammogram datasets, INbreast and MIAS, are used to test the proposed classification model. To give a comprehensive and objective evaluation towards the classification performance, different measure criteria are used. For each test method, classification accuracy (CA) and area under the ROC curve (AUCROC) are calculated as the main performance indices. Since this study investigates the effectiveness of different feature selection methods, the final selected features number (N) is considered. We also conduct two different test methods: leave-one-woman-out test [11,13] and 10-fold cross validation [21,23,25].

5.1. Classification Results Using Histogram Information

A histogram-based feature vector that contains 656 RIU4-LQP features is tested first on two datasets. Three feature selection methods (Section 3.6) re-sort the feature vector based on their importance. For each test iteration, a feature subset containing the first N features $\{f_1, f_2, \ldots, f_N\}$ ($N \leq 656$) is sent to the classifier for producing the classification result. Then, N is increased and the test procedure enters the next iteration, finally obtaining the curves of CA vs. N. Figure 14 shows the classification results on two datasets in which the highest CA is 88.16% and 81.06% on INbreast and MIAS, respectively. Meanwhile, we can also notice that different feature selection methods affect the classification results differently. For INbreast, RFE reduces the feature number N to 161 and obtains the highest AUCROC value (0.95 ± 0.02); for MIAS, the highest AUCROC value is 0.91 ± 0.03 when using FIR with $N = 214$.
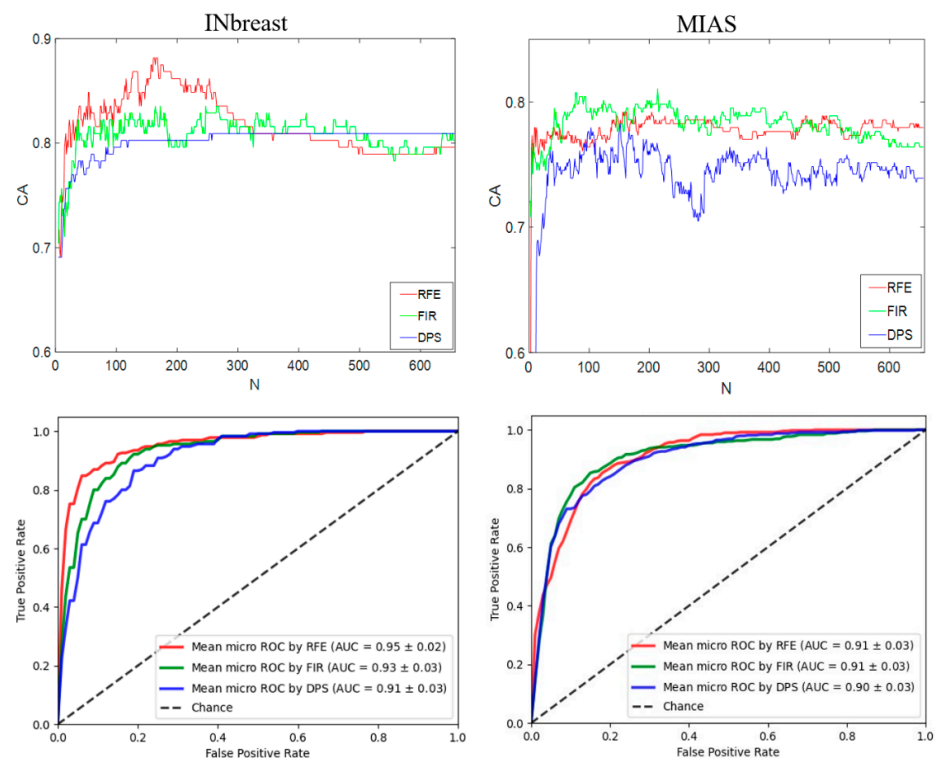
Figure 14. Classification accuracy (**first row**) and AUCROC values (**second row**) on two datasets using the feature vector based on histograms.

