Expert Systems With Applications 112 (2018) 388–400

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Three-Dimensional Local Energy-Based Shape Histogram (3D-LESH): A Novel Feature Extraction Technique

Summrina Kanwal Wajid [a,b], Amir Hussain [a,*], Kaizhu Huang [c,*]

[a] Division of Computing Science & Mathematics, School of Natural Sciences, University of Stirling, Stirling Scotland, UK
[b] International Riphah University, Islamabad, Pakistan
[c] Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

ARTICLE INFO

Article history:
Received 22 March 2017
Revised 20 October 2017
Accepted 30 November 2017
Available online 5 December 2017

Keywords:
Clinical decision support system (CDSS)
Echo state network (ESN)
Extreme learning machine (ELM)
Local energy-based shape histogram (LESH)
Magnetic resonance imaging (MRI)
Support vector machine (SVM)

ABSTRACT

In this paper, we present a novel feature extraction technique, termed Three-Dimensional Local Energy-Based Shape Histogram (3D-LESH), and exploit it to detect breast cancer in volumetric medical images. The technique is incorporated as part of an intelligent expert system that can aid medical practitioners making diagnostic decisions. Analysis of volumetric images, slice by slice, is cumbersome and inefficient. Hence, 3D-LESH is designed to compute a histogram-based feature set from a local energy map, calculated using a phase congruency (PC) measure of volumetric Magnetic Resonance Imaging (MRI) scans in 3D space. 3D-LESH features are invariant to contrast intensity variations within different slices of the MRI scan and are thus suitable for medical image analysis.

The contribution of this article is manifold. First, we formulate a novel 3D-LESH feature extraction technique for 3D medical images to analyse volumetric images. Further, the proposed 3D-LESH algorithmis, for the first time, applied to medical MRI images. The final contribution is the design of an intelligent clinical decision support system (CDSS) as a multi-stage approach, combining novel 3D-LESH feature extraction with machine learning classifiers, to detect cancer from breast MRI scans. The proposed system applies contrast-limited adaptive histogram equalisation (CLAHE) to the MRI images before extracting 3D-LESH features. Furthermore, a selected subset of these features is fed into a machine-learning classifier, namely, a support vector machine (SVM), an extreme learning machine (ELM) or an echo state network (ESN) classifier, to detect abnormalities and distinguish between different stages of abnormality. We demonstrate the performance of the proposed technique by its application to benchmark breast cancer MRI images. The results indicate high-performance accuracy of the proposed system (98%±0.0050, with an area under a receiver operating charactertistic curve value of 0.9900 ± 0.0050) with multiple classifiers. When compared with the state-of-the-art wavelet-based feature extraction technique, statistical analysis provides conclusive evidence of the significance of our proposed 3D-LESH algorithm.

1. Introduction

Extraction of distinctive features is a vital task in medical image analysis as it assists in the detection and diagnosis of chronic diseases. Just as human beings are capable of deriving a story from a picture using background information, machines can extract semantic knowledge from an image and build a computational model, by retrieving significant information that uniquely and precisely defines the object of interest (Kumar & Bhatia, 2014).

The success of accurate medical diagnosis from MRI images depends on reliable and accurate methods for extracting important patterns from a medical image sequence (or slices). A variety of methods have been proposed in the literature for extracting significant features for medical diagnosis, some of these are discussed in the *related works section*. A common problem with these methods is that they lack the capability to handle an image sequence, since a long sequence generates variations in the illumination condition.

Our proposed 3D-LESH feature extraction technique is based on the PC model. PC is a dimensionless quantity which represents fine detailed image patterns using a local energy map, depicting corners and edges. The resulting features are localized as they are perceived at points where the Fourier components are maximal and thus, are suitable for an extended image sequence. The PC model has a psychophysical impact on human feature perception (Kovesi, 2003). The next section outlines the choice of breast cancer detection using MRI images as our target case study, followed

* Co-corresponding authors.
E-mail addresses: skw1@cs.stir.ac.uk (S.K. Wajid), ahu@cs.stir.ac.uk (A. Hussain), kaizhu.huang@xjtlu.edu.cn (K. Huang).

https://doi.org/10.1016/j.eswa.2017.11.057

by review of related work. This is followed by a detailed description of the 3D-LESH feature extraction technique and CDSS system, and comparative simulation results.

1.1. Background

Despite advancements in medical science, breast cancer in women remains the second most fatal cause of death, accounting for almost 29% of recently diagnosed cancer cases in the US. It is estimated that approximately 40,290 women died from this disease in 2015, with rates increasing every year (American Cancer Society, 2016; Siegel, Miller, & Jemal, 2016; Shulman, Willett, Sievers, & Knaul, 2010). Individuals at an advanced cancer stage, with large tumour sizes at the time of diagnosis, have much lower levels of survival, indicating the urgency for early detection (American Cancer Society, 2016). Timely and effective treatment plans can be devised by Doctors through use of CDSS, coupled with advanced medical imaging techniques.

Mammography has been considered the mainstay for cancer diagnosis for decades, however it's low sensitivity rate in detecting lesions from dense breast tissue results in unnecessary biopsies, causing emotional and economic stress for patients. Furthermore, the requirement for breast compression can also cause discomfort for patients (Aminololama-Shakeri & Khatri, 2014)

To overcome the aforementioned limitations of mammography, new breast imaging modalities have emerged, such as ultrasonography, computed tomography (CT) and MRI. Over the past decade, MRI has evolved into a well-established breast imaging modality because of its low ionising radiation and high sensitivity in detecting in-situ and invasive breast cancer (Palestrant, Comstock, & Moy, 2014). It has also been found to be efficient for detecting cancer staging, and for prognosis and treatment monitoring (Palestrant et al., 2014).

Multiple research trials have shown that breast MRIs provide superior sensitivity for breast cancer detection when compared with mammographic and sonographic imaging (Menezes, Knuttel, Stehouwer, Pijnappel, & van den Bosch, 2014). MRI is not only resilient in detecting early stages of breast cancer but also effective in evaluating the size of the tumour, lymph node involvement, metastases and response to treatment (chemo- and radiation-therapy) (Aminololama-Shakeri & Khatri, 2014; Ojeda-Fournier, de Guzman, & Hylton, 2013). In the UK, MARIBS (Magnetic Resonance Imaging in Breast Screening) was conducted to evaluate the capacity of MRI for screening breast cancer in premenopausal women. It was found that MRI is more sensitive (77%) compared to mammograms (40%), as it can detect lesions that are left undetected using mammograms.

