Proceedings Paper:
Beltran Perez, C. and Wei, H.-L. (2018) Image classification using generalized multiscale RBF networks and discrete cosine transform. In: 2018 24th International Conference on Automation and Computing (ICAC). 24th International Conference on Automation and Computing (ICAC), 06-07 Sep 2018, Newcastle upon Tyne, UK. IEEE. ISBN: 9781862203419.

https://doi.org/10.23919/IConAC.2018.8748965

Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018

This is peer-reviewed final accepted manuscript

Image Classification Using Generalized Multiscale RBF Networks and Discrete Cosine Transform

Carlos Beltran Perez
Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
cbeltranperez1@sheffield.ac.uk

Hua-Liang Wei
Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
w.hualiang@sheffield.ac.uk

Abstract— **The use of the multiscale generalized radial basis function (MSRBF) network for image feature extraction is proposed for the first time. The MSRBF network holds a simple but flexible structure capable to modelling complex systems. However MSRBF is originally designed to identify observational-type input-output systems. We aim to use this efficient network to get to concise but accurate models of digital images thanks to: a) the use of multiple scales in the RBF kernel width, and b) the adoption of the forward regression orthogonal least squares (FROLS) algorithm to refine the model structure selection. Thereafter the new tailored model is excited to produce output signals aimed at be compressed by the discrete cosine transform (DCT), adopted in this work to compact signals' energy into a few coefficients. To recognise images as MSRBF networks, a mathematical modelling was done by considering the first ones as multiple-input single-output systems. Based on the new methodology a novel computer aided diagnosis (CAD) system for cancer detection in X-ray mammograms was designed. Classification results show that the new CAD method helped reach a competitive diagnostic accuracy of 93.5%. It was similarly found that the MSRBF network is able to construct tailored and precise image models.**

Keywords; Nonlinear system identification; Image processing; Discrete Cosine Transform; Radial Basis Functions; Computer Aided Diagnosis; Neural Networks

I. INTRODUCTION

Digital image processing techniques encompass an increasing variety of approaches that can help to recognize or extract patterns to enable a subsequent statistical inference such as the extraction of feature values for classification [1]. Among the latter, there has been a growing acceptance of system identification approaches, which are mainly oriented to build models based only on the historical record of system's inputs and outputs [2]. This kind of models may also recognize patterns from a system behaviour without prior knowledge of its structure. This pattern recognition capability is what make system identification models so attractive in image processing. Computer aided diagnosis (CAD) is a field of intense development that incorporates image processing in the medical field and has made the most of advances in intelligent systems to support radiologists in decision-making [3]. One of the most popular system identification approaches in CAD systems are the artificial neural network (ANN) given their remarkable modelling capability. We have observed that in CAD systems many authors like to use multi-layered ANN to try to obtain better approximations. However, the more hidden layers are included in the network the more complex the training becomes, even though it has been proved that one single hidden-layer is enough to estimate any continuous function [2][4]. A special class of ANN are the single-hidden layer networks also known as radial basis functions (RBF). These networks hold a linearity weighted structure that ease the training and avoids nonlinear procedures [5]. The typical architecture of RBF is illustrated in Figure 1. Although simple RBF sounds as a good choice, the models they produce may lack of flexibility to model dynamic or discontinuous systems. An alternative to this limitation is the *multiscale* version of RBF, termed Generalized Multiscale RBF Networks (MSRBF) that provide a trade-off among simple RBF Networks and the advantages provided by more complex networks [5]. To the best of our knowledge, MSRBF networks has not been employed in CAD systems or image processing techniques. In this work we adopt the MSRBF networks philosophy and combine it with DCT to extract information from images with classification purposes. The method was tested as a CAD system for breast cancer. Previous work on CAD systems involving system identification and/or neural networks is abundant, but we made a review of the most representative techniques. A technique unrelated to CAD but similar to our method uses RBF and DCT for face recognition [6]. The approach uses DCT as dimensionality reduction method,

