Article:

Beltran-Perez, C., Wei, H.-L. orcid.org/0000-0002-4704-7346 and Rubio-Solis, A. (2020) Generalized multiscale RBF networks and the DCT for breast cancer detection. International Journal of Automation and Computing, 17 (1). pp. 55-70. ISSN 1476-8186

https://doi.org/10.1007/s11633-019-1210-y

This is a post-peer-review, pre-copyedit version of an article published in International Journal of Automation and Computing. The final authenticated version is available online at: http://dx.doi.org/10.1007/s11633-019-1210-y.

Generalized Multiscale RBF Networks and the DCT for Breast Cancer Detection
Carlos Beltran-Perez, Hua-Liang Wei, Adrian Rubio-Solis
2020, vol. 17, no. 1, pp. 55-70, doi: 10.1007/s11633-019-1210-y

International Journal of Automation and Computing X(X), X X, X-X
DOI: XXX

Generalized Multiscales RBF Networks and the DCT for Breast Cancer Detection

Carlos Beltran-Perez[1,2] Hua-Liang Wei[1] Adrian Rubio-Solis[1]
[1]Automatic Control and Systems Engineering Department, The University of Sheffield, Sheffield UK. [2] Tecnologico de Monterrey, Campus Toluca, Mexico.

Abstract: The use of the multiscale generalized radial basis function (MSRBF) neural networks for image feature extraction and medical image analysis and classification is proposed for the first time in this work. The MSRBF networks hold a simple and flexible architecture that has been successfully used in forecasting and model structure detection of input-output nonlinear systems. In this work instead, MSRBF networks are part of an integrated Computer-Aided Diagnosis (CAD) framework for breast cancer detection, which holds three stages: an input-output model is obtained from the image, followed by a high-level image feature extraction from the model and a classification module aimed at predicting breast cancer. The first stage renders the image data into a multiple-input-single-output system. In order to improve the characterisation, the Nonlinear AutoRegressive with eXogenous inputs (NARX) model is introduced to rearrange the available input-output data in a nonlinear way. The Forward Regression Orthogonal Least Squares (FROLS) algorithm is then used to take advantage of the previous arrangement by solving the system as a model structure detection problem and finding the output layer weights of the NARX-MSRBF network. Once the network model is available, the feature extraction takes place by stimulating the input to produce output signals to be compressed by the Discrete Cosine Transform (DCT). Based on the new methodology, a CAD system for X-ray mammograms was integrated. To test the method performance, three different and well-known public image repositories were used: the mini-MIAS and the MMSD for mammography, and the BreaKHis for histopathology images. A comparison exercise was also made between different database partitions to understand the mammogram breast density effect in the performance since there are few remarks in the literature on this factor. Classification results show that the new CAD method reached an accuracy of 93.5% in mini-MIAS, 93.99% in DDSM and 86.7% in the BreaKHis. We found that the MSRBF networks are able to build tailored and precise image models and, combined with the DCT, to extract high-quality features from both black and white and coloured images.

Keywords: Nonlinear system Identification, Image Processing, Discrete Cosine Transform, Radial Basis Functions, Computer-Aided Diagnosis, Neural Networks.

---

1 Introduction

Digital image processing and computer vision techniques encompass an increasing variety of approaches to real-life problems. When it comes to image classification, image processing methods aim at recognising both visible and hidden patterns to enable a subsequent statistical inference process, oriented in the first place to extract feature values to feed such analytic process [1]. Among the last ones, there is increasing acceptance in the literature on system identification approaches, which are mainly focused on building models only based on the historical record of the system's inputs and outputs [2, 3]. These models are also capable of recognising and reproducing behavioural patterns from a system's behaviour without prior knowledge of its inner structure. Such pattern recognition capability is what makes system identification models highly appealing in image processing. Computer-aided diagnosis is another field of intense development that bridges image processing and computer vision disciplines to the medical field, especially in visualisation and diagnostic tasks. CAD has made the most of the current advances in intelligent systems. Examples are software supporting platforms for radiologists in decision-making [4, 6]. One of the most popular system identification approaches in CAD systems is represented by artificial neural networks (ANN) given their excellent modelling capacity. Moreover, many experts in CAD systems rely more frequently on the use of multi-layered ANN with the intent of obtaining even better approximations. However, the more the hidden layers are included in the network, the slower and more complex the model mapping becomes. Conversely, single hidden-layer networks, as radial basis functions, are known to be sufficient to estimate any continuous function, independently of the linearity degree [2, 5, 3]. Radial Basis Functions Networks (RBF) are popular kernel-based networks which represent a particular class of ANN. Kernels are mathematical functions contributing together to simulate a higher dimensional space from another one of a lower dimension to ease the setting up of relationships between the data by expressing it in a new way. RBF networks are efficient at solving nonlinear system identification problems in spite they hold a linearly weighted structure that eases the training and discards complex nonlinear procedures in the solution algorithm [5, 7]. Concerning nonlinear analysis, it is increasingly important for the study of real-life systems. As for digital image processing, its relevance has increased thanks to the proliferation of image usage in several application fields, the improvement of storage capacities and faster data transference speeds. An example of nonlinear analysis in image processing is that exclusively linear procedures in images may lead to poor operational results regarding edges, non-Gaussian noise and other random distortions, factors that can be especially ineffective when a high accuracy analysis is usually required. [8, 9, 10].

Research Article
Manuscript received date; revised date
Recommended by Associate Editor xxxx

Notwithstanding that RBF networks sound like a good choice due to their power of modelling and solving simplicity, the approximations they produce may lack the flexibility to model highly dynamic or rapid changing systems. An alternative to overcome such a limitation is the multiscale version of RBF, termed as generalized multiscale RBF networks, that provide a balance between the simplicity of modelling of RBF networks and the advantages provided by more complex networks [11, 12, 13].

Until the presentation of this work, MSRBF networks have not been used in image processing techniques nor in CAD systems. In this work, MSRBF networks are adopted and combined with the DCT to extract high-quality information from images with classification purposes. A basic outline of the proposed CAD system within the context of machine learning is shown in Fig. 1, where as detailed further in Section 3, training, and testing phases converge into classification to produce a computer-aided diagnosis using a clustering algorithm named *k-means++*. Also, a NARX-based input-output mapping of digital images is presented with the aim of making the digital image information consistent with nonlinear system identification problems. Tests results show that the new method is highly consistent as a CAD system in breast cancer image detection, an important and challenging public health problem, both in X-ray mammography and microscopy instances.

In terms of application, RBF networks have been used in a wide number of different real-world applications such as the modelling of complex systems [14], prediction of near-earth geomagnetic field [15], face recognition [16, 17], modelling and identification of dynamical systems [18], three-dimensional object recognition [19], and motor systems control [21], and in CAD systems involving pathological brain detection [22] and breast cancer detection [23, 24, 25, 26, 27].

This work puts forward a novel image processing framework for feature extraction based on an improved version of RBF networks. We add to this framework the advantages of the DCT to compress information. Finally, we successfully adapt the MSRBF methodology to CAD systems for breast cancer detection.

The rest of this paper is organised as follows: section 2 provides preliminaries while section 3 describes the proposed methodology for the classification and detection of breast cancer. In section 4, experiments and results are presented, where section 5 discusses the performance produced by the proposed methodology. Finally, section 6 draws conclusions.

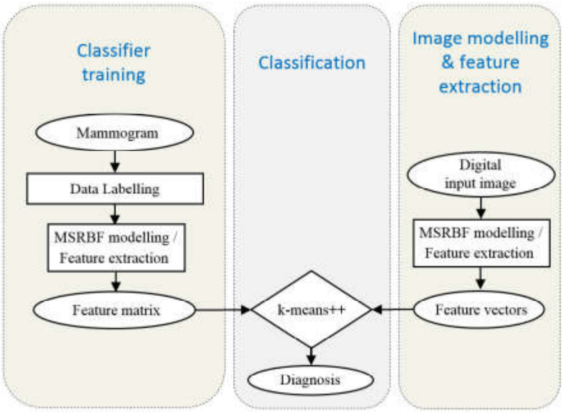


Figure 1 Role of the MSRBF-DCT into a classification-based CAD system.

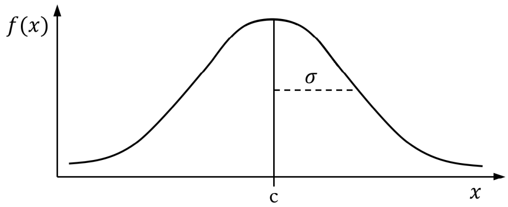


Figure 2 The shape of the Gaussian function contained in the RBF kernel.

2 Preliminaries

In this section a review of background material is briefly reviewed.

2.1 Traditional RBF and 2D MSRBF neural networks

Traditional RBF networks are known to be straightforwardly structured, but with a considerable power to identify a whole range of systems, including those with irregular data [20, 28]. However, single-scale RBF networks may have modest generalisation qualities [11]. MSRBF networks provide a favourable trade-off between easy to solve traditional RBF networks and the modelling advantages of multi-layer networks, which more than often include various hidden layers and involve nonlinear optimisation steps in the solution process [11]. MSRBF networks are multiscale because on the one hand, the kernel function included is Gaussian, and on the other, such Gaussian function has several widths or scales.