1.2. Related work

The challenge of distinguishing between malignant and benign lesions using MRI images has been addressed by many researchers (Seely, 2012). Keyvanfard et al. extracted the shape, texture and kinetic features from a segmented region of interest (ROI) containing the lesions, and experimented with different classifiers, namely, a multi-layer perceptron (MLP), probabilistic neural network (PNN), generalized regression neural network (GRNN), support vector machine (SVM) and radial basis function (RBF) networks, to distinguish between malignant and benign MRI lesions. They combined best results to develop a multi-classifier system that achieved an overall accuracy of 90% (Keyvanfard, Shoorehdeli, Teshnehlab, Nie, & Su, 2013). Levman et al. computed temporal features from ROIs extracted from dynamic contrast-enhanced magnetic resonance images (DCE-MRI), and applied SVM networks to distinguish between malignant and benign cases (Levman, Leung, Causer, Plewes, & Martel, 2008). Arbach, Stolpen and Reinhardt applied a back propagation neural network (BNN) to classify benign and malignant cases extracted from breast MRI images, and achieved an area-under a-receiver operating characteristic curve (AUROC) value of 0.913 (Arbach, Stolpen, & Reinhardt, 2004).

Aghaei et al. investigated the response of a tumour and its surroundings to chemotherapy in MRI images. To accomplish this task, they first extracted 39 kinetic features and analyzed these through feature fusion, by combining classification results generated from the application of multiple feature sets. Next, they combined an artificial neural network (ANN) and a wrapper subset evaluator, using a leave-one-case-out validation method for classification purposes and reported an AUROC value of 0.96 ± 0.03 (Aghaei et al., 2015). Yang et al. (2015) devised a technique for discriminating malignant and benign tumour cases from segmented MRI images, by extracting kinetic and background parenchymal enhancement (BPE) features and applying SVM using a leave-one-case-out strategy. They achieved a performance AUROC value of 0.919 ± 0.029, by combining multiple useful features. Huang extracted a morphological, grey-level co-occurrence matrix (GLCM) and ellipsoid fitting features from segmented lesions and applied a bi-serial correlation coefficient to select a resilient set of features, reporting a performance accuracy of 88.42% (Huang et al., 2013). Chen et al. extracted four features – time-to-peak, maximum contrast enhancement, washout rate of the lesion, and uptake rate – from the characteristic kinetic curve of breast lesions using DCE-MRI images, and then applied fuzzy c-means clustering to discriminate between benign and malignant cases (Chen, Giger, Bick, & Newstead, 2006). Szabo et al. extracted kinetic, morphological and combined MR features from MRI images of lesions. Application of an ANN-based classification model resulted in a performance as competent as that of an expert radiologist (Szab, Wiberg, Bon, & Aspelin, 2004). Meinel et al. presented a scheme to classify segmented breast MRI lesions on a scale from benign to malignant, with the help of 42 shape, texture and enhancement kinetic features. The best of these features were then selected and fed to a BNN. The performance was compared to that of a human observer and found to be significantly enhanced (Meinel, Stolpen, Berbaum, Fajardo, & Reinhardt, 2007). Honda et al. applied quadratic discriminant analysis (QDA) on features extracted from DCE-MRI breast image masses, namely dynamic changes in signal intensity, shape irregularity, and margin smoothness. They achieved a classification accuracy, sensitivity and specificity of 85.6%, 87.1%, and 82.1%, respectively (Honda, Nakayama, Koyama, & Yamashita, 2016). Lee et al. experimented with spatiotemporal association features extracted from time-series contrast-enhanced MRI images, with the tumours segmented using fuzzy c-means clustering. A least-squares support vector machine (LS-SVM) classifier was employed to differentiate between malignant and benign cases (Lee et al., 2010). Yang et al. extracted global kinetic features from DCE-MRI to classify malignant and benign tumours (Yang et al., 2016).

Hassanien and Kim proposed a system to distinguish between abnormal and normal breast MRI cases (Hassanien et al., 2012). They first enhanced the MRI using an adapted fuzzy type-II algorithm. Next, they segmented the ROI using a pulse-coupled neural network (PCNN), and extracted wavelet features from the segmented ROI. Finally, they applied SVM for classification, resulting in a classification error of 5.1% (Hassanien et al., 2012). Hassanien et al. extended this research to diagnose breast cancer from MRI images by applying ant-based clustering for ROI segmentation (Hassanien, Moftah, Azar, & Shoman, 2014). Subsequently, they extracted statistical features from the ROI and used a MLP neural network to perform classification, reporting a mean absolute error (MAE) of 0.0339 (Hassanien et al., 2014). Agliozzo et al. developed a computer-based system to classify malignant and benign masses in MRI images. They experimented with different feature subsets (morphological, kinetic and spatiotemporal features) in conjunction

with support vector regression, and reached a maximum area under the AUC curve value of 0.96 ± 0.02. A genetic algorithm was employed to select the best possible combination of these features (Agliozzo et al., 2012).

Anger et al. developed a system to distinguish between malignant and benign lesions in dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). Their technique applied an expectation-maximization-driven active contours scheme for automatic segmentation of breast lesions. Morphological, textural and kinetic features were extracted and classification performed using a SVM network. The system reached an accuracy of 83%, sensitivity of 79% and specificity of 88% (Agner et al., 2009). El Nawasany, Ali and Waheed proposed an algorithm for tumour detection from MRI images which applied a perceptron algorithm in conjunction with the scale-invariant feature transform (SIFT) feature extraction technique and achieved an accuracy of 86.76% (ElNawasany, Ali, & Waheed, 2014).

The transformation of images into feature sets aids in the classification and recognition of ROI in medical images [92]. Intensive research has been undertaken to develop new techniques and methods that can enable extraction of resilient features from medical images and, in turn, enhance the CDSS diagnosis system. Some of these state-of-the-art methods and approaches have been discussed above, where researchers have mainly experimented with diverse types of features to assess their impact on CDSS diagnosis performance. The relative merits and demerits of these tecniques are summarized in Table 1.

It is interesting to note that, all methods explored to-date have a rather large margin for performance enhancement, hence researchers have reported comparative results by assembling different subsets of features to enhance the diagnostic capabilities of the overall CDSS system.