5.2. Classification Results Using K-Spectrum

K-spectra are extracted in mammograms in two datasets and used as texture features in the classification model. As the extraction of K-spectra are based on the same RIU4-LQP operator and multiscale method used for collecting histogram information, the K-spectrum-based feature vector also contains 656 features. Classification performance is shown in Figure 15, with the highest CA of 82.89% and 73.60% obtained on two datasets.
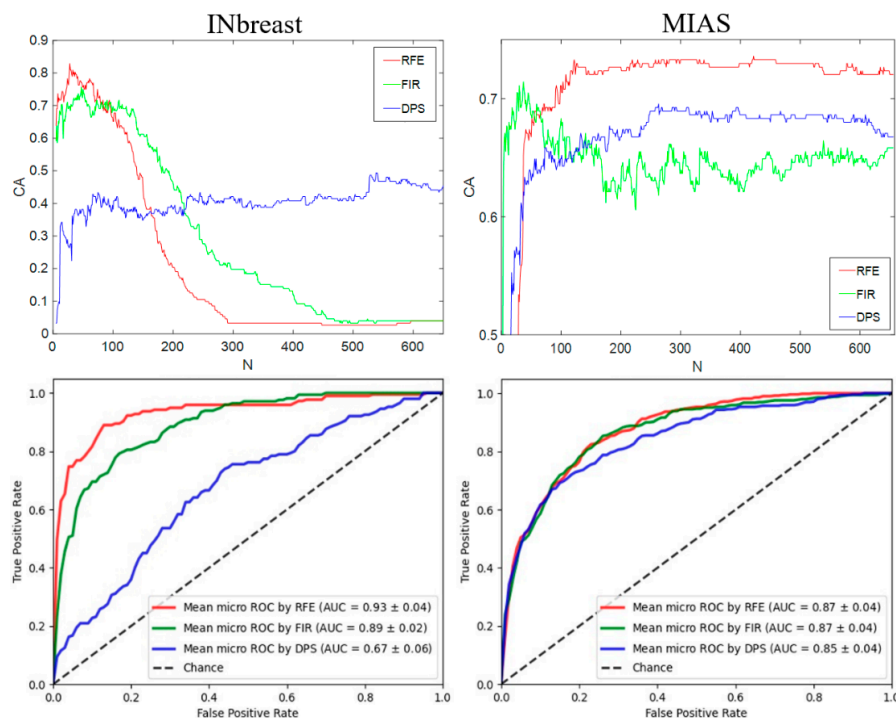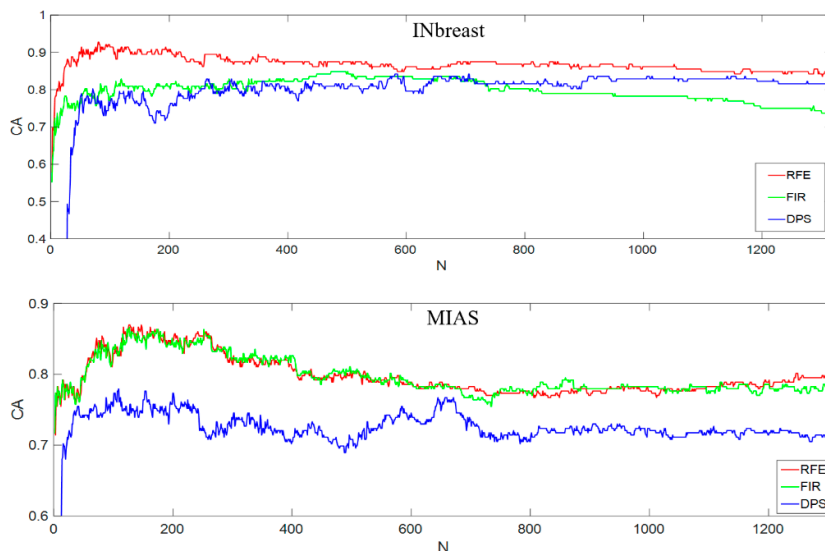
Figure 15. Classification accuracy (**first row**) and AUCROC values (**second row**) on two datasets using the feature vector based on K-spectra.

5.3. Classification Results Using Concatenated Features

The histogram and K-spectrum features are further concatenated to create a new hybrid feature space. In this step, feature selection procedure becomes more important, as the concatenation operation doubles the feature dimensionality from 656 to 1312. Figure 16 shows the classification results using the hybrid feature vector. A higher CA value 92.76% on INbreast and 86.96% on MIAS are obtained using RFE when $N = 80$ and $N = 127$, respectively, exceeding the best CA using only histogram or K-spectrum features. The AUCROC values are 0.95 ± 0.03 and 0.95 ± 0.02 on two datasets (Figure 17). Since the classification accuracy is improved significantly on both datasets after combining two feature sets, we can conclude that features extracted from K-spectra are complementary to histogram features and further improve the classification accuracy.

Figure 16. Classification accuracy on two datasets using combined texture features.
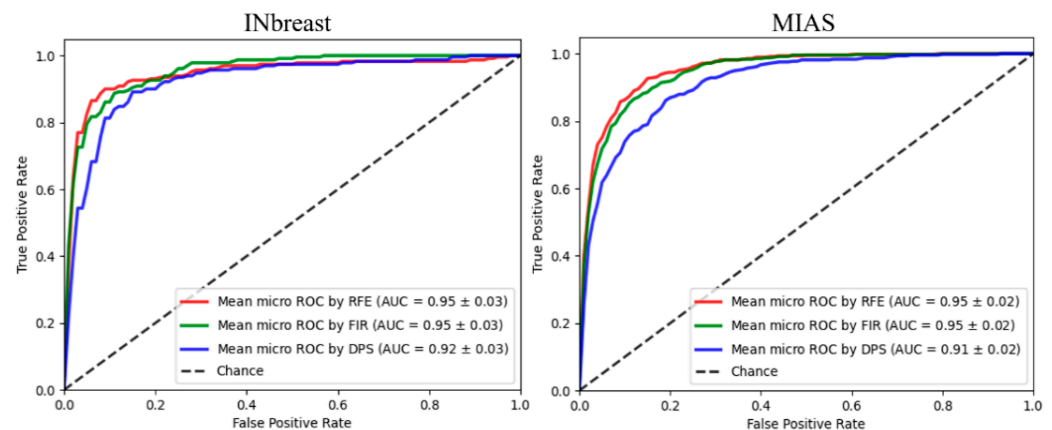
Figure 17. AUCROC values on two datasets using different feature selection methods.