As mentioned, the present work includes the Gaussian kernel, for it allows to easily use centres and widths for an added modelling flexibility, as it enables the structure detection algorithm to choose from more options for a better representation. Fig. 2 exemplifies how the Gaussian neuron-function processes the x input data according to the μ and σ parameters (mean or kernel centres and standard deviations, widths or scales, respectively) generating a bell-shaped distribution curve in the output .

The Gaussian kernel is known to be a multidimensional universal approximator of functions converting a dimensional space into another corresponding one, but with different dimension (usually longer) that helps to linearly separate any type of input data with non-linear dependencies (like most of the real-life problems) to make features or information easier to extract and interpret by machine learning algorithms. RBF networks base their effectiveness on such advantage and approximate the unknown nonlinear function $\hat{f}$ utilising a weighted sum of Gaussian radial functions. Fig. 3 shows the typical architecture of RBF networks.

The RBF structure consists of three layers, where the first one represents the input data linked to the independent variables $x_1, \cdots, x_m$. The first layer is fully connected to the second intermediate layer, formed by the Gaussian neurons $\phi_1, \cdots, \phi_n$. The second intermediate layer is in turn fully connected to the third layer or output layer, employing the kernel weights $w_1, \cdots, w_n$, which are part of the result of the network training. Note that In the context of the neural network, the Gaussian functions parameters c_i for the centres and μ_i, for the widths are not given in the problem and thus must be computed automatically from data. For this reason, RBF networks are nonparametric methods. The general formulation of the standard RBF for a one-dimensional system is the following:

$$\hat{f}(\vec{x}(t)) = \sum_{i=1}^{M} \theta_i \phi_i(\vec{x}(t); \vec{\sigma}_i, \vec{c}_i) \tag{1}$$

where ϕ_i is the i^{th} neuron or Gaussian kernel, subindex i denotes the neuron number, M is the total number of neurons or kernels, $\vec{x}(t)$ is the vector of independent variables (which in the NARX model are rather regressors), $\vec{\sigma}_i = [\sigma_1, \cdots, \sigma_n]$ is the vector of parameters of the scales (or widths) and $\vec{c}_i = [c_1, \cdots, c_n]$ is the vector of parameters of the kernel centres. In such a way, the Gaussian kernel function for a one-dimensional system is stated as follows:

$$\phi_i(\vec{x}(t); \vec{\sigma}_i, \vec{c}_i) = \exp\left[\sum_{b=1}^{d}\left(\frac{x_b(t) - c_{i,b}}{\sigma_i}\right)^2\right] \tag{2}$$

where $d = n_u + n_y$, being n_u, n_y the maximum lags for the system input and output and b is an auxiliary value for indexing the regressive variables contained in $\vec{x}(t)$.

2.1.1 2D MSRBF neural networks

The MSRBF network implemented in this framework adopts the multiscale approach as a primal contribution along with the 2D perspective to attack the image processing problem. The multiscale extension to RBF, as the name suggests, multiplies the scales or widths of each kernel function with the aim of expanding the flexibility of the single hidden-layer neural network and better approach the nonlinear function $\hat{f}$.

Fig. 4 describes the structure of MSRBF neural networks, where the vectors of the input layer are fully connnected to the Gaussian kernel functions $\phi_{p,q,m}$ (defined originally in trditional RBF networks as ϕ_i). The number of functions represents the number of kernel centres c_m. The hidden-layer neurons are fully connected to the output layer by means of a series of weights $w_{p,q,m}$ corresponding to the model parameters $\theta_{p,q,m}$ stated below in (Eq. 3). The 2D MSRBF version replaces the vector of regressors $\vec{x}(t)$ (Eq. 2) by the two-dimensional vector $\vec{x}(i,j)$ (Eq. 4).

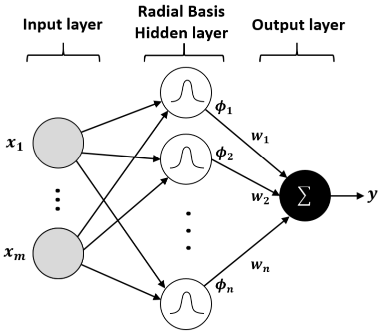


Figure 3 The multiple-input single-output architecture of a Gaussian RBFNN.

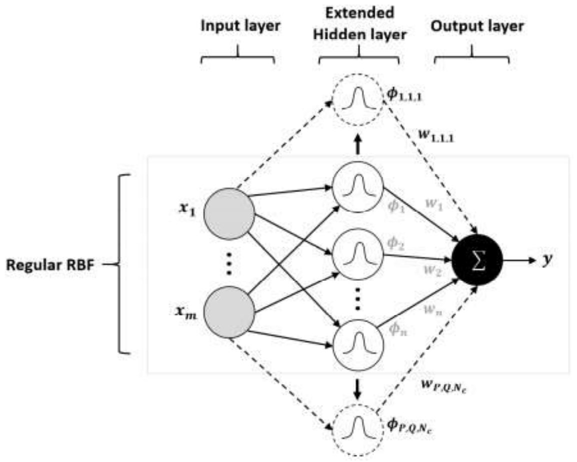


Figure 4 Increase in the number of RBF neurons produced by the multiscale approach.

The 2D MSRBF network implemented in this work presents the following mathematical structure, which is an adaptation of a definition presented in [11].

$$y(i,j) = \sum_{i=0}^{P}\sum_{j=0}^{Q}\sum_{m=1}^{N_c} \theta_{p,q,m}\phi_{p,q,m}(\vec{x}(i,j); \vec{\sigma}_m^{(p,q)}; \vec{c}_m) \tag{3}$$

where $y(i,j)$ is defined as $\hat{f}(\vec{x}(i,j))$ as is the system output, $\vec{x}(i,j)$ is the vector of bidimensional regressors composed of lagged inputs and outputs, $\sigma_m^{(p,q)}$ are the scales, $\vec{c}_m$ are the candidate centres with N_c representing their quantity in the network, $\phi_{p,q,m}$ are the basis functions and $\theta_{p,q,m}$ are the model weights to be estimated during training. In that way, the basis functions previously defined in traditional RBFs (Eq. 2), are defined in the 2D MSRBF network as:

$$\phi_{p,q,m}(\vec{x}(i,j); \vec{\sigma}_m^{(p,q)}, \vec{c}_m) = \exp\left[-\sum_{b=1}^{d}\left(\frac{x_b(i,j) - c_{m,b}}{\sigma_{m,b}^{(p,q)}}\right)^2\right] \tag{4}$$

where in the same fashion, $\phi_{p,q,m}$ is the general Gaussian kernel, $\sigma_m^{(p,q)}$ are the Gaussian multiscales, c_m are the Gaussian centres, b is an auxiliary value indexing the variables contained in vector $\vec{x}(i,j)$ and $d = n_y + n_{u1} + n_{u2} + n_{u3}$, being n_{u1},n_{u2},n_{u3} and n_y the regressive variables of the multiple-input-single-output network design suggested in this paper. However, special attention must be paid in the determination of the Gaussian parameters.

2.1.2 The discrete cosine transform

The discrete cosine transform [29] is a function that computes a sequence of discrete values out of a first sequence. The resulting coefficients are calculated by summing cosine functions valued at various frequencies, producing an oscillating effect in the resulting numbers. A relevant contribution of the DCT is the data compression capability for audio and image processing applications, including pattern recognition [1].

A simple way to explain the DCT is to imagine a vector of a certain length and the DCT as a transformation matrix so that the product of the first two results in a second vector of the same length but with the energy concentrated in fewer coefficients. Because of this quality, it is easy to reorder and leave out the less important values. More formally, the DCT for a data sequence $X(i), i = 0, 1, \cdots, (N-1)$ is:

$$F_x(u) = \begin{cases} \frac{\sqrt{2}}{N} \sum_{i=0}^{N-1} X(i), & u = 0 \\ \frac{2}{N} \sum_{i=0}^{N-1} X(i) \cos \frac{(2i+1)u\pi}{2N}, & u = 1, 2, \cdots, (N-1) \end{cases} \tag{5}$$

where $F_x(u)$ is the ith DCT coefficient and u is a vector of values to be compressed.

3 Methodology

In this section the proposed Generalized Multiscales RBF Network using a Discrete Cosine (MSRBF-DCT) Transform for the classification of breast cancer images is described.

3.1 The MSRBF-DCT methodology

The MSRBF-DCT feature value extraction method bases its logic on four main algorithms: conversion of the image data into the NARX format, the multiscale version of RBF networks, the FROLS algorithm and the discrete cosine transform. Fig. 5 shows the MSRBF network information flow within the new methodology.

The adaptation of the proposed methodology into the CAD point of view involved the image partition into subimages or regions of interest (ROIs) in the first place. In this work, ROIs are regarded here as the standard processing units, where a 64 x 64 pixel-size was assigned to better enclose the ROIs such as tumours and microcalcifications including the surrounding regions. Besides, a splitting process was included to deepen the analysis scope of this work as for the objects' position detection in the ROI area and to produce a two-fold and parallel characterisation, where a complete subimage is observed on the left side, followed by its dual partition on the right. The functional objective of this conversion is to diversify the features contained within the resulting vector, especially when the objects are out of the image centre. Please note that after the feature extraction of individual partitions, the values must form a single vector representing the original ROI.