In contrast, our proposed method, the 3D-LESH feature extraction technique, developed as part of a multi-stage framework based on MRI images, offers a significant increase in performance when compared to state-of-the-art techniques. Our main contribution in this paper is the development of 3D-LESH features and their application in conjunction with machine learning classifiers, for diagnosing cancer malignancies and stages of cancer using MRI images. The conventional LESH (for two-dimensional images) has been previously applied in diverse pattern recognition applications. In particular, Sarfraz and Hellwich initially developed this technique for facial recognition systems, with different face and head poses (Sarfraz & Hellwich, 2008a, b, 2009). Zakir, Zafar and Edirisinghe applied LESH to detect and recognize different road signs automatically (Zakir, Zafar, & Edirisinghe, 2011). Wajid and Hussain applied the LESH feature extraction technique in conjunction with SVM, both to detect malignancy and distinguish between various types of malignancy in mammogram images (Wajid & Hussain, 2015). This technique was further extended through application of an ESN classifier to mammogram images (Wajid, Hussain, & Luo, 2014). More recently, the LESH feature extraction technique, combined with SVM, ESN and ELM classifiers, was employed on chest radiographs to distinguish lungs with and without nodules, and to make a distinction between normal and abnormal lung nodules. All these experiments demonstrated the high distinguishing capability of the proposed system (Wajid, Hussain, Huang, & Boulila, 2016).

The technique proposed in this paper provides a novel contribution in the form of a new 3D-LESH algorithm, which computes features for MRI datasets in 3D, by convolving the image with a 3D gabor filter, and then calculates the 3D phase congruency. This technique is applied, for the first time, on 3D MRI images to detect malignancies. Further applications in CT scan

Table 1
An illustrative review of different feature extraction techniques for breast cancer detection from MRI images.

Feature type	Applied in	Strength	Weaknesses
Shape	(Keyvanfard et al., 2013).	Concise, robust and easy to compute	The initial locus of the beginning point may affect the correct shape feature approximation (M. Mahrooghy et al., 2013).
	(Meinel et al., 2007). (Honda et al., 2015)		
Kinetic	(Levman et al., 2008)	Effective to determine the speculated margin of malignant tumours	Not efficient in detecting intra-tumour heterogeneity, thus lacking the capability of tumour characterization (M. Mahrooghy et al., 2013).
	(Agliozzo et al., 2012), Agner et al., 2009), (Chen et al., 2006)		
Morphological	(Agliozzo et al., 2012), (Agner et al., 2009)	It helps to remove noise and create an outline of the objects in the image	Hard to generalize in a situation where micro-classification as well as speculated masses are involved
Texture Features	(Aghaei, Tan, Hollingsworth, Qian, Liu, and Zheng, 2015; Huang et al., 2013)	Simple and easy to compute	Not effective in discriminating different stages as well as types of malignancies.
Spatiotemporal association features	(Agliozzo et al., 2012; Lee et al., 2010).	Effective in predicting tumour malignancy	They are not based on heterogeneity partitioning and do not respond to enhancement in spatial frequency
Wavelet	(Hassanien et al., 2012).	Wavelet features also turned out to be excellent in discriminating among different types of abnormalities in mammograms. They are well localized in the time and frequency domain.	Wavelets suffer from a loss of generality.
			Wavelets have efficient image representation but fail to represent discontinuities along curves and edges.
Scale invariant feature transform (SIFT)	(ElNawasany et al., 2014).	Rotation and scale invariant	Mathematically complicated and computationally heavy.
3D-LESH	Proposed algorithm	Robust in discriminating malignant cases from benign ones as well as depicting the cancer stages.	They are not scale and rotation invariant.
		LESH is a histogram of local energy, which is at a maximum level at the abrupt change of image intensity. Accordingly, it marks the significant texture variations in the local area.	Further extensive assessment and clinical validation is essential, using additional clinical data-sets, to validate the clinical significance of the proposed algorithm.

images are also envisaged, to demonstrate the technique's wide applicability.

The rest of this paper is organized as follows: Section 2 introduces the proposed 3D-LESH algorithme; Section 3 presents the experimental work performed; Section 4 summarises the results and discussion; and finally Section 5 outlines conclusions and some future work directions.

2. Three-Dimensional local energy-based shape histogram (3D-LESH) feature extraction

A novel 3D-LESH technique is developed for extracting informative patterns from volumetric medical images, to assist in medical diagnosis. It builds on the local energy model of Morrone and Owens (1987), which considers image features at the maximum phase of the Fourier components (Kovesi, 2003), corresponding to the peak of the local energy. We compute these features using a dimensionless measure, the PC, in three-dimensional space. The 3D-LESH features are normalized within the range of 1 (highly significant) to 0 (not significant features).

The technique investigates valuable image statistics by computing a histogram of the local energy pattern within the image (Wajid et al., 2014). 3D-LESH features are calculated using the following steps.

2.1. Convolution with 3D log-Gabor

The original approach to compute the local energy of the signal involves computation of a Hillbert transform (using Fourier transform) of the signal, before convolving it with a pair of filters in quadrature, to remove the DC component, resulting in a band pass version of the original signal. The conventional approach has been extended to compute the PC of two-dimensional images. In this approach the image is convolved with filters in quadrature with the first Hillbert transform, producing a 90-degree phase shift of the other image. These two resultant outputs are squared and summed to produce a measure of the local energy for the overall image (Kovesi, 1999).

However, this method has the drawback that the Fourier transform is not considered an appropriate measure for the computation of local energy, and is not capable of considering the spread of frequencies congruent at a particular point. Considering these defects, we adopt the log-Gabor filter-based convolution (wavelet transform) for obtaining local frequency information from the 3D MRI images, after computing the Fourier transform. Application of wavelet filters for obtaining local frequency information was originally proposed as a Morlet wavelet in (Kovesi, 1999). In this approach, the local energy is computed by first convolving the Fourier transform of the volumetric image I with the bank of three-dimensional wavelet-based log-Gabor filters, in quadrature with different orientations n and scales c, hence calculating the phase and amplitude of the image.