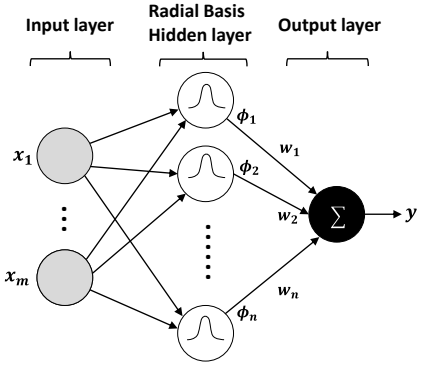


Figure 1. Multiple-input single-output architecture of a RBFNN

Fisher's linear discriminant to extract feature values and RBF for classifying the vectors. Although metrics are not comparable the authors were able to create an insensitive

to luminosity variations method. In [7] a CAD technique for breast cancer based on the polynomial NARX model, the FROLS algorithm and the k-means classifier is presented. The method propose the modelling of an image as an input-output dynamic system in order to put together a parametric model based on the image to be characterized. The classification results demonstrated a competitive performance. In [8] a CAD system using the extreme learning machine (ELM) for breast cancer detection was introduced. This unsupervised training method beaten the metrics of previous methods, but the algorithm depends on a very accurate parameter setting. In [9] RBF neural networks are used for classification and the gray-level co-occurrence matrix for feature extraction in a CAD system for breast cancer detection. The classification results of RBFNN and the back propagation NN was compared, resulting the proposed method to be more accurate in accuracy (93.98% vs.79.5%) and tumour distinction (100% vs. 89.47%). In [10] an easy to implement CAD approach uses independent component analysis for feature extraction and RBFNN for classification to attain an accuracy of 88.23% and abnormality distinction rate of 79.31%. In [11] an NN technique for breast cancer detection is introduced. It uses a gray level co-occurrence matrix for feature extraction and the scaled conjugate gradient back propagation to train the network. The classification results were positive for accuracy and sensitivity (93.1% of, 99%) but only moderately good for specificity (83%). Finally, an integrated CAD system for breast cancer detection using a particular network kind is presented [12]. The authors propose the generalized pseudo-Zernike moment for feature extraction which is claimed to be robust to noise, and a novel adaptive differential evolution wavelet neural network is recommended as classifier. In the tests, two mammogram databases are used, MIAS and DDSM, attaining accuracy rates of 89% and 87% respectively.

This work puts forward a novel image processing framework for feature extraction based on an improved version of RBF networks, adds to them the advantages of DCT information compression and adapts successfully the new methodology into a new CAD systems for breast cancer detection. The following chapter describes the information flow and the logic behind the proposed method, including new and adopted procedures. Chapter III shows the experiments and results of the methodology. Chapter IV presents a discussion on the findings, difficulties and future work.

II. METHODOLOGY

The DCT MSRBF feature value extraction method is mainly based on three algorithms: RBF neural network in the multi-scale version, the FROLS algorithm and the Discrete Cosine Transform. However, to adapt this

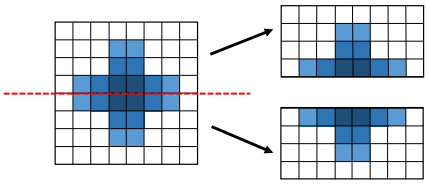


Figure 2. Subimage splitting for a two-fold characterization

methodology to a CAD system, the image is initially divided into sub-images, which is the level at which the computational process is performed, as we considered that the size of a subimage (64 x 64 pixels) is suitable for containing the region of interest (ROI) such as a tumours or microcalcifications. To improve the focus range regarding tumour location within the image, an additional image splitting was implemented to boost the characterization through a two-fold simultaneous processing (Figure 2). Then, each sub-subimage is read and structured as a nonlinear autoregressive with exogenous input (NARX) [13] model array, which is in turn processed by the MSRBF network to produce a new arrange of candidate terms made of kernel functions. Then, the FROLS algorithm evaluates the candidates and creates a compact but accurate image model. When the model is ready, a series of input reference signals are used to excite the model to generate a corresponding series of output signals, which are finally processed via the DCT and put together to obtain a feature vector. This same process is performed with all mammography's subimages in order to compare, by means of a classification algorithm, their values with those of other samples previously tagged as healthy, benign or malignant). The CAD classification scheme is resumed in Figure 3.