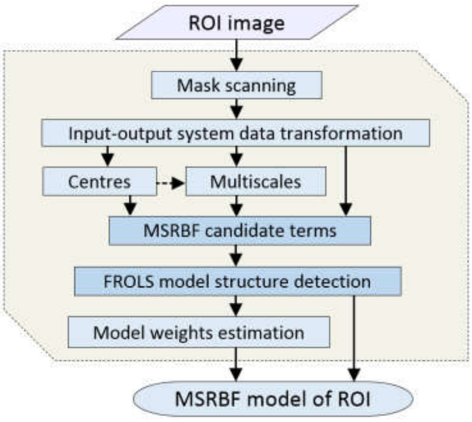


Figure 5 MSRBF model approximation flowchart.

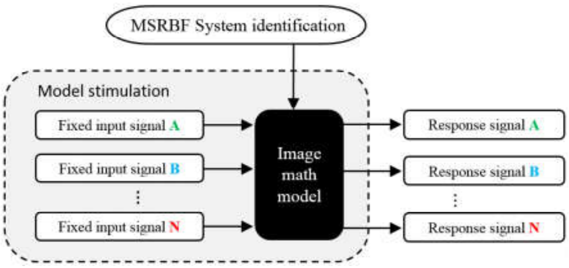


Figure 6 MSRBF model stimulation procedure.

As for the subject of the image processing, each ROI split is read and stored according to the input-output system format at first. Then, such data must be processed to derive a convenient number of data centres. The referred centres represent artificial neurons or functions contained in the singleton hidden network layer. In this paper, the mathematical structure of each neuron-function is the standard Gaussian function, defined alongside the complete processing in the following section.

At the end of the image modelling process, the structure selection algorithm FROLS comes into play to assess the candidate neurons and include the most representative terms into the model.

Once the model is available, a set of input signals is used to excite the model and generate a corresponding output signal series, whose values are processed via the DCT and assembled to obtain a set of feature vectors. As illustrated in Fig. 6, this process of feature extraction is repeated over all the mammogram's ROIs to compare the final vectors to pre-tagged samples corresponding to healthy, benign or malignant class utilising a distant-based classification algorithm.

3.2 Discrete-time system structuring

At this stage, the new method aims to scan the image data similarly to a time series, where instead of discrete time periods, adjacent pixel neighbourhoods without overlap lay distributed along the image (Fig. 7).

From the input-output systems perspective, the way of representing such data must be congruent with the following equation:

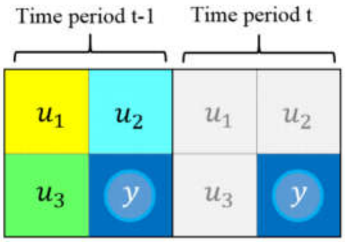


Figure 7 2D pixel-level equivalence of time domain input-output variables in the new framework.

$$y(t) = \hat{f}(\mathbf{x}(t)) + e(t) \tag{6}$$

in which the output $y(t)$ is explained by a nonlinear function $\hat{f}$ and an error sequence $e(t)$. In the case of the RBF, the nonlinear function can be conveniently represented as linear-in-the-parameters as follows:

$$\hat{f}(\vec{x}(t)) = \sum_{i=1}^{M} \theta_i \phi_i(\vec{x}(t); \vec{\sigma}_i, \vec{c}_i) \tag{7}$$

where ϕ_i is the Gaussian function (defined later in Section 3.3.2), $\vec{x}(t)$ is the vector of independent variables (also known as regressors), $\vec{\sigma}_i = [\sigma_1, \cdots, \sigma_n]$ is the vector of parameters of the scales or widths and $\vec{c}_i = [c_1, \cdots, c_n]$ is the vector of parameters of the kernel centres. Based on the 2D-NARX model describing a single-input-single-output (SISO) system, the nonlinear function is compound together by a list of input-output regressors as follows [30, 2]:

$$\begin{aligned} y(t) =& \hat{f}[y(t-1), y(t-2), \cdots, y(t-n_y), \\ & u(t-d), u(t-d-1), \cdots, u(t-d-n_u)] + e(t) \end{aligned} \tag{8}$$

where $\hat{f}$ is an unknown nonlinear function, $y(t)$ is the sequence of the system output, $u(t)$ is the sequence of the system input, n_u and n_y are the maximum lags for the system inputs and output (in this work fixed at 1), and d is a time delay auxiliary value, set here to $d = 1$. Based on more complex NARX representations for a multiple-input-single-output (MISO) system that aim at producing a richer feature extraction, the vector $\vec{x}(t) = [x_1(t), \cdots, x_d(t)]^T$ is defined as a set of regressors in the following ways:

$$\vec{x}_b(t) = \begin{cases} y(t-b), & 1 \le b \le n_y \\ u(t-(b-n_y)), & n_y + 1 \le b \le n_y + n_u \end{cases} \tag{9}$$

where n_u and n_y are the maximum lags for the input u and the output y respectively and b is an auxiliary value. Taking into account the previous description, the vector of basic cross-coupled regressors to be combined in the 2D case within the function expansion and defined as $x_b(i,j)$. The last representation shift aimed at addressing the bi-dimensional image processing problem usually present in the field of medical image processing, compared to the simpler time series problem that depends on a single variable. Please note that Section 3.2.2 details the 2D modelling of this process. With this in mind, Eq. (10) defines the set of regressors of as follows:

$$\vec{x}_b(i,j) = \begin{cases} y(i, j-2b), & 1 \le b \le 1 \\ u_1(i-1, j-1-2(b-1)), & 2 \le b \le 2 \\ u_2(i-1, j-2(b-2)), & 3 \le b \le 3 \\ u_3(i, j-1-2(b-3)), & 4 \le b \le 4 \end{cases} \tag{10}$$

where the maximum lags $n_y, n_{u1}, n_{u2}, n_{u3}$ were fixed in 1 and b is an auxiliary value, following the actual model set up to be seen in Section 4.1.

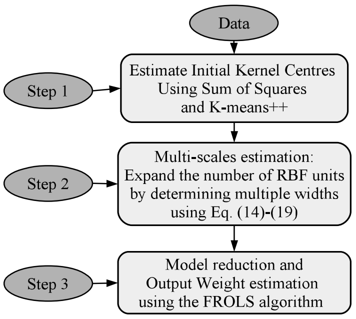


Figure 8 Flowchart for the MSRBF parameter identification.

3.3 Parameter Estimation for the MSRBF networks

As described in Fig. 8, the new framework takes advantage of the MSRBF networks efficiency to build image models from which functional feature vectors are extracted with the help of the DCT. The vectors are then included in the training matrix of the CAD system. It is noted here that the training of the MSRBF networks and the CAD system are separate but consecutive processes. While the kind of learning of the former is unsupervised, the latter is supervised. Unlike multilayer networks, where habitually all parameters are simultaneously optimised via backpropagation (which is slow and can get stuck in local minima), RBF networks can take in many other training schemes [43, 2]. In this work, the parameter estimation of the MSRBF networks consists of three separate steps, (a) centres estimation, (b) widths estimation and (c) output layer weights estimation. In the last step, we used a popular OLS-based training strategy (detailed in Section 3.4) that uses a linear-in-the-parameters representation of the network that makes it easy to solve [2].

3.3.1 Kernel centres estimation

The proposed method includes the implementation of an adaptive algorithm to determine the number of centres N_c (and therefore, the number of Gaussian functions in the hidden layer), taken from the work of [11] and [31]. In the first place, *the sum-of-squares clustering* algorithm acts as a criterion for estimating the number of centres. The algorithm includes the following steps:

1. The input data, composed of N rows and p columns, is divided into an arbitrary number of k initial groups $G_1, \cdots, G_k$.

2. The geometry centre (centroid) $\vec{c}_j$ of each group G_j is obtained.

3. The variability d_j per group is estimated by summing all distances of $\vec{z}_i$ with respect to the centroid $\vec{c}_j$:

$$d_j = 2 \sum_{i \in I_j} \|\vec{z}_i - \vec{c}_j\|^2 \tag{11}$$

where the vector $\vec{z}_i$ is the ith row of input data belonging to the group G_j.

4. The variability function of k, W_k, is estimated by summing the d_j of all groups.

5. The process is repeated from step 1 to 4 using different k values to estimate their variability function W_k.

6. The difference in the variability function of k values involves the following formula:

$$\mathrm{DIFF}(k) = (k-1)^{2/p}W_{k-1} - k^{2/p}W_k \qquad (12)$$

7. The following equation helps to compute the effectiveness of each k by comparing the values obtained in step 6:

$$E(k) = |\mathrm{DIFF}(k)/\mathrm{DIFF}(k+1)| \qquad (13)$$

8. Finally, the recommended k value, or number of centres N_c, is that one maximising the function $E(k)$.

After the estimation of the number of kernel centres, the *k-means++* algorithm [32] is used to compute a corresponding number of centroids from the $N \times p$ size input data matrix.