The 3D log-Gabor filter is computed by multiplying the one-dimensional log-Gabor function (in the radial frequency) and the Gaussian (in the angular distance with rotational symmetry in spherical coordinates), given as follows (Ferrari et al., 2011):

$$G(\omega,\ \Theta) = \exp\left\{-\frac{\left(log\left(\frac{\omega}{\omega_i}\right)\right)^2}{2\left(log\left(\frac{\sigma_\omega}{\omega_i}\right)\right)^2}\right\} \times exp\left(-\frac{\alpha(\Theta)^2}{2\ \sigma_\alpha^2}\right) \quad (1)$$

where $\Theta=(\theta,\ \phi)$ represent the filter orientation for the sphere of unit radius, θ is the elevation angle and ϕ is the azimuth angle. ω_i corresponds to the central radial frequency of the ith filter and σ_ω and σ_α are standard deviations controlling the filter bandwidth (Dosil, Pardo, & Fdez-Vidal, 2005; Ferrari et al., 2011). The ratio $(\frac{\sigma_\omega}{\omega_i})$is kept constant to keep the shape ratio filters persistent, hence given as $(\frac{\sigma_\omega}{\omega_i}) = -\frac{1}{4}\sqrt{log2}\beta$ within the range 0- 1, and β is in octave (Ferrari et al., 2011).

$\alpha(\Theta)=\arccos(.\frac{\omega.\ v}{v}), v=(\cos\phi.\ \cos\theta,\ \cos\phi.\ \sin\phi,\ \sin\theta)$ and ω is a point in the frequency space in Cartesians (Dosil et al., 2005; Ferrari et al., 2011).

In this case, DC components are removed from the image while retaining other frequency components intact. The bank of wavelet filters is designed so that each filter overlaps its neighbours and the entirety of all the transfer functions provides even coverage of the spectrum, when reconstructing the original image over a frequency band, with a certain scale factor (Kovesi, 2000). Here, the frequency component is constructed by convolving the image with even and odd filter components and the scale of the filter controls the frequency range for calculating PC (Kovesi,1999). If we let G_{cn}^{even} and G_{cn}^{odd} be the even-symmetric and odd-symmetric wavelet filters at scale c and orientation n, the convolution results in a response vector given as:

$$[e_n(s,\Theta), o_n(s,\Theta)] = [\mathrm{I(s)} * \mathrm{G}(\omega,\Theta)_{\mathrm{cn}}^{\mathrm{even}}, \mathrm{I(s)} * \mathrm{G}(\omega,\Theta)_{\mathrm{cn}}^{\mathrm{odd}}] \quad (2)$$

where s indicates a location (x,y,z) in the image I. Hence, the amplitude of the response at a given scale and orientation is computed as:

$$A_n(s,\ \Theta) = \sqrt{(\boldsymbol{e_n}(s,\ \Theta))^2 + (\boldsymbol{o_n}(s,\ \Theta))^2} \quad (3)$$

And the sensitive phase deviation measure (Kovesi,1999) is given as:

$$\Delta\Phi_n(s,\ \Theta) = \cos\left(\phi_n(s,\ \Theta) - \bar{\phi}_n(s,\ \Theta)\right) - \left|\sin\left(\phi_n(s,\ \Theta) - \bar{\phi}_n(s,\ \Theta)\right)\right| \quad (4)$$

$$= \boldsymbol{e_n}(s,\ \Theta)\bar{\phi}_e(s,\ \Theta) + o_n(s,\ \Theta)\bar{\phi}_o(s,\ \Theta) - \left|\boldsymbol{e_n}(s,\ \Theta)\bar{\phi}_o(s,\ \Theta) - o_n(s,\ \Theta)\bar{\phi}_e(s,\ \Theta)\right| \quad (5)$$

$$where\ \bar{\phi}_e(s,\ \Theta) = \sum\nolimits_n \boldsymbol{e_n}(s,\ \Theta)/E(s,\ \Theta), \quad (6)$$

$$\bar{\phi}_o(s,\ \Theta) = \sum_n \boldsymbol{o_n}(s,\ \Theta)/E(s,\ \Theta) \quad (7)$$

Finally, the local energy is computed as:

$$E(s,\ \Theta) = \sqrt{\left(\sum_n \boldsymbol{e_n}(s,\ \Theta)\right)^2 + \left(\sum_n \boldsymbol{o_n}(s,\ \Theta)\right)^2} \quad (8)$$

2.2. Computation of 3D phase congruency

PC is only useful if it occurs on a wide range of frequency; thus, it is computed as the local energy normalized by the sum of Fourier amplitude components as:

$$PC_{3D}(s,\ \Theta) = \frac{E(s,\ \Theta)}{\sum_n A_n(s,\ \Theta) + \epsilon} \quad (9)$$

$$PC_{3D}(s,\ \Theta) = \frac{\sum_n \mathrm{W}(s,\ \Theta)\lfloor A_n(s,\ \Theta)\Delta\Phi_n(s,\ \Theta) - T\rfloor}{\sum_n A_n(s,\ \Theta) + \epsilon} \quad (10)$$

Where the enclosed measure $\lfloor\bullet\rfloor$ is equal to itself, if positive, and is zero otherwise for the spatial location $s=(x,\ y,\ z)$ in the 3D MRI image, and $\Theta=(\theta,\ \Phi)$ is the filter orientation with unit radius for the sphere. T is the noise level threshold of the image energy map and ε is a small constant for avoiding division by zero (Ferrari et al., 2011).

$W(s,\ \Theta) = \frac{1}{1+e^{\gamma(d-l(s,\Theta))}}$ is the phase congruency weighting function (Kovesi, 1999, 2000), where d represents the lower bound

Table 2
3D-LESH feature extraction algorithm.

Algorithm: Let I be a breast MRI image with a certain pixel location given as $s = (x, y, z)$.

Begin:
1. Convolve the Fourier transform of the image I with a bank of 3D log-Gabor filter $G(\omega, \Theta)$ given in "(1)," with different orientations n and scales c. The resultant convolution response vector is given in "(2)".
2. Calculate the amplitude of the response using "(3)".
3. Calculate the sensitive phase deviation measure using "(5)".
4. Calculate the local energy using "(8)".
5. Calculate the 3D phase congruency for the image using "(10)".
6. Compute the 3D-LESH feature vector using "(11) & (12)".

End

for the filter response spread (threshold). Below this, the phase congruency value is penalized, and γ, the gain factor, controls the sharpness of the filter response spread (Ferrari et al., 2011).