5.4. Methods Comparison on the INbreast Dataset

In this section, different feature extraction methods are compared for the INbreast dataset. As INbreast has not been used widely with the breast density classification task in the literature, the only available experimental results based on this dataset are 86% and 80.5%, as reported in [25,26], with half images (MLO views) of the dataset tested. We compare five progressive transformations of LBP/LQP methods and implement corresponding algorithms to classify INbreast mammograms, performing comparative analysis based on their classification results. We use RIU4-LQP-K and RIU4-LQP-HK to denote texture feature sets based on the K-spectra and the combination of histograms and K-spectra. Classification performance is evaluated by classification accuracy (CA), AUCROC, Kappa coefficient, and F1 score. In addition, a statistical hypothesis test is conducted. The '10-fold cv *t*-test' method in [43] with a significance level of 0.05 (i.e., *alpha* = 0.05) is employed between RIU4-LQP-HK and every other method to calculate *p*-value, which shows the statistical difference between them. Table 2 shows that our proposed method outperforms other approaches, with the highest CA (91.87 $\pm$ 6.28), Kappa (0.89), and F1 score (0.92). In the *t*-test, all other methods present low *p*-values (<0.05), which means the difference in classification performance is statistically significant.

Table 2. Classification performance comparison on INbreast using different methods.

	CA (%)	AUC (%)	Kappa	F1	MCC	*p*-Value
RIU2-LBP	78.75 $\pm$ 9.35	94.93 $\pm$ 1.97	0.71	0.79	0.79	0.0002
RIU2-LQP	83.75 $\pm$ 7.50	94.09 $\pm$ 2.19	0.78	0.84	0.78	0.0002
RIU4-LQP	85.62 $\pm$ 12.20	95.79 $\pm$ 2.07	0.80	0.86	0.80	0.0251
RIU4-LQP-K	84.38 $\pm$ 8.03	95.37 $\pm$ 2.02	0.79	0.85	0.79	0.0028
RIU4-LQP-HK	91.87 $\pm$ 6.28	95.36 $\pm$ 2.52	0.89	0.92	0.90	-

5.5. Methods Comparison on MIAS Dataset

Since MIAS has been used in several research works with the breast density classification task, we first compare the five progressive transformations of LBP/LQP methods on it, and then our proposed method is compared with other approaches in the literature. In Table 3, we can see that the proposed RIU4-LQP-HK method provided the best classification performance on the MIAS dataset, with the highest CA (90.61 $\pm$ 4.87), AUCROC (94.39 $\pm$ 2.88), Kappa (0.86), and F1 score (0.90). In addition, we notice that the classification accuracy obtained can be affected by a few factors, such as the image number used in the test, the results evaluation method, and target categories. Therefore, when comparing our method with other approaches, these factors are considered for presenting an objective and fair comparison. In Table 4, the CA values are compared between different methods, and we can see that the proposed RIU4-LQP-HK method is very competitive with state-of-the-art approaches, providing the highest CA when testing all images in the MIAS dataset.

Table 3. Classification performance comparison on MIAS using different methods.

	CA (%)	AUC (%)	Kappa	F1	MCC	*p*-Value
RIU2-LBP	69.38 ± 7.68	84.14 ± 4.54	0.57	0.70	0.57	$<10^{-4}$
RIU2-LQP	72.73 ± 7.55	85.60 ± 4.56	0.59	0.72	0.60	$<10^{-4}$
RIU4-LQP	80.91 ± 8.89	90.72 ± 3.53	0.71	0.81	0.71	0.0099
RIU4-LQP-K	73.33 ± 6.03	87.39 ± 4.15	0.60	0.73	0.60	$<10^{-4}$
RIU4-LQP-HK	90.61 ± 4.78	94.39 ± 2.88	0.86	0.90	0.86	-

Table 4. Classification results comparison on the MIAS dataset.