3.3.2 Multi-scales estimation

As for the scales or kernel widths, a two stage process was carried out according to the strategy recommended in [11]. The idea behind aims at estimating a single scale by basis function ϕ_i in the first place followed in turn by the computation of the quantiles (points taken at regular intervals) resulting from the first scale. Thus, the equations below define the first single scales.

$$\sigma_y = max\{y(i,j)\} - min\{y(i,j)\} \qquad (14)$$

$$\sigma_{u_r} = max\{u_r(i,j)\} - min\{u_r(i,j)\} \qquad (15)$$

where σ_y is the initial scale for the output and σ_{u_r} are the initial scales for the inputs $u_r = [u_1, \cdots, u_R]$ of the MISO system. For the calculation of the multiple final scales, the following formula is used to expand σ_y and σ_{u_r}:

$$\Lambda_m^{(p,q)} = diag\big[\underbrace{(\sigma_{y,m}^{(p)})^2,\cdots,(\sigma_{y,m}^{(p)})^2}_{\text{output y}},\underbrace{(\sigma_{u_r,m}^{(q)})^2,\cdots,(\sigma_{u_r,m}^{(q)})^2}_{\text{input } u_r}\big] \qquad (16)$$

where m connotes the kernel centre, $\Lambda_m^{(p,q)}$ are the covariance matrices for the values $p = 0, \cdots, P$ and $q = 0, \cdots, Q$, u_r are the system inputs and $\sigma_{y,m}^{(p)} = 2^{-p}\sigma_y$ and $\sigma_{u_r,m}^{(q)} = 2^{-q}\sigma_{u_r}$ are the quantiles linked to the output and input initial scales. In this work, the values of P and Q were fixed in 1 and the number of system inputs, R, in 3. Therefore the scales contained in (16) can be disaggregated as follows:

$$\sigma_{y,m}^{(p)} = \big[(\sigma_y 2^0)^2, (\sigma_y 2^{-1})^2\big] = \big[(\sigma_y)^2, (\sigma_y/2)^2\big],\ \forall\ m \qquad (17)$$

$$\sigma_{u_r,m}^{(q)} = \big[(\sigma_{u_r} 2^0)^2, (\sigma_{u_r} 2^{-1})^2\big] = \big[(\sigma_{u_r})^2, (\sigma_{u_r}/2)^2\big],\ \forall\ m \qquad (18)$$

where $u_r = [u_1, u_2, u_3]$ corresponds to the system inputs, σ_y and σ_{u_r} are the initial scales obtained in (14) and (15) and m indicates the kernel centre. With the above definitions, a more explicit representation of the multiscale radial basis functions expressed in Eq. (7) is:

$$\phi_{p,q,m}(\vec{x}(i,j);\vec{\sigma}_m^{(p,q)},\vec{c}_m) = \exp\Big[-\sum_{b=1}^{n_y}\Big(\frac{x_b(i,j)-c_{m,b}}{\sigma_{y,m}^{(p)}}\Big)^2 - \sum_{b=n_y+1}^{d}\sum_{r=1}^{R}\Big(\frac{x_b(i,j)-c_{m,b}}{\sigma_{u_r,m}^{(q)}}\Big)^2\Big] \qquad (19)$$

where $d = n_y + n_{u1} + n_{u2} + n_{u3}$ and $R = 3$ is the number of system inputs. After the kernels' definition and estimation, a matrix of resulting values is built to allow the FROLS algorithm to detect the model structure and ease the output weights estimation.

3.3.3 Output-layer weights estimation

After the calculation of the Gaussian functions, the associated firing strengths are taken as candidates during the parameter estimation of the weights connecting the hidden and the output layers. The estimation starts with the selection of the candidate neurons contributing most to explain the system output $y(t)$ by means of the FROLS algorithm (Section 3.4) so that a much smaller subset of candidate neurons remains available. Thanks to the linear-in-the-parameters representation of the RBF defined in (Eq. 7), the vector of weights is easily obtained by solving the problem as a system of linear equations for $\boldsymbol{\theta}$:

$$\boldsymbol{h\theta} = \boldsymbol{y} \qquad (20)$$

where $\boldsymbol{h}$, $\boldsymbol{\theta}$ and $\boldsymbol{y}$ are the vectors of firing strengths, weights and system output observations respectively.

3.4 Model structure detection

Following up the definitions in Section 3.3.2, the number of scales or RBF widths for a MISO system is $N_s = (P+1)(Q+1)^R$. The proposed model was set up at $P = Q = 1$ and $R = 3$, so the initial number of Gaussian centres k was scaled up $N_s = (2)(2)^3 = 16$ times in the MSRBF network, in a similar fashion to the architecture expansion shown earlier in Fig. 4. The final number of M candidate neurons of the MSRBF network is $M = N_cN_s$, where N_c is equal to the number of initial centres k recommended by the sum-of-squares algorithm (Section 3.3.1).

The listing of the candidate neurons M gains importance in the structure detection algorithm since it makes use of a D dictionary containing M candidate functions, from which the selection process is carried out. The D dictionary enlists the basis functions in the following manner:

$$D = \{\phi_m(\cdot); m = 1, \cdots, M\} \qquad (21)$$

where $\phi_m \in \phi_{p,q,m}(\cdot,\vec{\sigma}_m^{(p,q)},\vec{c}_m); p = 0,\cdots,P; q = 0,\cdots,Q; m = 1,\cdots,N_c$. The FROLS algorithm [33] is designed to build, term by term, the best and most concise models from D, the pool of candidate terms. It bases initially on the original OLS estimator [34], which iteratively looks for the candidate terms that best minimise the error respecting the model output $y(t)$ by using the ERR estimator. The orthogonalisation algorithm helps to exclude from the selection the candidate terms which content is redundant to that already included in the model.

However, the ERR estimator in the OLS is biased towards the inclusion of terms sorted first in the model equation [2]. The FROLS algorithm contributes to removing that shortcoming by adding a reordering of the candidate terms within the equation, leaving out biases of any kind in the inclusion of the most significant candidates. The stop-criterion of the FROLS algorithm changed to an *IF* function to limit the number of terms. Thus, the model detection ends up when the error tolerance is satisfied or when the model is long enough.

3.5 Feature extraction and the DCT

The feature extraction module of this framework works out from the image models estimated by the MSRBF network. In this work a finite number of fixed signals are used to obtain responses from the MSRBF model. However, unlike the 2D NARX model, the featuring process of the model's response signal includes the discrete cosine transform (DCT) to improve the representativeness of the image values concerning the quality and the quantity. This improvement is because the featuring of the model's output response signal takes place through a direct data transformation instead of external measures based on statistical measures, which can be useful but can ignore information when measuring from the outside. Section 2.1.6 and the following subsection explain the basics of the DCT algorithm.

Fig. 9 shows a scheme of the MSRBF-based image processing method, where the stimulus of the image model and the DCT play essential roles to produce feature values. Another advantage behind using the DCT is the ease to obtain even-sized feature vectors given that it allows the choice of an identical number of coefficients per image.

3.6 Classification and detection

The classification module is the connection between the feature extraction process and CAD systems. It links the feature vectors from the supervised, pre-labelling task with the unlabelled feature vectors of the image to classify according to the case study's classes. Fig. 1, shown in the first section, aims to ease the information flow visualisation of the proposed framework, where we observe two separate parallel processes of image data extraction converging into the detection/diagnosis module, based on classification. The difference between classification and diagnosis is that the first one associates the input vector with a class. The diagnosis module uses the classification results to interpret the patient's condition and displays a message easy to understand.

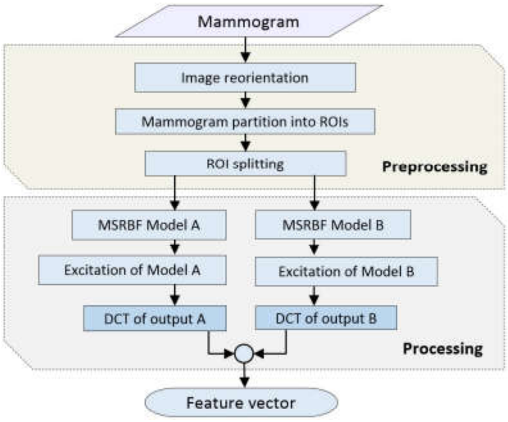


Figure 9 Flowchart of MSRBF-based image processing for feature extraction.

For classification, the distance-based *k-means++* algorithm was selected [32]. The standard algorithm *k-means* inspired this technique. However, *k-means++* holds the advantage of using an improved seeding method to choose centres, producing an efficient classification up to 70% faster [32].

4 Experiments and Results

In this section we evaluated the proposed MSRBF-DCT framework performance through three popular medical image data sets: two consisting of X-ray mammograms (mini-MIAS and DDSM) and one made up from histopathological (microscopy-level) samples (BreaKHis). The reason for choosing the first two case studies follows that both mini-MIAS and DDSM are high quality benchmarks that are also public and have been widely used, which facilitated comparison with previous work. The foregoing is also relevant because there are quite a few methods in the literature for breast cancer detection that claim interesting results but use databases that are not publicly available. Despite the advantages, we note that mini-MIAS and DDSM are not recent, so we included a recently published database related to breast cancer which could allow us to test our method both with non-mammographic and non-grayscale images at once to broaden the experimental scope. Thus, we incorporated the BreakHis database to our tests. As for the experimental conditions, all programs were coded in MATLAB R2014b 64-bit and executed in a computer running in Windows 7 Professional operating system with an Intel (R) Core (TM) i5-4590 processor at 3.30GHz speed, running MATLAB 2014b.