Given that $l(s, \Theta) = \frac{1}{N}\left(\frac{\sum_n A_n(s, \Theta)}{A_{max}(s, \Theta)+\epsilon}\right)$ is the filter response spread measure, calculated by summing the amplitudes (A_n) and dividing by the highest individual response (A_{max}). This produces the width of the distribution (Ferrari et al., 2011).

2.3. Design of 3D-LESH feature extraction

The orientation label map for each voxel is computed, and each voxel is assigned a label of the orientation which exhibits the largest energy across all scales.

Finally, a histogram of the local energy of each sub-region, along each orientation of the 3D image, is generated and combined to preserve the relationships between different regions. The resultant 3D-LESH feature vector is calculated as follows:

$$h_{r,b} = \sum V_r \times PC_{3D}(s, \Theta) \times \delta_{Lb} \tag{11}$$

$$V_r = \frac{1}{\sqrt{2\pi}\sigma} e^{[(x-r_{x0})^2+(y-r_{y0})^2+(z-r_{z0})^2]/\sigma^2} \tag{12}$$

Where V_r is the Gaussian weighting function of region r with standard deviation σ, centered at (r_{x0}, r_{y0}, r_{z0}), in the 3D MRI image, $PC_{3D}(s, \Theta)$ represents the local energy computed by "(1)," and δ_{Lb}, represents Kronecker's delta of the orientation label map L and current bin b. In order to keep the spatial relation intact, a 8 bin local histogram is computed to correspond to the 8 filter orientations for each of the 16 sub-regions of the 3D image, hence producing a 128-dimensional feature vector. Table 2 illustrates the overall algorithm for 3D-LESH feature extraction.

2.4. Computational complexity of 3D-LESH feature extraction technique

Here we analyze the computational complexity of our proposed 3D-LESH algorithm. Assuming a standard volumetric MRI image is of size $D_x \times D_y \times D_z$ where D_z represents the number of slices, and $D_x \times D_y$ is a slice/image in the xy-plane. Assuming $D = D_x = D_y = D_z$ (for the sake of simplicity), where D is always within a limit, the overall algorithm will converge in all cases. In the case of segmented ROI, $D = 64$, where in the case of breast area analysis, detecting a stage of breast cancer D would equate to 128.

First, the Fourier transform of the image is computed. For a one-dimensional signal of size D, the Fourier transform takes $O(D \log D)$ operations. Hence, in the case of an image of size D^3, it will take $O(D^3 \log D)$ (Ferrari et al., 2011).

Next, the transformed image is convolved with a 3D-log Gabor filter in the frequency domain, resulting in D^3 operations. Thus, $O(D^3 \log D + D^3)$ operations are performed per filter bank. As the total number of filters is c (scale) $\times\, \theta$ (the elevation angle) $\times\, \phi$ (the azimuth), leading to $O(\mathrm{D}^3 \log \mathrm{D} + \mathrm{D}^3)c\theta\phi$. Next, the computation of $PC_{3D}(s, \Theta)$ as well as $h_{r,b}$ requires D^3 operations. The total times complexity can thus be written as:

$$T(D, c, \theta, \phi) = O\left(\mathrm{D}^3 \log \mathrm{D} + \mathrm{D}^3\right)c\theta\phi + \mathrm{D}^3 + \mathrm{D}^3 = O\left(\mathrm{D}^3 \log \mathrm{D}\right)c\theta\phi = O\left(\mathrm{D}^3 \log \mathrm{D}\right) \tag{13}$$

Values of scales, elevation and azimuth can be variables, but for our work they are kept fixed and provided in the experimental section.

In order to analyze the space complexity of the algorithm, we can see that, initially, to keep the Fourier transform of the image as well as the result of convolution of the Fourier transform with the Gabor filter, an $O(D3)$ space is required. Similarly, the space required to compute $\mathrm{PC_{3D}(s, \Theta)}$ and $\mathrm{h_{r,b}}$ is $O(D3)$: hence the total space complexity can be stimated as:

$$\mathrm{Sp} = \mathrm{O}\left(\mathrm{D}^3 + \mathrm{D}^3 + \mathrm{D}^3\right) = \mathrm{O}\left(\mathrm{D}^3\right) \tag{14}$$

The next section provides details of various stages of the proposed CDSS framework and their experimental evaluation.

3. Experimental work

The workflow of the proposed CDSS framework for breast cancer diagnosis, based on the 3D-LESH feature extraction technique, is depicted in Fig. 1.

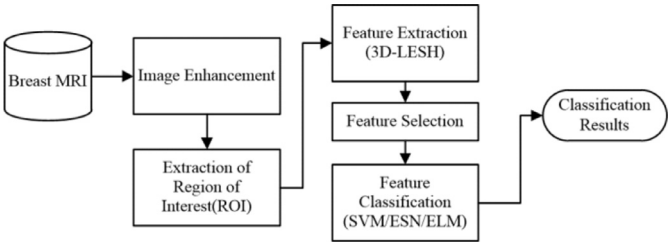

Fig. 1. Workflow of the proposed CDSS framework.

3.1. Breast MRI dataset

For the evaluation of our proposed system, we collected 137 breast cancer MRI scans from the cancer atlas for breast cancer, specifically the TCGA-BRCA data set, provided by The Cancer Imaging Archive (TCIA) of the Frederick National Laboratory for Cancer Research (Clark et al., 2013). TCGA-BRCA is the largest publicly available data set of breast MRIs. Each MRI scan is accompanied by clinical and pathological data, giving histopathological information about the cancer's location and its stage, as diagnosed by the radiologist. These stages are related to the lesion size, its growth and the area of spread. Once the stage is clear, it helps the doctor decide which treatment is the most appropriate to enhance the patient's prognosis and promote a faster recovery. The TCGA-BRCA data set contains MRI image samples for stages I–III. Figs. 2–4 show sample MRI images for stages I, II and III, respectively. Of the abnormal cases identified in the MRI data set, the majority belong to stage II (Clark et al., 2013).

The selected MRI images are in standardized digital (DICOM) format, obtained using a standard double breast coil on a 1.5T GE whole-body MRI system (GE Medical Systems) (Li et al., 2016). The data set consists of T1-weighted, T2-weighted, pre-contrast and post-contrast images (Clark et al., 2013). Geir Torham et al. suggest a T2-weighted image modality for detecting abnormalities in breast parenchyma; hence only T2-weighted images are used for this study, with spacing between slices ranging from 2 to 3 mm. Most of them are size 512×512 or 256×256 with an average of 100 slices (Ali & Ray, 2013).