A. *Discrete-time system structuring*

At this stage the method scans the image data and store it as a time series so that pixel neighbourhoods along the image put together a list of input-output observations with a structure based on the NARX model as follows [7][2]:

$$y(t) = F[y(t-1), y(t-2), \dots, y(t-n_y), u(t-d), u(t-d-1), \dots, u(t-d-n_u)] + e(t) \quad (1)$$

where $F[\cdot]$ is a nonlinear function, $y(t)$ the system output sequence, $u(t)$ the system input sequence, n_y and

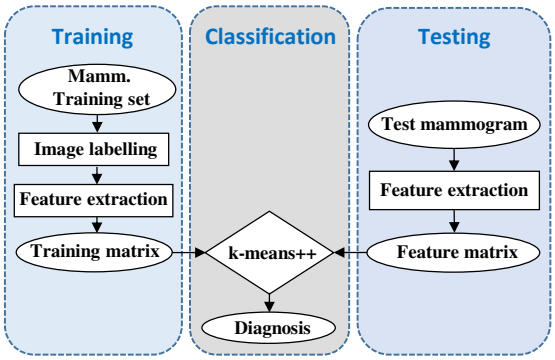


Figure 3. Adaptation of MSRBF DCT to a CAD system

n_u the maximum lags for the system output and input (set in this work =1), and d a time delay set here to $d = 1$.

B. MSRBF network

The generalized radial basis function networks (MSRBF) were proposed as a trade-off alternative between the simple RBF networks and the more complex neural networks involving nonlinear optimization [5]. Besides, this networks hold a simple structure while are capable to identify and modeling complex nonlinear systems. The MSRBF network information flow in our methodology is shown in Figure 4.

The MSRBF network implemented in this work presents the following structure [5],[20]:

$$y(t) = \hat{f}\big(x(t)\big) = \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{m=1}^{n_c}\theta_{i,j,m}\varphi_{i,j,m}(\boldsymbol{x}(t): \boldsymbol{\sigma}_m^{(i,J)}, \boldsymbol{c}_m) \quad (2)$$

Having the basis functions $\varphi_{i,j,m}(x(t): \sigma_m^{(i,J)}, c_m$ defined as

$$\varphi_{i,j,m}(\boldsymbol{x}(t): \boldsymbol{\sigma}_m^{(i,J)}, \boldsymbol{c}_m) = exp\left[-\sum_{k=1}^{d}\left(\frac{x_k(t)-c_{m,k}}{\sigma_{m,k}^{(i,J)}}\right)^2\right] \quad (3)$$

where f is the unknown function, $\varphi_{i,j,m}$ is the general Gaussian kernel, $\theta_{i,j,m}$ are the model parameters, $\boldsymbol{\sigma}_m^{(i,J)}$ the Gaussian scales and $\boldsymbol{c}_m$ the Gaussian centres. Special attention must be put in the Gaussian parameter determination. In our method we implemented an adaptive method to determine the number of centres and the k-means++ algorithm to estimate each. For the scales we chose $\boldsymbol{\sigma}_m^{(1,1)}$, so that we had 16 scales by Gaussian kernel. After the kernel function calculation, a matrix of functions was constructed in order to allow the FROLS algorithm to select the model structure.

C. Model structure selection

The forward orthogonal least squares regression (FROLS) algorithm [14] is designed to build, term by term, the best and most concise models from a pool of candidate terms. This process is equivalent to the neural network training, since model centres and scales are determined through the minimization of the error with respect to the output vector $y(t)$, which is a known pattern. The FROLS algorithm is based on the original OLS estimator [15]. It greedily looks for the term that best minimizes the error of the explanatory variables with respect to the model output $y(t)$ by taking as reference the error reduction ratio (ERR) estimator. However, FROLS adds to OLS the ability to add exclusively terms that provide information than has not been previously included in the model.
In order to reduce the computational cost and warranty good and concise models, we defined the number of model terms to be included based on two conditions: to reach a minimum global accuracy $(1 - \sum ERR)$ of 0.9985, or otherwise add another term up to a maximum of two.