Methods	Image Used	Results Evaluation Method	Results CA (%)	Density Categories
Intensity-based features + SVM [11]	322	Leave-one-image-out	85.7	3 (F, G, D)
Intensity-based features + KNN [11]	322	Leave-one-image-out	78.6	3 (F, G, D)
Intensity-based features + SVM [11]	322	Leave-one-woman-out	77.0	3 (F, G, D)
Intensity-based features + KNN [11]	322	Leave-one-woman-out	76.4	3 (F, G, D)
GLCM + KNN [6]	322	Leave-one-image-out	82.0	3 (F, G, D)
LQP + SVM [25]	322	10-fold cross validation	86.13	4 (BI-RADS)
LBP + Bayesian Network [21]	321	10-fold cross validation	69.4 ± 0.92	3 (F, G, D)
ELBP + Bayesian Network [21]	321	10-fold cross validation	75.4 ± 1.05	3 (F, G, D)
u-ELBP + Bayesian Network [21]	321	10-fold cross validation	73.3 ± 0.64	3 (F, G, D)
M-ELBP + Bayesian Network [21]	321	10-fold cross validation	77.4 ± 1.06	3 (F, G, D)
LDP + Bayesian Network [21]	321	10-fold cross validation	76.0 ± 0.96	3 (F, G, D)
RIU2-LBP + SVM [26]	322	5-fold cross validation	73.8 ± 10.6	4 (BI-RADS)
RIU2-LTP + SVM [26]	322	5-fold cross validation	81.0 ± 9.5	4 (BI-RADS)
RIU2-LQP + SVM [26]	322	5-fold cross validation	82.1 ± 7.1	4 (BI-RADS)
RIU2-LSP + SVM [26]	322	5-fold cross validation	83.3 ± 8.8	4 (BI-RADS)
RIU4-LQP-HK + SVM	322	Leave-one-woman-out	86.96	3 (F, G, D)
RIU4-LQP-HK + SVM	322	Leave-one-image-out	93.21	3 (F, G, D)
RIU4-LQP-HK + SVM	322	10-fold cross validation	90.61 ± 4.78	3 (F, G, D)
RIU4-LQP-HK + SVM	322	5-fold cross validation	86.25 ± 5.24	3 (F, G, D)

5.6. Effect of Feature Selection

Three feature selection methods are used and compared in all experimental analyses in this paper. In related work [25,44], the DPS method was used to optimise texture features for analysing mammograms and other texture images. However, there was no comparative analysis including other feature selection methods. To bridge this gap, we use the selection results by DPS and corresponding classification accuracy as the base line in this work, and use two other feature selection methods, RFE and FIR, to repeat the feature selection and breast density classification procedures. Comparisons are given in Table 5, from which we can see that RFE works better than the other two methods, with a low number of features used and higher classification accuracy obtained.

Table 5. The number (*N*) of features selected by different methods and corresponding CA values.

Dataset	Feature Selection Methods	Histograms-Based Feature Vector		K-Spectra-Based Feature Vector		Combined Texture Features	
		N	Best CA	N	Best CA	N	Best CA
INbreast	RFE	161	88.16	28	82.89	80	92.76
	FIR	164	83.55	47	75.66	473	84.87
	DPS	260	80.92	538	49.34	581	84.21
MIAS	RFE	156	79.19	422	73.60	127	86.96
	FIR	214	81.06	37	71.43	127	86.65
	DPS	152	78.26	263	69.57	109	77.95

5.7. Running-Time Comparison

This section compares running-time cost using the proposed classification model. Note that the time cost is recorded for the 10 run 10-fold cross validation in which one fold images are used as the test set and remaining images are used to train the classifier for each iteration. The total time is divided by the number of training and testing images separately in a dataset, and we obtain the average time for processing one image. The methods outlined in this paper are implemented in Matlab R2017b on a desktop computer with Intel core i7 3.6 GHz CPU, 16 GB memory. In Table 6, we can see that the RIU2-LQP- and RIU4-LQP-based methods have similar time cost, requiring around 13~15 ms per image for training. By contrast, the use of LQP costs around 50 times longer time than the other methods on the two datasets due to its high feature dimensionality.

Table 6. Time cost for training and classifying procedures by using different methods.

Dataset		Average Time for Classifying Images (Milliseconds per Image)					
		LQP	RIU2-LQP	RIU4-LQP	RIU4-LQP-HK + RFE	RIU4-LQP-HK + FIR	RIU4-LQP-HK + DPS
INbreast	Training	715.26	13.96	14.03	13.62	13.93	14.25
	Test	176.71	3.67	3.97	2.90	3.77	4.19
MIAS	Training	717.05	14.85	14.78	14.63	14.23	14.35
	Test	181.34	3.71	3.99	3.56	3.47	3.49

6. Discussion

This paper introduces two improvements of the basic LQP method. The first is the RIU4-LQP, which is an extension of the LQP. The second is a novel spatial feature extraction method based on Baddeley's K-inhom function.

The proposed RIU4-LQP decreases the high feature dimensionality of LQP from 2^{P+2} to $P^2 + 11$, as discussed in Section 3.2. The significantly reduced dimension of the feature space makes it possible to consider a larger neighbourhood of pixels (i.e., higher value of P) when computing local texture patterns, while avoiding the exponential rise of the number of features. From the running-time cost shown in Table 6, we can see that the RIU4-LQP-based method only uses around 2% of the time consumed by the basic LQP method. In addition, the proposed RIU4-LQP is rotation invariant, and more texture information related to microstructures can be captured by it as wider spatial transition numbers (T) are considered. As discussed in Section 3.3, the texture feature discrimination capability is improved in RIU4-LQP, based on which we assume that a better classification performance for grouping mammogram images can be obtained. The experimental results in Tables 2 and 3 support this argument: comparing to RIU2-LQP, the classification accuracy is improved by 2% on the INbreast dataset and 8% on the MIAS dataset. Such improvement shows empirical evidence that the proposed RIU4-LQP indeed captures more texture features effectively and helps classify mammograms with higher accuracy.