4.1 The mini-MIAS Database

The assessment of the MSRBF-DCT method started with the mini-MIAS database of mammograms [35], a public repository including 322 high-quality grayscale X-ray films of 1024 × 1024 pixels of the medio-lateral oblique view of the breast in PGM format. In spite of the high-quality database, numerous artefacts and scanning imperfections within images were present, as unknown breast position and orientation (leftnor right), duct tapes, orientation tags, low-intensity labels and scanning artefacts. The evaluation goal was to assess the quality of the feature extraction method by evaluating its classification quality for a defined set of mammograms with information attached to them regarding the medical condition class and the background tissue type.

The database distribution regarding the breast tissue type is detailed in Table 1. A randomised data-splitting of the 322 breast scans of the database was made, following a 65% to 35% ratio for training and testing with the aim of reducing the chance of attaining biased performance metrics. Furthermore, to counteract the high image variability

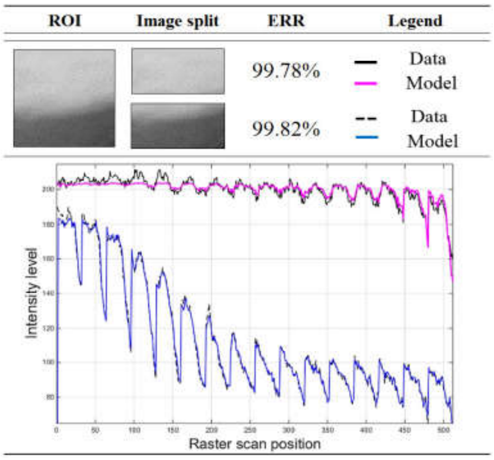


Figure 10 Two pairs of fit-to-data curves and ERR values. ROI from [35].

regarding the breast tissue type, $n = 4$ different training and testing scenarios with different tissue background composition were carried out aimed at, on the one hand leaving in evidence potential differences in the classification results and on the other to get a set of final performance measures with minimal bias. In that way, the global accuracy for a n number of training and testing scenarios is defined by $Accuracy_n = average(Accuracy(i))$, where $i = [1, \cdots, n]$ symbolises the ith test.

Given the high image resolution and the reduced dimension of several ROIs, we decided to make the processing at the subimage (namely ROI) level. During training, we assembled a matrix of 21,637 feature vectors, which data labelling produced 95.5% of these belonging to normal and 4.5% to abnormal, from which 2.29% belonged to benign and 2.21% to malign. The error tolerance of the ERR stop-criterion was 0.15%, and the maximum number of terms was 2.

After the partitioning of mammograms (divide and conquer strategy), the ROI feature extraction and the labelling of the complete database by following the database documentation, the construction of a training matrix for a specific training-testing partition made only necessary to generate a subset of the full training matrix by removing from it the mammogram-related vectors selected for testing. The initial evaluation aimed at judging the ability of the model to fit the observational data. Figure 10 shows the example of a dense tissue-type sub-image or ROI, its subdivisions (for a two-fold characterisation) and the error reduction ratio (ERR) of the models concerning the data of each case. The table also includes a plot overlying the fit of both models versus the original data. It is possible to observe from the chart that the model adjustment is reliable in both cases since the curves of the predicted output and the original data overlap each other in both pairs of curves.

Table 1 mini-MIAS database breast-type distribution [35].

	Fatty	**Fatty-Glandular**	**Dense**	**Total**
Cases	106	104	112	322
%	32.92	32.3	34.78	100

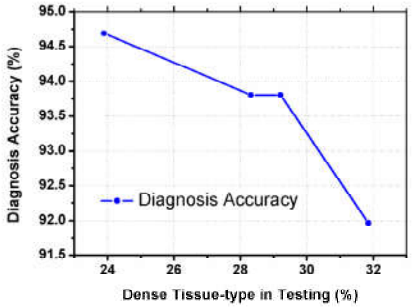


Figure 11 Accuracy as a function of the presence of dense mammograms in the test set.

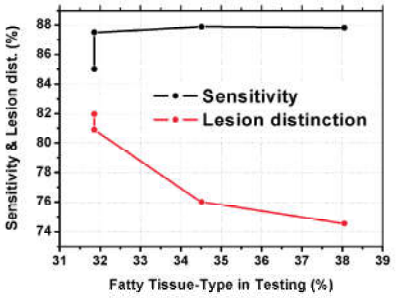


Figure 12 Sensitivity and lesion distinction accuracy as functions of the presence of fatty mammograms in the test.

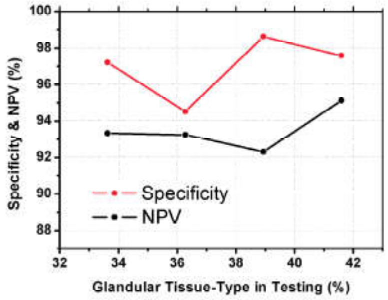


Figure 13 Specificity and NPV as functions of the presence of glandular mammograms in testing.

Table 2 Classification results in mini-MIAS database [35].

Statistical Measure	**Average result (%)**
Accuracy	93.57
Sensitivity	87.05
Specificity	96.97
PPV	93.79
NPV	93.50
Lesion Distinctinon	78.35

Table 3 MSRBF-DCT performance results by breast-tissue-type ratio (mini-MIAS Database [35]).

		Test 1	**Test 1**	**Test 1**	**Test 4**
Ratio	Fatty (%)	31.86	31.86	38.05	34.51
	Dense (%)	29.20	31.86	28.32	23.89
	Glandular (%)	38.94	36.28	33.63	41.59
Measure	Accuracy (%)	93.81	91.96	93.81	94.69
	Sensitivity (%)	85.00	87.50	87.80	87.88
	Specificity (%)	98.63	94.52	97.22	97.50
	PPV (%)	97.14	89.74	94.74	93.55
	NPV (%)	92.31	93.24	93.33	95.12
	Pathology Identification (%)	81.97	80.88	74.55	76.00

After this point, the assessment of the feature extraction included the consistency tracking regarding the distance between the feature vectors. This relationship was directly proportional to the visual image similarity.

The overall performance numbers of this study are presented in Table 2. It is noticeable that values of sensitivity, PPV and lesion distinction are not as high as expected, possibly because of the high similitude of dense-healthy and glandular-healthy tissue with many abnormal tumours. Among all the values, it stands out that the tumour distinction was the lowest value of all the registered ones. This difference is due on the one hand to the relative scarcity of abnormal samples, which represented only 4.5% of the total of the labelled samples and on the other to the fact that machine learning methods are generally more efficient to a more significant number of samples available for training [38].

The experimental performance results of the four tests from different partitions of the database are described in Table 6. In the first place, the percentages by mammogram-type included in each one of the tests are displayed. The overall results are quite encouraging in the four tests, especially regarding accuracy, specificity and NPV.

As assumed, it is possible to note that the composition of the elements in the test impacts in a visible manner the classification results. The latter is an interesting point to take into account as this factor may well lead various breast cancer classification studies to confusing results. To ease the analysis of the resulting variations of the classification concerning the mammogram-type composition in the testing set, exciting trends in the results were found and plotted. Fig. 11 shows a negative relationship found between the presence of dense mammograms in the test set and the classification accuracy. Such divergence can be the result of that dense-healthy images are visually similar to tumours of high density, producing false detections.

On the other hand, Fig. 12 suggests that there was a lessening ability to distinguish the abnormality class with the increase of fatty mammograms presence in the test, which was opposite to the expected result, given that fatty tissue tends to have translucence, which would make the classification procedures easier.

However, and in favour of the latter hypothesis, the change of the sensitivity values in the different set compositions suggested a positive trend between fatty tests sets and the effective detection of any abnormalities (benign or malign). The last point, together with the accuracy decrease in denser compositions, led to finding a positive relationship between the MSRFB DCT classification accuracy with fatty mammograms and a negative relationship with dense mammograms. Although these results may seem intuitive, it is necessary to carry out more discriminative studies of this type in the future, especially with other methods of featuring and classification to draw more generalised conclusions regarding the breast cancer detection.

As for the variation of the presence of glandular mammograms in the testing set, Fig. 13 shows a very light direct relation of specificity and NPV with the presence of glandular tissue. Although the trend was not significant enough to be taken into account, it was expected, however, that the presence of glandular tissue, on the contrary, would actively impede the quality of the classification results.

4.2 The DDSM Database

The second database used for evaluation is the Digital Database for Screening Mammography, DDSM [36], which together with mini-MIAS, is one of the best-known sources of images in the development of breast cancer detection methods. This repository of the University of Florida contains 2,620 clinical cases, where each one presents two mediolateral oblique views (MLO) and two caudal cranial to total 10,480 mammograms. The collection provides more precisely the density of the breast tissue, defined in BI-RADS categories of density ranging from one to four. To improve the comparison with previous databases, we decided to consider only images with the MLO view. We also used the CBIS-DDSM [37] collection, which is a curated version of DDSM and verified by medical experts, containing only benign and malignant cases of masses and calcifications.