D. DCT & Feature extraction

The main contribution of the discrete cosine transform (DCT) is its data compression capability [16]. To explain the DCT straightforwardly we can imagine a vector $\boldsymbol{v}$ of certain length and the DCT as matrix transformation $\boldsymbol{C}$, such that by multiplying them we will obtain a second vector $\boldsymbol{u}$ of smaller length than less than $\boldsymbol{v}$, such tat:

$$\boldsymbol{vC} = \boldsymbol{u} \quad (4)$$

Another way of looking at it is to think that $\boldsymbol{C}$ breaks down $\boldsymbol{v}$ into a weighted sum of basis cosine sequences[17]. The idea behind using the DCT is to apply it to the MSRBF model output signals to obtain another vector from which only a few numbers will serve as feature values (Figure 5). A graphic example of the DCT compression is illustrated in Figure 6.

E. Classification and detection

The final step of the tests lay upon the classification algorithm. For this purpose the algorithm k-means++ was

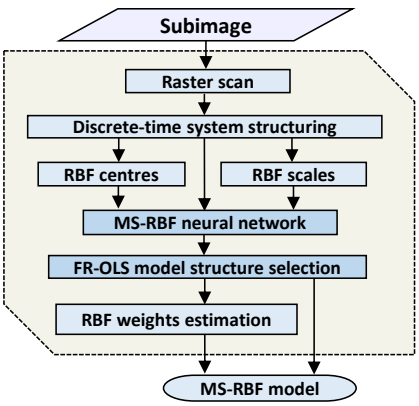


Figure 4. MSRBF-DCT flowchart

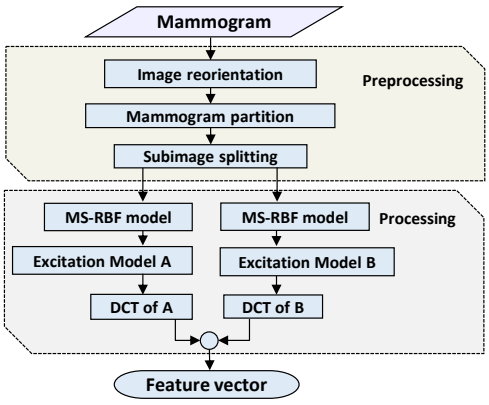


Figure 5. DCT feature extraction flowchart

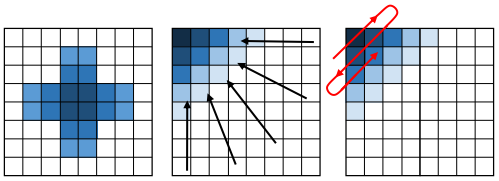


Figure 6. DCT energy compression illustration. (a) Input signal, (b) signal compression, (c) data collection priority.

selected. This technique is inspired in the classical algorithm with the advantage of using an improved seeding method to choose centres, producing a classification up to 70% faster [18].

III. Experiments and Results

To assess the MSRBF DCT method, the experiments engaged various steps. The chosen public repository was the mini-MIAS database of mammograms [19]. It includes 322 grayscale X-ray films of 1024 pixels x 1024 pixels of the medio-lateral oblique view in PGM format. The evaluation aim is to assess the characterization quality of the feature extraction method with supervised learning by evaluating its classification quality for a defined set of images. To reduce the chance of attaining biased performance values, we made a randomized data-splitting of the 322 breast scans following a 65% to 35% ratio for training and testing Furthermore, to counteract the high image variability, we carried out $n = 4$ different training and test scenarios aimed at averaging the final performance measures. For instance, the global accuracy for n scenarios of training and testing is defined by:

$$Accuracy_n = \frac{1}{n}\sum_{i=1}^{n} Accuracy_i \qquad (5)$$

To train the classifier, a matrix of 21,637 feature vectors was assembled, of which 95.5% belonged to *healthy* and 4.5% were identified as abnormal, being 2.29% *benign* and 2.21% *malign*.