For the design of a spatial feature extraction method, our main consideration is that every pixel in an image has its own RIU4-LQP code related to the local texture structure. Previous work in this area used the histogram to represent texture features that only gave the frequency of each code and ignored their spatial distribution characteristics. In Sections 3.4 and 3.5, we address a novel spatial feature extraction framework using Baddeley's K-inhom function to collect spatial distribution information based on RIU4-LQP. Here, we assume that the extra spatial features provide complementary information about image texture structures that are not obtained from histograms. From experimental results in Sections 5.4 and 5.5, we can see that the classification accuracy using only RIU4-LQP-based K-spectrum (i.e., RIU4-LQP-K) does not surpass the results by RIU4-LQP-based histograms, with the CA of 84.38% vs. 85.62% on INbreast and 73.33% vs. 80.91% on MIAS. Meanwhile, we observe that the ability of RIU4-LQP-K for the mammogram classification task is similar to RIU2-LQP: 84.38% vs. 83.75% on INbreast and 73.33% vs. 72.73% on MIAS. This indicates

that the spatial features extraction using our proposed method performs well, yielding slightly higher performance than RIU2-LQP. Therefore, we further concatenate RIU4-LQP and RIU4-LQP-K to construct a new feature set, RIU4-LQP-HK, which contains both histogram statistics and the spatial distribution features. As shown in Tables 2 and 3, the improvement on classification performance is significant: taking the RIU4-LQP results as the baseline, CA is improved by 6.25% on INbreast and 9.70% on MIAS. Such improvement is the result of adding the extra spatial features (i.e., RIU4-LQP-K), thus supporting the argument that spatial distribution characteristics can offer some supplementary features that are ignored by histogram statistics. Note that a recent work [18] using a CNN model to classify mammograms reached 88% of accuracy on a large clinical dataset containing 1985 images, but obtained 70% of accuracy on INbreast. The authors in that work argue that the reason comes from the use of a small dataset for training deep CNNs (Figure S2). This also indicates that when training data are limited, our proposed texture feature descriptor is competitive to produce better results, even comparing to neural network models.

Regarding the choice of classification model, various classifiers have been used in the literature, including SVM [8,9,23,25], KNN [6,13], RF [10], Bayesian [21], and MLP [26]. SVM was commonly used and produced higher classification accuracy than others [11,26]. This study therefore uses SVM to process the final feature set, and the grid-search method to select the optimum parameters for it. By contrast, feature selection was not considered and analysed carefully in related work. Some work [8–13,23] did not design a feature selection step, instead directly input initial extracted features into a classifier, while others [21,25,26] selected features prior to using a classifier but did not present a comparative analysis. Note that feature selection discussed in this paper refers to the removal of redundant features and compressing feature space for presenting the best classification accuracy. Redundant features that do not closely relate to the target density category can increase the complexity of the classification model and decrease its classifying accuracy. The work reported in this paper particularly gives attention to this point by imposing three feature selection methods and comparing their influences on classification performance.

As shown in Figures 14–16, our experimental results provide detailed comparisons when applying different feature selection schemes to two mammogram datasets. From these results, the influence on classification accuracy caused by using different methods is apparent. For example, the CA results on INbreast in Figure 15, the accuracy exceeds 80% when using the top 30 features selected by the RFE method, while it declines to less than 10% after using more than 250 features. This finding also demonstrates that the initial extracted feature space contains some redundant features that do not closely relate to the target class properties (e.g., mammographic density categories in this work). Therefore, removing those redundant features and only keeping the target-related features in the final feature vector is important for producing desirable classification performance. Our analysis on different feature selection methods and feature descriptors are displayed in Table 5, based on which we conclude that for the breast density classification work the RFE performs better than the other two approaches after testing two mammogram datasets. RFE is not used in past work, and we suggest that this can be an important factor to improve the classification results.

As the RFE and other two feature selection methods, work by re-sorting the extracted features to position the most powerful features ahead and redundant features back, it is unclear of how much proportion the spatial features account for in the final selected feature set. We therefore investigate how different features contribute to the final classification results by tracking their orders in the selection procedure. Table 5 shows that there are 80 features in the final feature vector used to obtain the best CA on INbreast, and we find out that 42% of features (i.e., 34 out of 80) contributed by spatial characteristics and the remaining are from histograms: for MIAS, K-spectrum spatial features account for 48% (i.e., 61 out of 127) of the optimised feature set. This gives further evidence that the spatial features extracted by our proposed method create an important and complementary

feature set compared to the commonly used histograms, thus improving the capacity of representing image texture features.