In the CBIS-DDSM data, images are split into training and testing sets and by type of abnormality, where images of the ROIs are included to help the user to locate tumours. As for the healthy images, we used the utility provided in [41] to convert images from the DDSM native format (LJPEG) to a format readable by MATLAB. With the above conditions we obtained a set of 1783 mammograms randomly divided into 65% training and 35% testing, where 38,776 ROIs and feature vectors were extracted. Note that the original distribution of DDSM regarding breast-type is displayed in Table 7. All ROIs with some pathology obtained during training were increased in number through image rotations, flips and shift positioning around the ROI. Similarly to the previous study, we chose to split the set of tests into four equal segments, although with a different BI-RADS distribution of breast density. Although the increase in the dimension of the images of DDSM contributed to obtaining more ROIs for training, this factor played against the processing time per mammogram, which extended 30% on average. In the case of microcalcifications, an exhaustive training was required, since their detection during early stages of training is very limited, especially when the tissue is dense and the appearance of microcalcifications seems to be higher. The average results for all tests are reflected in Table 8, while the results by test also including pathology identification are shown in Table 9. Similarly to the mini-MIAS study case, the diagnosis accuracy in DDSM was inversely proportional to the presence of dense samples in the testing partition (see Fig. 14), confirming the importance of tissue-type distribution in mammogram classification problems.

Likewise, there was a notable increase in sensitivity and specificity regarding the absence of dense images in the test partition (Fig. 15). For this reason, the MSRBF-DCT-based CAD system, which has shown to be quite credible with all kind of samples, is highly reliable when fewer dense mammograms are present in the test set.

We also noticed that although DDSM provides a higher number of samples for training, there was no notable improvement in the statistical measures concerning the mini-MIAS experiment. It is possible that this is due to a much more significant presence of calcifications in the DDSM repository. A comparison of the new method with previous work is presented in Table 10. In general, the new method is competitive but did not reach such a high sensitivity as some other approaches. The authors believe that this imperfection is the product of the training strategy carried out, since to avoid a high occurrence of false positives caused by the resemblance between healthy dense tissue and some types of a tumour, it was necessary to increase the number of dense samples in the training database.

Table 4 DDSM breast-type distribution of converted images.

	Dens.1	**Dens.2**	**Dens.3**	**Dens.4**	**Total**
Cases	429	511	420	423	1,783
%	24.05	28.64	23.58	23.73	100

Table 5 Classification results in the DDSM database [36].

Statistical Measure	**Average Result %**
Accuracy	93.99%
Sensitivity	92.65%
Specificity	94.35%
PPV	81.97%
NPV	97.90%
Pathology Identif	85.72%

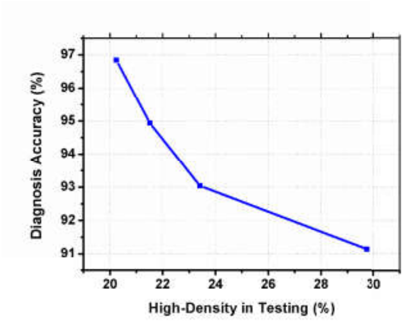


Figure 14 Accuracy as a function of the presence of high-density mammograms in the test (DDSM database [36]).

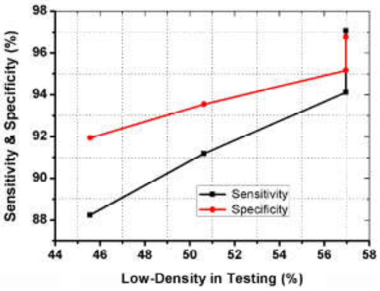


Figure 15 Sensitivity and Specificity as a function of the presence of low-density mammograms in the test (DDSM database [36]).

4.3 The BreaKHis Dataset

The last image set for assessment of the new CAD system is the BreakHis dataset [46], a more recent public repository (2015) for breast cancer image classification, which unlike the collections previously utilised in the manuscript, images are available at the cellular (histopathological) level [44]. Unlike X-ray mammography, BreaKHis images are in colour and only benign and malignant images are presented, so the tests in this study case are centred in the ability of the framework to do an adequate binary classification.

Table 6 MSRBF-DCT performance results by breast-tissue-type ratio (DDSM database [36]).

		Test 1	**Test 1**	**Test 1**	**Test 4**
Ratio	Density 1 (%)	29.11	21.52	23.42	21.52
	Density 2 (%)	27.85	35.44	27.22	24.05
	Density 3 (%)	22.78	20.89	25.95	24.68
	Density 4 (%)	20.25	21.52	23.42	29.75
Measure	Accuracy (%)	96.84	94.94	93.04	91.14
	Sensitivity (%)	97.06	94.12	91.18	88.24
	Specificity (%)	96.77	95.16	93.55	91.94
	PPV (%)	89.19	84.21	79.49	75.00
	NPV (%)	99.17	98.33	97.48	96.61
	Pathology Identification (%)	87.88	81.25	87.10	86.67

Table 7 Comparison of MSRBF-DCT with previous work in mammography databases.

Model	Image Set	**Accuracy %**	**Specificity %**	**Sensitivity %**
2D-NARX[30]	mini-MIAS	91.00	93.00	89.50
ELM[23]	mini-MIAS	91.00	90.00	98.00
GLCM[24]	mini-MIAS	93.90	97.20	91.50
ICA-RBF[25]	mini-MIAS	88.20	N/A	N/A
LDA-ANN[26]	mini-MIAS	93.10	99.00	83.00
GPZM[27]	mini-MIAS	89.30	83.50	93.40
GPZM[27]	DDSM	87.27	82.51	90.33
MSRBF-DCT	**mini-MIAS**	**93.50**	**87.00**	**96.90**
MSRBF-DCT	**DDSM**	**93.99**	**92.65**	**94.35**

The images were acquired from samples (histology slides) obtained by the *P&D* Lab, Brazil, using open surgical biopsy. The final diagnoses were made by expert pathologists. The samples were shot with an Olympus BX-50 microscope and a Samsung SCC-131AN digital colour camera to obtain 3-channel RGB images in TrueColor (image composed by pixels defined by three values) [46, 45]. The samples are distributed into four magnification factors: 40×, 100×, 200× and 400×, summing 7,909 items. The BreaKHis image distribution is shown in Table 8.

As for image feature extraction, a first challenge was to make the MSRBF to extract features from a 3-channel RGB colour space, so we chose to preprocess colour images by dividing them into three channels (R,G,B) to make the MSRBF-based feature extraction work at the channel level (Fig. 16).

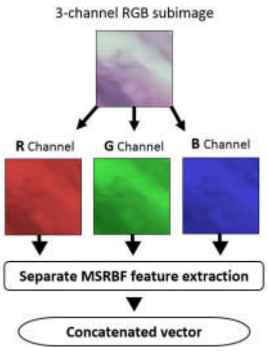


Figure 16 Colour image decomposition into 3 channels. ROI from [46]

Table 8 BreaKHis database distribution by magnification factor and class.

Magnification	**Ben.**	**Mal.**	**Total**
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909

Table 9 Comparison of depicting features with methods described in [46] that use the BreaKHis database

Method	**No. of features**
CLBP	1352
GLCM	13
LBP	10
LPQ	256
ORB	32
PFTAS	162
MSRBF-DCT-3-channels	**150**

Table 10 MSRBF-DCT accuracy results by test and magnification factor (BreaKHis database [46])

	Class	40×	100×	200×	400×
Test 1	B	0.83	0.87	0.80	0.80
	M	0.86	0.87	0.84	0.82
Test 2	B	0.85	0.83	0.87	0.82
	M	0.85	0.86	0.82	0.80
Test 3	B	0.88	0.84	0.83	0.82
	M	0.88	0.86	0.85	0.83
Test 4	B	0.86	0.89	0.82	0.83
	M	0.86	0.88	0.82	0.83
Test 5	B	0.82	0.86	0.85	0.78
	M	0.86	0.87	0.83	0.83

Table 11 MSRBF-DCT average accuracy and standard deviations per magnification factor

Class	Magnification Factor			
	40 ×	100 ×	200 ×	400 ×
Benign	84.6 ± 2.5	85.9 ± 2.4	83.5 ± 2.6	81.1 ± 1.8
Malign	86.3 ± 1.0	87.0 ± 1.0	83.4 ± 1.1	82.1 ± 1.3

Table 12 Comparison of the MSRBF-DCT method with previous work using the BreaKHis database[46].