The average time to train the MSRBF network of a subimage was 5.8 seconds. Instead, the time required to build the classifier training matrix was of several weeks due to the need for a careful selection of the medical samples and the difficulty implied by the similarity between glandular and dense healthy samples with the suspicious ones. After that, to put together a training matrix for a specific training-testing partition it was only necessary to generate a subset of the complete training matrix by removing from it the mammogram-related vectors selected for testing. All programs were coded in MATLAB R2014b 64-bit and executed in a computer running the Windows 7 Professional operating system with Intel (R) Core (TM) i5-4590 processor at 3.30GHz speed processing and 4 GB of RAM.

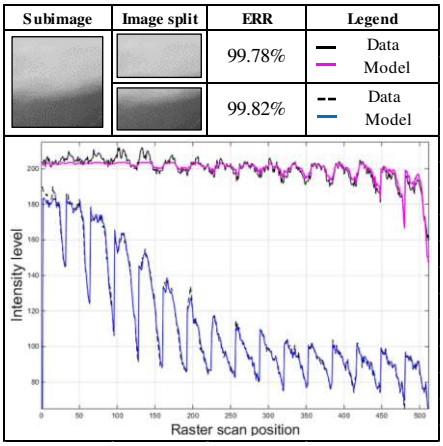


Figure 7. Two pairs of fit-to-data curves and ERR values.

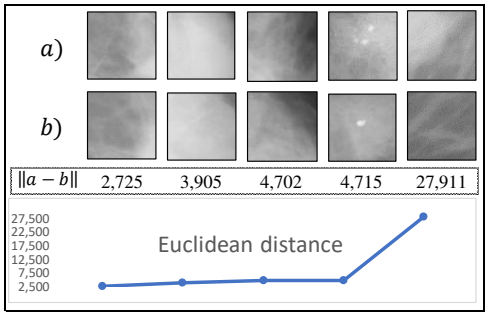


Figure 8. Subimage pairs examples (coupled vertically) and MSRBF DCT feature vectors pair-distance. The more the visual dissimilarity, the larger the distance.

The initial evaluation is aimed at judging the ability of the model to fit the observational data. Figure 7 shows the example of a dense tissue-type subimage, its subdivisions (two-fold characterization) and the ERR of models with respect to the data of each case.

Also a plot overlying the fit of both models versus the original data is shown. It can be observed that the model adjustments are indeed reliable since the curves of prediction and data in the two pairs of curves are nearly overlapped.

After this we tracked if the characterization method was capable to provide coherence in terms of distance. We found this relationship favourable, as we exemplify in Figure 8. In it, five pairs of images (a) and (b) holding different concordance degree are represented.

Table I. MS-RBF DCT performance results and breast tissue-type ratio by test.

		Test 1	Test 2	Test 3	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
Statistical 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
	Lesion distinctinon %	81.97	80.88	74.55	76.00

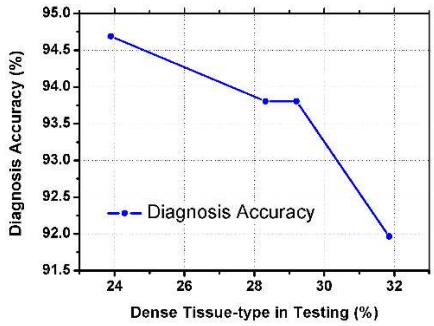
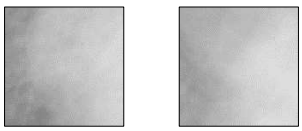


Figure 9. Accuracy as function of the presence of dense mammograms in the test.