Some inconsistent result patterns between the two mammogram datasets are noticed. For example, in Figure 14, while presenting the classification accuracy vs. selected features based on RIU4-LQP histogram features, the RFE method gave the highest CA on INbreast, while the FIR led to the best result on MIAS. Figure 15 shows different CA trends along with the feature number when using RIU4-LQP-K: for INbreast, CA curves saw a significant fall after peak points observed with RFE and FIR methods, while for MIAS, CA curves did not slide too much, but remained comparatively flat until the maximum feature number was used. Such inconsistency between the two datasets can be ascribed to the difference in image types and their classification criteria. As introduced in Section 2, mammograms in INbreast are FFDM images and MIAS is an SFM dataset, with different image qualities. INbreast mammograms are classified based on four BI-RADS density categories, while MIAS uses a three-class (i.e., F-G-D) standard. However, when we test the datasets using the concatenated feature set in Figure 16, a trend in consistent results is seen, with the best performance obtained using the RFE and the worst produced by the DPS.

The sensitivity of machine-learned solutions developed as diagnostic aids must be valid and applicable in clinical practice. The proposed method is based on a novel texture feature descriptor that extracts image microstructure patterns from mammograms. This procedure will not be affected by diagnostic ability of radiologists. However, our method needs a training set with radiologists' annotations for breast density, based on that a classifier is trained and used to predict the density labels for unseen mammograms. The proposed model was tested on two mammogram datasets that used different density classification criteria and ground truth annotated by different radiologists. Experimental results show that the method maintains good sensitivity on both datasets. Note that the mammogram images in the two datasets were carefully interpreted and annotated by at least two radiologists, and if there was a disagreement, a diagnosis from the third radiologist was required. As such, a comparatively accurate ground-truth set could be built for training the classification model, which guarantees the sensitivity of the method. If the annotations were given by only one radiologist who poses different diagnostic ability, this could lead to some inconsistent density labels compared to other radiologists' diagnoses and cause a variation in the sensitivity of the classification method.

The proposed method has some distinct advantages. Firstly, the proposed texture feature descriptor is found to be more powerful than the basic LBP/LQP methods in capturing local texture patterns. Secondly, our method, similar to LBP, is easily adaptable to similar applications requiring texture feature extraction. Finally, a novel aspect of this approach is the spatial distribution analysis of texture features, which is important but ignored in past work. Experimental results demonstrate that spatial distribution characteristics provide supplementary image features and effectively improve the classification performance.

In addition to the state-of-the-art classification results obtained, there are some limitations using the proposed methods. The proposed method requires a number of parameters (listed in Table 1) to be optimised first for achieving the best classification results. This is a one-off task which is not repeated in training or testing cycles. In addition, this work was initially inspired by related work of breast density classification, so we continue it with two improvement designs based on the LQP method. Our experimental settings also focus on the mammogram dataset test and its results analysis. We have not tested the effectiveness of the proposed methods on other image datasets. Although we have demonstrated that the proposed RIU4-LQP is rotation invariant and is powerful to capture more microstructures, different image types should be used to test it and give more evidence. We shall address this part of the research in future work.

7. Conclusions

This paper presents a robust texture feature descriptor for analysing mammograms and classifying breast density. Based on the conventional LQP operator, the rotation

invariant method and different transition number conditions are considered to develop the novel texture descriptor, RIU4-LQP. This paper employs Baddeley's K-inhom function to capture spatial distribution information of texture feature points, which is used to construct a new feature vector called K-spectrum. After concatenating the histogram and K-spectrum information, this paper also investigates three different feature selection schemes for optimising the initial feature set and improving the classification result. An SVM classifier is trained and used to predict the density labels for test images. Classification results are evaluated by classification accuracy (CA) and AUC, and statistical analysis is conducted between different methods. Two mammogram datasets, INbreast and MIAS, are used to test the proposed methods in our experiments. Experimental results demonstrate that the proposed method extracts robust and effective texture features in mammograms, which improve the classification performance significantly. Comparing with state-of-the-art methods, the classification accuracy by using our proposed approaches is competitive, reaching the best CA of 92.76% and 86.96% on INbreast and MIAS datasets.