Method	Magnification Factor			
	40×	100×	200×	400×
PFTAS-QDA[46]	83.8 ± 4.1	82.1 ± 4.9	84.2 ± 4.1	82.0 ± 5.9
PFTAS-SVM[46]	81.6 ± 3.0	79.9 ± 5.4	85.1 ± 3.1	82.3 ± 3.8
PFTAS-RF[46]	81.8 ± 2.0	81.3 ± 2.8	83.5 ± 2.3	81.0 ± 3.8
GLCM-1-NN[46]	74.7 ± 1.0	76.8 ± 2.1	83.4 ± 3.3	81.7 ± 3.3
AlexNet[45]	85.6 ± 4.8	83.5 ± 3.9	83.1 ± 1.9	80.8 ± 3.0
DeCAF[47]	84.6 ± 2.9	84.8 ± 4.2	84.2 ± 1.7	81.6 ± 3.7
CSDCNN[48]	95.8 ± 3.1	96.9 ± 1.9	96.7 ± 2.0	94.9 ± 2.8
MSRBF-DCT-3-Channels	$\mathbf{85.8 \pm 2.0}$	$\mathbf{86.7 \pm 1.8}$	$\mathbf{83.4 \pm 1.9}$	$\mathbf{81.8 \pm 1.6}$

Hence, the features extracted per channel were joined back to derive an expanded feature vector representing the colour space image. We found that such a shift did not affect considerably the computational time in any of the method stages. With the adjustment from one to to three channels, the number of features describing the image was increased in the same relation from 50 to 150 features. Table 9 represents this attribute and compares it with other methods presented in [46] also using the study case of this section. The training in the BreaKHis database aimed at reducing the effect of the high image resolution and the great number of samples. into the computational burden and training time. At the same time, we followed a strategy similar to that of previous work for favouring the comparability of results. Similarly to the work presented in [46] and [45] and to the evaluations of MSRBF-DCT with mammography databases, we followed a 70% and 30% ratio to break apart at random the samples into training and testing respectively into each of the two classes and four magnification factors. In that way, four different random partitions were made, from which the average accuracy accuracy by class and magnification factor was estimated.

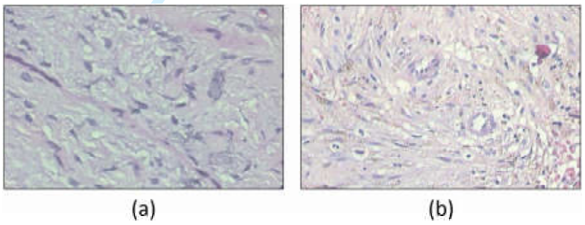


Figure 17 Example of a benign tumour (a) which wrongly fell into the malign class by cause of malign samples belonging to (b). Images from BreaKHis [46].

In favour of reducing the computational load, we chose to reduce the dimension of the original images by 50%, without this significantly affecting the fitting capacity of the model, as has been reported in [44] and [45]. In our case, the image reduction went from 700 × 400 to 350 × 200 pixels by using bicubic interpolation as a resizing procedure, for it gives a good balance of output quality and execution time. As with the X-ray images, the divide and conquer scheme produced per image 15 subimages, without overlap, of 64 × 64 pixels. To determine what class an image belongs to, it suffices that 8 of 15 subimages fall into the same class.

Taking into account the number of divisions per image (15), the learning-testing partition ratio (0.7 and 0.3) and the total number of mutually exclusive images per benign and malignant class in Table 8 (2,480 and 5,429), we obtained 26,040 and 57,000 images for training and 11,116 and 24,435 images for testing throughout each random partition.

Concerning the performance evaluation after the five random database partitions, the training data was presented to the CAD system, so that it could gather the features related to images from each class. Table 13 summarizes the classification results of the CAD framework of 5 tests using different random partitions.

Table 11 reflects the average results of the five tests performed with different partitions. It is noteworthy that the magnification factor that obtained the highest accuracy was 100×, while the most difficult to predict was that of 40×. This trend is also reflected in the previous work shown in Table 12. The comparison shows that our CAD system separated remarkably well benign and malignant samples, compared to most existing methods, including convolutional neural networks like DeCAF and AlexNet. However, our method, along with the rest, was clearly surpassed by a more recent convolutional network called CSDCNN. Despite this, we noticed that the MSRBF-DCT networks are generally speaking a reliable method. According to our observations, the above is due to the fact that the histology slides of the samples with the smaller zoom (40×) have more distinctive features such as nuance, texture and colour between classes, while those of greater zoom (200× and 400×) show more defined morphological features like shapes and dark patterns, but these are quite similar between the two classes in many instances. An example of the visual parallelism between high magnification samples is that of Fig. 17, in which a benign image of a phyllodes tumor (a) fell wrongly into the malignant class thanks to samples belonging to a mucinosis carcinoma malign image (b) stored in the training set. It is visible that, although the hue is different, the morphological similarity between (a) and (b) is relevant.

5 Conclusions

The proposed methodology presents a convenient neural network-based modelling framework, originally designed to approximate nonlinear observational input-output series, as a novel contribution to digital image feature extraction and CAD system. Furthermore, the Discrete Cosine Transform algorithm was successfully incorporated to make the most of the MSRBF networks.

The experiments aimed at appraising the tumour detection in X-ray mammograms and histopathological images for three different public databases showed up that the method is competitive compared to well-known previous CAD systems for breast cancer based on system identification and artificial neural networks. Among other performance values, the proposed method reached a classification accuracy above 93% in the two mammography databases and 86.7% in BreaKHis. While the MSRBF-DCT method is not perfect, some below-average metrics results may be also explained by an imperfect data labelling strategy. The improvement in the training strategy from the first database to the second can be appreciated in more balanced statistical measures. On the other hand, a null incidence of identical feature vectors for a multitude of visually similar images (except from the totally black or white samples) in the three databases lead us to think that the two-fold ROI characterization coupled with the DCT increased the model ability to extract both size and object position features from the image effectively, information which otherwise could be lost.

As regards the comparative performance between the two mammography databases, more similarities than differences were found. Although the DDSM includes more data for a proper training, the increase in accuracy was not significant (less than 1%) compared to mini-MIAS, perhaps because DDSM also presents a higher proportion of microcalcifications, which are difficult to diagnose. As for the statistical measures, there were notable improvements from the first to the second study, especially in sensitivity (5%), PPV (11%) and pathology-type identification (7%). These differences may be due to several factors, such as a better documentation on the pathology localization, thanks to the CBIS-DDSM repository.

Regarding the breast density in mammograms, several machine learning studies for breast cancer detection do not report the proportion of dense samples in their experiments, which is to the best of our knowledge a central factor that is capable of producing changes in the global performance. In relation to comparisons with previous work, the CAD system performance showed to rise to the challenge in most cases.

Concerning the BreaKHis, subimage level analysis allowed high-resolution images to be processed by dividing the computational load into smaller processes as with mammograms repositories. Regarding colour images, we noticed that 150 feature vectors, derived through MSRBF-DCT, are efficient to represent 3-channel images. We realized that higher magnification factors like 200× and 400× produced more classification errors. We attributed this trend to the fact that numerous samples between classes are morphologically similar at that level. We also observed that the benign class produced more errors, possibly because it had fewer samples to train compared to the malignant class and because certain types of tumours were easier to be misclassified. For instance, we found that phyllodes tumours, from benign, were mistaken several times because of their strong similitude to specific malignant subclasses such as mucinosis carcinoma.

In databases two and three (DDSM and BreaKHis), the comparison of the model performing with different training-testing compositions led the writers to infer that getting results with a single partition in heterogeneous databases is undesirable as it may generate unwanted trends in dependence on the percentage of challenging elements.

Future work includes the use of a receiver operating characteristic (ROC) curve to determine the best decision threshold regarding the training strategy to best balance sensitivity and specificity. The transfer of the recommended methodology to other medical study areas such as brain diseases and lung cancer detection is desirable. Also, the successful expansion of the characterization method to colour images enables the use of the MSRBF-DCT approach to case studies involving real-world object detection or skin and face recognition. Finally, we could couple the present feature extraction procedure with more advanced classifiers in order to enhance the integration of MSRBF networks with CAD systems.

References

[1] Bovik, A.C. ed., The essential guide to image processing. Academic Press, 2009.

[2] Billings, S.A., Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains, John Wiley & Sons, 2013.

[3] Liu, H., Xiao, G.F., Tan, Y.L. and Ouyang, C.J., 2018. Multi-source remote sensing image registration based on contourlet transform and multiple feature fusion. International Journal of Automation and Computing, pp.1-14.

[4] Doi, K., Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics, vol. 31, no. $4-5$, pp. 198–211, 2007.

[5] Cybenko, G.. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, vol. 2, no. 4, pp. 303–314, 1989.

[6] Vipparthi, S.K. and Nagar, S.K., 2016. Local extreme complete trio pattern for multimedia image retrieval system. International Journal of Automation and Computing, 13(5), pp.457-467.

[7] Li, A., Lu, Z., Wang, L., Xiang, T., Li, X. and Wen, J.R., 2017. Zero-shot fine-grained classification by deep feature learning with semantics. arXiv preprint arXiv:1707.00785.

[8] Sonka, M., Hlavac, V. and Boyle, R. Image processing, analysis, and machine vision. analysis, and machine vision. Cengage Learning, 2014.

[9] Kotropoulos, C. and Pitas, I., Nonlinear model-based image/video processing and analysis, 2001.

[10] Pessoa, L.F. and Maragos, MRL-filters: A general class of nonlinear systems and their optimal design for image processing, IEEE Transactions on Image Processing, vol. 7, no. 7, pp. 966–978, 1998.

[11] Billings, S.A., Wei, H.L. and Balikhin, Generalized multiscale radial basis function networks, IEEE Neural Networks, vol. 20, no. 10, pp. 1081–1094, 2007.