Below each pair, the Euclidean distance between the elements is presented as well as a plot showing how the distance increases positively as the images display greater disparity. The experimental performance results of the four tests from different partitions of the database are described in Table I. It also displays percentages by mammogram-type included in each test. The overall results are very positive, especially in terms of accuracy, specificity and NPV. As assumed, it may be noted that the composition of the elements in the test affects in some proportion the classification result. This is an interesting point to take into account because it could lead some classification studies to confusing results.

To ease the analysis of the resulting variation of tests in dependence of the mammogram-type, we plotted interesting trends. Figure 9 shows a negative relationship between the appearance of dense mammograms and the classification accuracy. This may be due to dense healthy images resemble some tumours also having high density.



a) b)

Figure 10. False malign detection occurred in a healthy dense subimage (b) attracted by a malign sample (a).

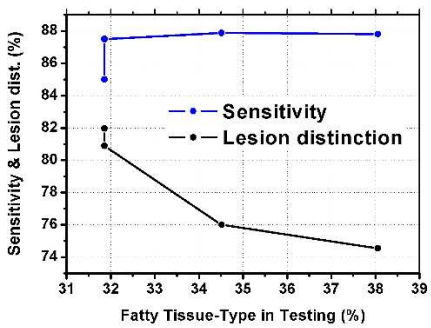


Figure 11. Sensitivity and lesion distinction accuracy as functions of the presence of fatty mammograms in the test.

Table II. MS-RBF DCT average performance results.

Statistical Measure	Avg. result %
Accuracy	93.57
Sensitivity	87.05
Specificity	96.97
PPV	93.79
NPV	93.50
Lesion distinctinon	78.35

Table III. Comparison of MSRBF-DCT with previous work.

Model	Image Set	Acc. %	Sens. %	Spec. %
2D-NARX [7]	mini-MIAS	91	93	89.5
ELM [8]	mini-MIAS	91	90	98
GLCM [9]	mini-MIAS	93.9	97.2	91.5
ICA-RBF [10]	mini-MIAS	88.2	--	--
LDA-ANN [11]	mini-MIAS	93.1	99	83
GPZM [12]	mini-MIAS	89.3	83.5	93.4
MSRBF-DCT	mini-MIAS	93.5	87	96.9

As an example, Figure 10 shows the case of a false positive detected in a healthy dense subimage. It can be seen that malignant tumour tissue (a) and dense healthy tissue (b) can have similar composition and intensity levels.

Figure 11 suggets that there is a lessen ability to distinguish the class of abnormality in the presence of fatty mammograms in the test, which is opposite to what was expected. However, a positive trend between fatty tests sets and the effective detection of any kind of abnormalities was as well suggested.

The overall rating numbers of this study are presented in Table II. It is noticeable that sensitivity, PPV and lesion distinction values are not as high as expected, possibly because the high similitude between dense and glandular healthy tissue with certain abnormality variations.

Finally, a comparison of our method and previous work is presented in Table III. In general our method is competitive but did not reach such a high sensitivity as some most approaches. We believe this imperfection could be yielded by our training strategy, since to avoid a high occurrence of false positives caused by the resemblance between healthy dense tissue and some types of tumour, it was necessary to increase the amount of dense samples in the training database.

IV. Conclusion and Future Work

This paper presents an advantageous modelling neural network framework designed originally to model nonlinear observational input-output series as a novel image feature extraction method and CAD system. Furthermore, we incorporated the Discrete Cosine Transform algorithm to make the most of the MSRBF image models.

Experiments to classify tumours in X-ray mammograms showed that the method is capable to contend with well-known previous CAD systems based on system identification. The proposed method reached a classification accuracy above 93%. While we recognise that the MSRBF DCT method is not perfect, we consider that some below-average metrics may be due to a fault in the training strategy and an important similarity of dense tumours and healthy dense tissue.

As regards the comparison with previous work no reference about the tissue-type composition included in the test set was found, a factor we think can produce changes in the global performance. The comparison exercise of the model performance taking into account different training-testing compositions lead us to infer that getting results with a single partition in a heterogeneous database 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 balance in the training matrix composition to get to an optimal balance between sensitivity and specificity. In addition, the transfer of our methodology to other medical study areas such as brain diseases and lung cancer detection is desirable.