References

1. Breast Cancer, U.S. Breast Cancer Statistics. Available online: http://www.breastcancer.org/symptoms/understand_bc/statistics (accessed on 20 May 2020).
2. Boyd, N.F.; Byng, J.W.; Jong, R.A.; Fishell, E.K.; Little, L.E.; Miller, A.B.; Lockwood, G.A.; Tritchler, D.L.; Yaffe, M.J. Quantitative classification of mammographic densities and breast cancer risk: Results from the canadian national breast screening study. *J. Natl. Cancer Inst.* **1995**, *87*, 670–675. [CrossRef] [PubMed]
3. Wolfe, J.N. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* **1976**, *37*, 2486–2492. [CrossRef]
4. Sickles, E.A.; D'Orsi, C.J.; Bassett, L.W.; Appleton, C.M.; Berg, W.A.; Burnside, E.S.; Feig, S.A.; Gavenonis, S.C.; Newell, M.S.; Trinh, M.M. ACR BI-RADS® mammography. In *ACR BI-RADS® Atlas, Breast Imaging Re-Porting and Data System*; American College of Radiology: Reston, VA, USA, 2013.
5. Brandt, K.; Scott, C.; Ma, L.; Mahmoudzadeh, A.; Jensen, M.; Whaley, D.; Vachon, C. Comparison of clinical and automated breast density measurements: Implications for risk prediction and supplemental screening. *Radiology* **2016**, *279*, 710–719. [CrossRef]
6. Mario, M.; Mislav, G.; Krešimir, D. Breast density classification using multiple feature selection. *Automatika* **2012**, *53*, 362–372. [CrossRef]
7. Li, S.; Wei, J.; Chan, H.-P.; Helvie, M.A.; Roubidoux, M.A.; Lu, Y.; Zhou, C.; Hadjiiski, L.M.; Samala, R.K. Computer-aided assessment of breast density: Comparison of supervised deep learning and feature-based statistical learning. *Phys. Med. Biol.* **2017**, *63*, 025005. [CrossRef]
8. Oliver, A.; Tortajada, M.; Lladó, X.; Freixenet, J.; Ganau, S.; Tortajada, L.; Vilagran, M.; Sentis, M.; Martí, R. Breast density analysis using an automatic density segmentation algorithm. *J. Digit. Imaging* **2015**, *28*, 604–612. [CrossRef]

9.   Subashini, T.; Ramalingam, V.; Palanivel, S. Automated assessment of breast tissue density in digital mammograms. *Comput. Vis. Image Underst.* **2010**, *114*, 33–43. [CrossRef]

10. Qu, Y.; Fu, Q.; Shang, C.; Deng, A.; Zwiggelaar, R.; George, M.; Shen, Q. Fuzzy-rough assisted refinement of image processing procedure for mammographic risk assessment. *Appl. Soft Comput.* **2020**, *91*, 106230. [CrossRef]

11. Tzikopoulos, S.D.; Mavroforakis, M.E.; Georgiou, H.V.; Dimitropoulos, N.; Theodoridis, S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput. Methods Programs Biomed.* **2011**, *102*, 47–63. [CrossRef] [PubMed]

12. Muhimmah, I.; Zwiggelaar, R. Mammographic density classification using multiresolution histogram information. In Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, Ioannina, Greece, 26–28 October 2006; pp. 26–28.

13. Chen, Z.; Denton, E.; Zwiggelaar, R. Local feature based mammographic tissue pattern modelling and breast density classification. In Proceedings of the Presented at the 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, China, 15–17 October 2011.

14. Zheng, Y.; Keller, B.M.; Ray, S.; Wang, Y.; Conant, E.F.; Gee, J.C.; Kontos, D. Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment. *Med. Phys.* **2015**, *42*, 4149–4160. [CrossRef]

15. Mohamed, A.A.; Berg, W.A.; Peng, H.; Luo, Y.; Jankowitz, R.C.; Wu, S. A deep learning method for classifying mammographic breast density categories. *Med. Phys.* **2018**, *45*, 314–321. [CrossRef] [PubMed]

16. Ahn, C.K.; Heo, C.; Jin, H.; Kim, J.H. A novel deep learning-based approach to high accuracy breast density estimation in digital mammography. In Proceedings of the SPIE Medical Imaging 2017: Computer-Aided Diagnosis, Orlando, FL, USA, 11–16 February 2017; Volume 10134. [CrossRef]

17. Lee, J.; Nishikawa, R.M. Automated mammographic breast density estimation using a fully convolutional network. *Med. Phys.* **2018**, *45*, 1178–1190. [CrossRef] [PubMed]

18. Li, C.; Xu, J.; Liu, Q.; Zhou, Y.; Mou, L.; Pu, Z.; Xia, Y.; Zheng, H.; Wang, S. Multi-view mammographic density classification by dilated and attention-guided residual learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1003–1013. [CrossRef] [PubMed]

19. Hamidinekoo, A.; Denton, E.; Rampun, A.; Honnor, K.; Zwiggelaar, R. Deep learning in mammography and breast histology, an overview and future trends. *Med. Image Anal.* **2018**, *47*, 45–67. [CrossRef] [PubMed]

20. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [CrossRef]

21. George, M.; Zwiggelaar, R. Comparative study on local binary patterns for mammographic density and risk scoring. *J. Imaging* **2019**, *5*, 24. [CrossRef] [PubMed]

22. Tan, X.; Triggs, W. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650. [CrossRef] [PubMed]

23. Rampun, A.; Morrow, P.; Scotney, B.; Winder, J. Breast density classification using local ternary patterns in mammo-grams. In Proceedings of the Image Analysis and Recognition 14th International Conference, ICIAR 2017, Montreal, QC, Canada, 5–7 July 2017; pp. 463–470. [CrossRef]