[12] Rubio-Solis, A., Martinez-Hernandez, U. and Panoutsos, G., 2018, July. Evolutionary extreme learning machine for the interval type-2 radial basis function neural network: A fuzzy modelling approach. In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.

[13] Rubio-Solis, A. and Panoutsos, G., 2014. Interval type-2 radial basis function neural network: a modeling framework. IEEE Transactions on Fuzzy Systems, 23(2), pp.457-473.

[14] Rubio-Solis, A., Melin, P., Martinez-Hernandez, U. and Panoutsos, G., 2018. General Type-2 Radial Basis Function Neural Network: A Data-Driven Fuzzy Model. IEEE Transactions on Fuzzy Systems, 27(2), pp.333-347.

[15] Wei, H.L., Billings, S.A. and Balikhin, M., Prediction of the Dst index using multiresolution wavelet models, Journal of Geophysical Research: Space Physics, 109(A7), 2004.

[16] Er, M.J., Chen, W. and Wu, S., High-speed face recognition based on discrete cosine transform and RBF neural networks, IEEE Transactions on neural networks, vol. 16, no. 3, pp. 679–691, 2005.

[17] Er, M.J., Wu, S., Lu, J. and Toh, H.L., Face recognition with radial basis function (RBF) neural networks, IEEE transactions on neural networks, vol. 13, no. 3, pp. 697–710, 2002.

[18] Chen, S.A.B.S. and Billings, S.A., Neural networks for nonlinear dynamic system modelling and identification, International journal of control, vol. 56, no. 2, pp. 319–346, 1992.

[19] Poggio, T. and Edelman, S., A network that learns to recognize three-dimensional objects. Nature, vol. 343, no. 6255, pp. 263–266, 1990.

[20] Yao, Z.J., Bi, J. and Chen, Y.X., 2018. Applying deep learning to individual and community health monitoring data: a survey. International Journal of Automation and Computing, 15(6), pp.643-655.

[21] Yang, R., Er, P.V., Wang, Z. and Tan, K.K., An RBF neural network approach towards precision motion system with selective sensor fusion. Neurocomputing, vol. 199, pp. 21–39, 2016.

[22] Lu, Z., Lu, S., Liu, G., Zhang, Y., Yang, J. and Phillips, P., A pathological brain detection system based on radial basis function neural network. Journal of Medical Imaging and Health Informatics, vol. 6,no. 5, pp. 1218–1222, 2016.

[23] Vani, G., Savitha, R. and Sundararajan, N., 2010, December. Classification of abnormalities in digitized mammograms using extreme learning machine. 11th International Conference on Control Automation Robotics & Vision, pp. 2114–2117, 2010.

[24] Pratiwi, M., Harefa, J. and Nanda, S., Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. Procedia Computer Science, vol. 59, pp. 83–91, 2015.

[25] Christoyianni, I., Koutras, A., Dermatas, E. and Kokkinakis, G., Computer aided diagnosis of breast cancer in digitized mammograms. Computerized medical imaging and graphics, vol. 26,no. 5, pp. 309–319, 2002.

[26] Iftikhar, K., Anwar, S., Haq, I.U., Khan, M.T. and Akbar, S.R., An Optimal Neural Network Based Classification Technique for Breast Cancer Detection. Journal of Engineering and Applied Sciences, vol. 35,no. 1, pp. 51–58, 2016.

[27] Singh, S.P. and Urooj, S., An improved CAD system for breast cancer diagnosis based on generalized pseudo-Zernike moment and Ada-DEWNN classifier. Journal of medical systems, vol. 40,no. 4, pp. 105, 2016.

[28] Poggio, T. and Girosi, F., A theory of networks for approximation and learning (No. AI-M-1140). MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 1989.

[29] Ahmed, N., Natarajan, T. and Rao, K.R., Discrete cosine transform. IEEE transactions on Computers, vol. 100,no. 1, pp. 90–93, 1974.

[30] Beltran-Perez, C. and Wei, H.L., Digital image classification and detection using a 2D-NARX model. In 2017 23rd International Conference on Automation and Computing (ICAC), pp. 1–6, 2017.

[31] Krzanowski, W.J. and Lai, Y.T., A criterion for determining the number of groups in a data set using sum-of-squares clustering. Biometrics, pp. 23–34, 1988.

[32] Arthur, D. and Vassilvitskii, S., k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, pp. 1027–1035, 2007.

[33] Billings, S.A., Korenberg, M.J. and Chen, S., Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. International Journal of Systems Science, vol. 19,no. 8, pp. 1559–1568, 1988.

[34] Korenberg, M., Billings, S.A., Liu, Y.P. and McIlroy, P.J., Orthogonal parameter estimation algorithm for nonlinear stochastic systems. International Journal of Control, vol. 48,no. 1, pp. 193–210, 1988.

[35] Suckling J, P., The mammographic image analysis society digital mammogram database. Digital Mammo, pp. 375–386, 1994.

[36] Heath et al.(2000)Heath, Bowyer, Kopans, Moore, and Kegelmeyer Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore, and W. Philip Kegelmeyer. The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography*, pages 212–218. Medical Physics Publishing, 2000.

[37] Lee et al.(2017)Lee, Gimenez, Hoogi, Miyake, Gorovoy, and Rubin Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L. Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4:170177, 2017.

[38] Whitehill et al.(2009)Whitehill, Wu, Bergsma, Movellan, and Ruvolo Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R. Movellan, and Paul L. Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.

[39] Bowyer, K., Kopans, D., Kegelmeyer, W.P., Moore, R., Sallam, M., Chang, K. and Woods, K., The digital database for screening mammography. In Third international workshop on digital mammography, vol. 58, pp. 57, 1996.

[40] D'Orsi, C.J., Mendelson, E.B. and Ikeda, D.M. eds., ACR breast imaging and reporting data system: breast imaging atlas. American College of Radiology, 2003.

[41] Sharma, A., DDSM Utility. GitHub, GitHub repository, https://github. com/trane293/DDSMUtility, 2015.

[42] Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M. and Rubin, D.L., A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data, vol. 4, pp. 170–177, 2017.

[43] Schwenker, Kestler, and Palm Friedhelm Schwenker, Hans A. Kestler, and Günther Palm, Three learning phases for radial-basis-function networks. Neural networks, 14(4-5):439–458, 2001.

[44] Hafemann, L.G., Oliveira, L.S. and Cavalin, P., 2014, August. Forest species recognition using deep convolutional neural networks. In 2014 22nd International Conference on Pattern Recognition (pp. 1103-1107). IEEE.

[45] Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L., 2016, July. Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) (pp. 2560-2567). IEEE.

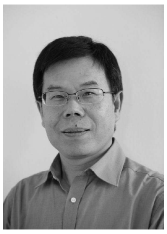
[46] Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L., 2016, July. Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) (pp. 2560-2567). IEEE.

[47] Fabio A. Spanhol, Luiz S. Oliveira, Paulo R. Cavalin, Caroline Petitjean, and Laurent Heutte. Deep features for breast cancer histopathological image classification. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1868–1873. IEEE, 2017.

[48] Zhongyi Han, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. Breast cancer multi-classification from histopathological images with structured deep learning model. In *Scientific reports*, 7(1):4172, 2017.

Carlos Beltran-Perez received a B.Eng. and the M.Sc. from the Autonomous Metropolitan University, Mexico and from the Autonomous University of Nuevo Leon, Mexico, respectively. He recently completed a Ph.D. in Automatic Control and Systems Engineering from the University of Sheffield, UK. He is currently a Senior Lecturer at the Departments of Industrial Engineering and Computer Engineering at Tecnologico de Monterrey University, Campus Toluca, Mexico and a graduate researcher at the University of Sheffield, UK. His current research interests include identification and modelling for complex systems, medical image processing, NARMAX methodologies, multilayer neural networks, forecasting and time series analysis.

E-mail: carlos.beltran@tec.mx (Corresponding author)



Hua-Liang Wei received the Ph.D. degree in the Department of Automatic Control and Systems Engineering, the University of Sheffield, UK, in 2004. He is currently a Senior Lecturer in the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, U.K. His current research interests include system identification and data analytics for complex systems, data driven modeling and data mining, NARMAX methodology and its applications, statistical digital signal processing, machine learning and neural networks, spatio-temporal system modeling, neuro-wavelet models for learning, nonstationary (time varying) process modeling, forecasting of complex dynamic processes, generalized regression analysis, linear and nonlinear optimization, and multidisciplinary applications in medicine and biomedicine, medical informatics, space weather, and environmental sciences.

E-mail: w.hualiang@sheffield.ac.uk

Adrian Rubio-Solis received the B.Eng. degree in robotics and the M.Sc from the Instituto Politecnico Nacional, Mexico City, Mexico, in 2003 and 2005 respectively. He also completed a PhD in Automatic Control and Systems Engineering from the University of Sheffield, UK in 2014. He is currently working as research fellow at the submarine robotics Laboratory, CIDESI Queretaro, Mexico. His current research include data-driven modelling, image processing, fuzzy control theory and machine learning applied to submarine robotics.

E-mail: a.rubiosolis@sheffield.ac.uk