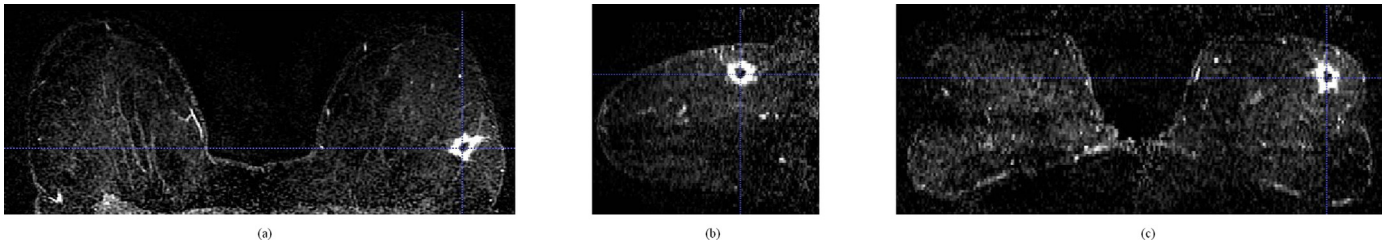

Fig. 2. Sample T2-weighted MRI displaying Stage I from the TCGA-BRCA data set, (a) axial view, (b) sagittal view, and (c) coronal view.

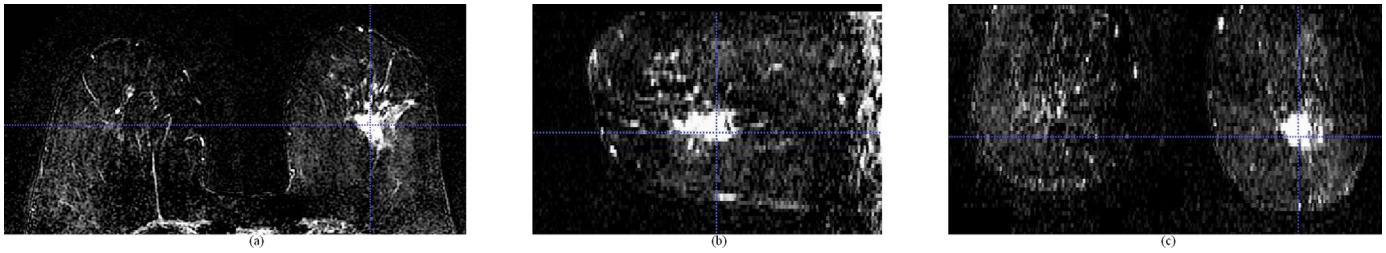

Fig. 3. Sample T2-weighted MRI displaying Stage II from the TCGA-BRCA data set, (a) axial view, (b) sagittal view, and (c) coronal view.

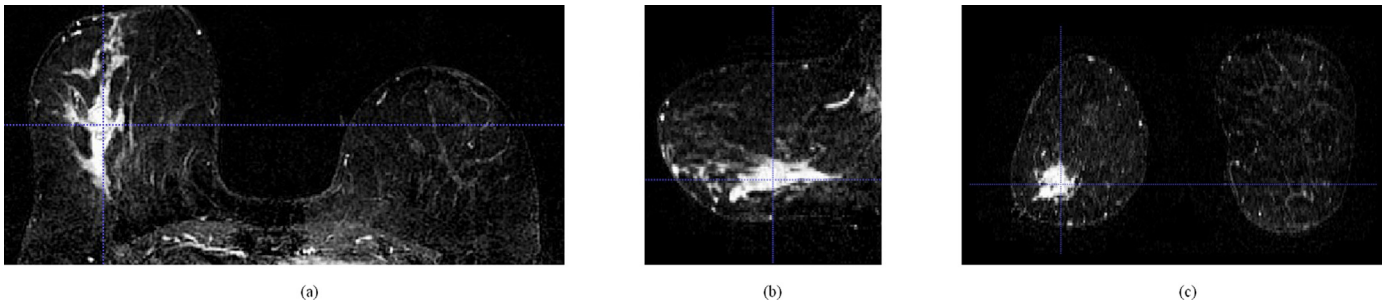

Fig. 4. Sample T2-weighted MRI displaying Stage III from the TCGA-BRCA data set, (a) axial view, (b) sagittal view and, (c) coronal view.

3.2. Image enhancement

To improve the diagnosis capability, the MRI images were first enhanced by the application of the CLAHE technique on a slice-by-slice basis.

CLAHE first partitions a slice/image into contextual sub-regions and calculates their respective histogram equalization (HE) using a specific number of bins, by applying an experimentally determined threshold. Lastly, each region's intensity values are mapped with respect to the new histogram results. The final CLAHE image is then reconstructed by interpolating the grey-level mapping (Sundaram, Ramar, Arumugam, & Prabin, 2011). The calculation of HE is described below.

3.2.1. Histogram equalization (HE)

HE is a conventional technique to adjust image intensities which applies non-linear, monotonic mapping of intensity values of the pixels so that the transformed image has a uniform distribution of intensity values. Let I be the image with pixel intensity values ranging from *0 to L-1*, where L is 256, then the probability of occurrence of intensity value r_k can be written as (Gonzalez & Woods, 2002):

$$p_r(r_k) = \frac{number\ of\ pixels\ with\ intensity\ k}{total\ number\ of\ pixels\ in\ the\ image} \quad k : 0\ to\ L-1 \tag{15}$$

The cumulative distribution function corresponding to p_r is given as:

$$cdf\ (k) = \sum_{i=0}^{k} p_r(r_k) \tag{16}$$

Thus, the histogram equalization acts as an image transformation method so that the transformed image is obtained by mapping each pixel with intensity r_k, to the corresponding level s_k given as:

$$s_k = (L-1) * cdf(k) \tag{17}$$

This transformation is termed histogram equalization.

3.3. Image segmentation

The regions of interest (ROI) were manually segmented from the surrounding parenchyma by sketching a boundary around the lesion, in three-dimensional space (using axial, sagittal and coronal views) identified by the radiologist, as provided in the corresponding pathology report. All the ROIs were resized as $64 \times 64 \times 64$. We extracted 193 (110 abnormal and 93 normal) ROIs from TCGA-BRCA MRI cases using the ITK-SNAP tool version 3.2.0 (Yushkevich et al., 2006). These ROIs were used to test the performance of the CDSS for distinguishing between normal and abnormal cases.