24. Nanni, L.; Lumini, A.; Brahnam, S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.* **2010**, *49*, 117–125. [CrossRef] [PubMed]

25. Rampun, A.; Scotney, B.W.; Morrow, P.J.; Wang, H.; Winder, J. Breast density classification using local quinary patterns with various neighbourhood topologies. *J. Imaging* **2018**, *4*, 14. [CrossRef]

26. Rampun, A.; Morrow, P.J.; Scotney, B.W.; Wang, H. Breast density classification in mammograms: An investigation of encoding techniques in binary-based local patterns. *Comput. Biol. Med.* **2020**, *122*, 103842. [CrossRef]

27. Moreira, I.C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M.J.; Cardoso, J. INbreast: Toward a full-field digital mammo-graphic database. *Acad. Radiol.* **2012**, *19*, 236–248. [CrossRef] [PubMed]

28. Suckling, J.; Parker, J.; Dance, D.; Astley, S.; Hutt, I.; Boggis, C.; Ricketts, I.; Stamatakis, E.; Cerneaz, N.; Kok, S.; et al. Mammographic Image Analysis Society (MIAS) Database v1.21 [Dataset]. Available online: https://www.repository.cam.ac.uk/handle/1810/250394 (accessed on 1 June 2021).

29. Baddeley, A.; Møller, J.; Waagepetersen, R. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **2000**, *54*, 329–350. [CrossRef]

30. Li, H.; Mukundan, R.; Boyd, S. A novel application of multifractal features for detection of microcalcifications in digital mammograms. In *Medical Image Understanding and Analysis. MIUA 2019. Communications in Computer and Information Science*; Zheng, Y., Williams, B., Chen, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1065, pp. 26–37. [CrossRef]

31. Muštra, M.; Grgic, M.; Rangayyan, R. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. *Med. Biol. Eng. Comput.* **2015**, *54*, 1003–1024. [CrossRef]

32. Slavkovic-Ilic, M.; Gavrovska, A.; Milivojević, M.; Reljin, I.; Reljin, B.; Marijeta, S.-I.; Ana, G.; Milan, M.; Irini, R.; Branimir, R. Breast region segmentation and pectoral muscle removal in mammograms. *Telfor J.* **2016**, *8*, 50–55. [CrossRef]

33. Li, H.; Mukundan, R.; Boyd, S. Robust texture features for breast density classification in mammograms. In Proceedings of the 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, 13–15 December 2020; pp. 454–459. [CrossRef]

34. George, M.; Denton, E.; Zwiggelaar, R. Mammogram breast density classification using mean-elliptical local binary patterns. In Proceedings of the SPIE 10718, 14th International Workshop on Breast Imaging, Atlanta, GA, USA, 6 July 2018. [CrossRef]
35. Guo, Z.; Zhang, L.; Zhang, D.; Zhang, S. Rotation invariant texture classification using adaptive LBP with directional statistical features. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010. [CrossRef]
36. Ripley, B. The second-order analysis of stationary point processes. *J. Appl. Probab.* **1976**, *13*, 255–266. [CrossRef]
37. Baddeley, A.; Turner, R. Practical maximum pseudolikelihood for spatial point patterns. *Aust. N. Z. J. Stat.* **2000**, *42*, 283–322. [CrossRef]
38. Baddeley, A.; Edge, R.; Turner, R. *Spatial Point Patterns: Methodology and Applications with R. Chapman & Hall/CRC Interdisciplinary Statistics*; CRC: Boca Raton, FL, USA, 2015.
39. Guo, Y.; Zhao, G.; Pietikäinen, M. Discriminative features for texture description. *Pattern Recognit.* **2012**, *45*, 3834–3843. [CrossRef]
40. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
42. Peng, L.; Lin, L.; Hu, H.; Ling, X.; Wang, D.; Han, X.; Chen, Y. Joint weber-based rotation invariant uniform local ternary pattern for classification of pulmonary emphysema in CT images. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017. [CrossRef]
43. Bouckaert, R. Choosing between two learning algorithms based on calibrated tests. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003.
44. Mehta, R.; Egiazarian, K. Dominant rotated local binary patterns (DRLBP) for texture classification. *Pattern Recognit. Lett.* **2016**, *71*, 16–22. [CrossRef]
45. Sreedevi, S.; Sherly, E. A novel approach for removal of pectoral muscles in digital mammogram. *Procedia Comput. Sci.* **2015**, *46*, 1724–1731. [CrossRef]
46. Rampun, A.; López-Linares, K.; Morrow, P.J.; Scotney, B.W.; Wang, H.; Ocaña, I.G.; Maclair, G.; Zwiggelaar, R.; Ballester, M.A.G.; Macía, I. Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network. *Med. Image Anal.* **2019**, *57*, 1–17. [CrossRef] [PubMed]