ACKNOWLEDGEMENTS

C.B.P. thanks the funding of this work by the National Council for Science and Technology (CONACYT). The authors gratefully acknowledge that part of the work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and Platform Grant EP/H00453X/1.

REFERENCES

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

[2] S. Billings, *Nonlinear System Identification NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*, vol. 21, no. I. 2013.

[3] K. Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,” *Comput. Med. Imaging Graph.*, vol. 31, no. 4–5, pp. 198–211, 2007.

[4] G. Cybenko, “Approximations by superpositions of sigmoidal functions,” *Approx. Theory its Appl.*, vol. 9, no. 3, pp. 17–28, 1989.

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

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

[7] C. Beltran Perez and H. L. Wei, “Digital image classification and detection using a 2D-NARX model,” in *ICAC 2017 - 2017 23rd IEEE International Conference on Automation and Computing: Addressing Global Challenges through Automation and Computing*, 2017.

[8] G. Vani, R. Savitha, and N. Sundararajan, “Classification of abnormalities in digitized mammograms using Extreme Learning Machine,” *Control Autom. Robot. Vis. (ICARCV), 2010 11th Int. Conf.*, no. December, pp. 2114–2117, 2010.

[9] M. Pratiwi, Alexander, J. Harefa, and S. Nanda, “Mammograms Classification Using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network,” in *Procedia Computer Science*, 2015, vol. 59, pp. 83–91.

[10] I. Christoyianni, a Koutras, E. Dermatas, and G. Kokkinakis, “Computer aided diagnosis of breast cancer in digitized mammograms.,” *Computerized medical imaging and graphics the official journal of the Computerized Medical Imaging Society*, vol. 26, no. 5. pp. 309–319, 2002.

[11] K. Iftikhar, S. Anwar, I. U. Haq, M. T. Khan, and S. R. Akbar, “an Optimal Neural Network Based Classification Technique,” *J. Eng. Appl. Sci.*, vol. 35, no. 1, pp. 39–46, 2016.

[12] S. P. Singh and S. Urooj, “An Improved CAD System for Breast Cancer Diagnosis Based on Generalized Pseudo-Zernike Moment and Ada-DEWNN Classifier,” *J. Med. Syst.*, vol. 40, no. 4, pp. 1–13, 2016.

[13] I. J. Leontaritis and S. A. Billings, “Input-output parametric models for non-linear systems Part I: deterministic non-linear systems,” *Int. J. Control*, vol. 41, no. 2, pp. 303–328, 1985.

[14] S. A. Billings, M. J. Korenberg, and S. Chen, “Identification of non-linear output-affine systems using an orthogonal least-squares algorithm,” *Int. J. Syst. Sci.*, vol. 19, no. 8, pp. 1559–1568, 1988.

[15] M. Korenberg, S. A. Billings, Y. P. Liu, and P. J. McIlroy, “Orthogonal parameter estimation algorithm for non-linear stochastic systems,” *Int. J. Control*, vol. 48, no. 1, pp. 193–210, 1988.

[16] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete Cosine Transform,” *Comput. IEEE Trans.*, vol. C-23, no. 1, pp. 90–93, 1974.

[17] Z. M. Hafed and M. D. Levine, “Face recognition using the discrete cosine transform,” *Int. J. Comput. Vis.*, vol. 43, no. 3, pp. 167–188, 2001.

[18] D. Arthur and S. Vassilvitskii, “k-means++: the advantages of careful seeding,” in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1025.

[19] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, “The Mammographic Image Analysis Society Digital Mammogram Database,” *Expert. Medica, Int. Congr. Ser.*, vol. 1069, no. January 1994, pp. 375–378, 1994.

[20] H.- L. Wei., D. Q. Zhu, S. A. Billings, M. A. Balikhin, "Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks," Advances in Space Research, vol. 40, no. 12, pp.1863-1870, 2007.