We extended our experiment further to identify cancer stages by analyzing the breast parenchyma as a whole, since stages are recognized by analyzing the size and the location spread of the lesion singly as well as multi-focally. The segmented breast region was resized to $128 \times 128 \times 128$. The statistics of the abnormal MRIs comprised 22 stage I, 70 stage II and 18 stage III cases.

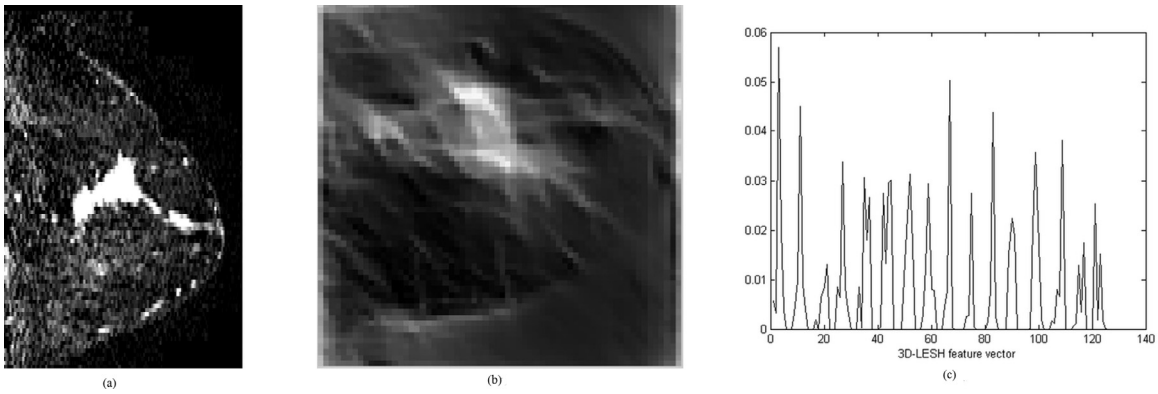

Fig. 5. Sample T2-weighted MRI image, its local energy map and 3D-LESH feature vector plot, (a) Sample T2-weighted MRI image in sagittal view, (b) Its local energy map, and (c) 3D-LESH feature vector.

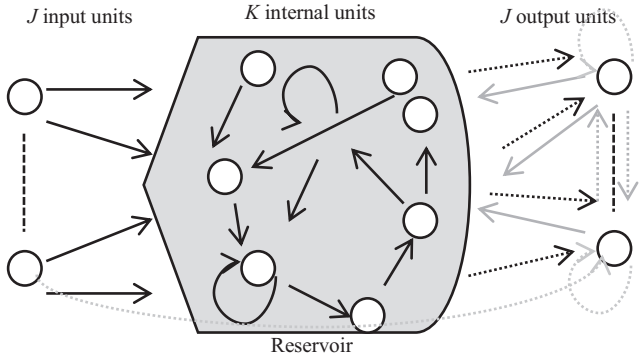

Fig. 6. Basic ESN architecture (Dotted lines are optional connections) (Jaeger, 2001).

3.4. Feature extraction

We applied the 3D-LESH feature extraction technique to ROIs extracted from the MRI images. For 3D-LESH computation, the Gabor filter was designed to control all the filters in the equator of the sphere, while orientations were integrated into the xy-plane. This set-up is appropriate for analyzing MRI images as their in-slice resolution is less significant than slice thickness (Ferrari et al., 2011).

We experimentally determined an appropriate spectrum coverage of the filters for the number of wavelet scales c as *5*, and the number of filter orientations n as 8. The wavelength of the minimal scale filter was set to *3*, and the scaling factor between successive filters was *2.1*. The ratio factor ($\frac{\sigma_\omega}{\omega_i}$) was taken as 0.55 and θ (the elevation) and ϕ (the azimuth) were each set to *1.2*.

The noise standard deviation was set to *2.0*, with a threshold of *0.5*, below which the phase congruency values are penalized. Fig. 5 shows a sample MRI image, the local energy map and its associated 3D-LESH feature plot.

The proposed technique was implemented in MATLAB R2010a and executed multiple times on a system that had an Intel Pentium i7- running at 2.3 GHz and 16GB RAM. Computation of 3D-LESH for a segmented breast region (size $128 \times 128 \times 128$) took 30.23 CPU time on average whereas in the case of segmented ROI of size $64 \times 64 \times 64$, it took 4.031 CPU time on average. The results are given in detail in Section 4.

3.5. Feature selection

We selected a subset of the N largest 3D-LESH coefficients, say h_N from 3D-LESH feature vectors, ($h_{r,b}$), for classification purposes. These selected feature coefficients were fed to a machine learning classifier to distinguish between the different stages and discriminate among normal and abnormal cancer cases.

3.6. Classification with selected machine learning classifiers

Classification is an integral part of CDSSs for the prediction and diagnosis of a specific disease. It assigns a class/label (from predefined classes) to new cases in a data set, on the basis of the information incorporated into their respective feature sets.

The selected 3D-LESH features h_N were fed to three widely used classifiers, namely SVM, ESN and ELM, and their comparative performance evaluated. A short description of each classifier is given below.

3.6.1. Support vector machine (SVM)

SVM, proposed by Vapnik (2000), is a supervised learning classification technique with a high generalization capability. Given a set of feature vectors of the training set in h_N, the SVM classifier draws a hyperplane which separates them into two distinct classes: abnormal (class label 1) and normal (class label 0). The hyperplane is given as below:

$$g(h_N) = w^T h_N + w_0 = 0 \tag{18}$$

where, w represents a vector normal to the hyperplane. The training of SVM produces a model which is used for testing the model capability to assign a specific class to a new unknown feature vector. If features are not linearly separable, they are transformed to a higher dimensional space using some kernel methods. We experimented with linear, polynomial and RBF kernels for this purpose, as described below.

3.6.1.1. Linear kernel. It is the most suitable choice if a feature set is linearly separable.

$$k(h_N, h_{N'}) = {h_N}^T h_{N'} + C \tag{19}$$

where C is the optimal constant (Wajid & Hussain, 2015).

3.6.1.2. Polynomial kernel. The polynomial kernel with degree d can be written as:

$$k(h_N, h_{N'}) = \left(\gamma {h_N}^T h_{N'} + C\right)^d \tag{20}$$

where γ and C are adjustable constants and d is the degree of the polynomial (Wajid & Hussain, 2015).

3.6.1.3. Gaussian, radial basis function (RBF) kernel.

$$k(h_N, h_{N'}) = exp\left(-\gamma \|h_N - h_{N'}\|^2\right) \ \gamma > 0 \tag{21}$$

where γ is a positive parameter to control the radius (Wajid & Hussain, 2015).

3.6.2. Echo State Network (ESN)

We demonstrate here the capability of ESN to classify abnormal/normal cases and to distinguish between different stages of cancer (for the case of binary as well as multiclass classifiers).

Given an ESN with J input network units (feature vectors) $h_N(m)=(h_{N1}(m), h_{N2}(m)\ldots h_{NJ}(m))^T$, a large, fixed-size ($K$) reservoir of recurrent neural networks (RNN), also termed internal units $x(m)=(x_1(m),x_2(m),\ldots, x_K(m))^T$, are generated from input signals. Further, J output signals $y(m)=(y_1(m),y_2(m),\ldots, y_J(m))^T$ are produced as a linear combination of these response signals from the reservoir (Jaeger H., 2001; Løkse S. et al., 2017; Zhong et al., 2017). The overall basic ESN structure with activation of the units at time step m is depicted in Fig. 6.

The internal units of the network are updated as:

$$x(m+1) = f\left(W^{in}hN(m+1) + W^{t}x(m) + W^{back}y(m)\right) \quad (22)$$

where $f=(f_1\ldots\ldots f_k)$ is the internal unit output function {mostly a sigmoid function tanh} at time step m. Further, the output is calculated as:

$$y(m+1) = f^{out}\left(W^{out}(hN(m+1), x(m+1), y(m))\right) \quad (23)$$

where these units within ESN are connected through real-valued connection weights as: W^{in} for the connection between input and internal units, W^{t} for the weights for the reservoir, W^{out} for the connection between the internal reservoir and outer units, and W^{back} for the weights for back propagation.

$f^{out} = (f_1^{out} \ldots\ldots\ldots f_L^{out})$ are output functions applied to combinations of the input, internal and previous output vectors at time step m (Jaeger H., 2001).

We conducted experiments using the ESN code produced by Jaeger (2001). Once ESN is trained, testing is performed and results are evaluated using performance measures described in the next section. For extending ESN to a multiclass problem (classification of multiple stages of cancer), the one-vs-all scheme is applied.

3.6.3. Extreme learning machine (ELM)

Huang (2015) presented ELM as a novel learning paradigm for a single-hidden-layer feed-forward neural network (SLFN), in which feature mapping between input and hidden layers is random and only weights between hidden and output layers are trained. Given J training samples $(h_{Ni},\ y_i)_{i=1}^{J} \in R^N \times R^1$, the output of an ELM, with $\tilde{N}$ hidden nodes, can be written as follows (Huang, 2015):

$$y_i = \sum_{j=1}^{\tilde{N}} \beta_j g_j(h_{Ni}) \ = \sum_{j=1}^{\tilde{N}} \beta_j g\left(h_{N_i}; a_j, b_j\right) \quad (24)$$

Where $a_j=[a_1, a_2\ldots.., a_N]^T$ and b_j are learning parameters for the jth hidden neuron and $\beta_j=[\beta_1, \ldots\ldots, \beta_{\tilde{N}}]^T$ is the weight vector between the hidden and output layers; $g(h_{Ni};\ a_j, b_j)$ is the activation function, which can be a non-linear piecewise continuous function, e.g. sigmoid, Fourier, hard-limit, Gaussian. The ELM structure is depicted in Fig. 7. A number of ELM variants have been reported in the literature to-date (Liu Y. et al., 2017; Guo T. et al., 2017).

4. Results and discussion

Experiments were conducted with the TCGA-BRCA data set to distinguish abnormal and normal cases and to diagnose the stage of abnormality. We report results using a classification accuracy measure which accounts for the number of correct predictions made from overall predictions. We also employ a receiver operating characteristic (ROC) curve to measure the significance of the results generated. The ROC is plotted as the true positive rate (TPR) against the false positive rate (FPR). The area under the curve (AUC) A_z lies within the range 0 to 1 (1 being the highest performance), and is the probability of how effective a classifier can rank

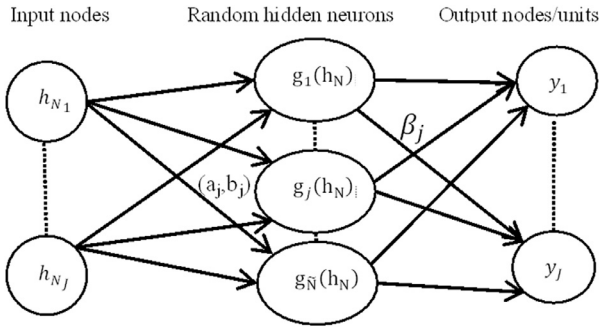

Fig. 7. A typical extreme learning machine structure with input nodes h_N, hidden nodes $g(h_N)$ and output nodes y (Huang, 2015).

a randomly selected positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative') (Fawcett, 2006).

We applied a grid search with ten-fold cross-validation to determine suitable parameters for SVM classifiers. In the case of SVM with RBF kernel, the slope parameter γ and the penalty parameter C were tuned, while in the case of SVM with the linear kernel, only C was tuned and the polynomial kernel of degree two was selected for experimentation. All results were averaged using ten-fold cross-validation and are presented in Tables 3 and 4.

The results indicate that 3D-LESH features combined with SVM classifiers provide an efficient classification performance (Table 3). The most significant performance was achieved by the SVM linear classifier, which had a classification accuracy of $99.47\pm0.47\%$, when 100 or all 3D-LESH features were selected (Table 3), and an A_z value of 0.9956 ± 0.0044 (Table 3) for classification of normal and abnormal cases. Similarly, SVM with the RBF kernel also performed well, attaining an accuracy of 98.95 ± 0.052 (Table 3) and an A_z value of 0.9909 ± 0.056 (Table 3).

As can be seen in Table 3, ELM resulted in a high classification accuracy of $99.47\pm1.03\%$ and A_z value of 0.9856 ± 0.0033 overall. Finally, the SVM with polynomial kernel had a maximum performance of 98.48% and A_z value of 0.9909, when 100 3D-LESH features were selected. The ESN similarly showed a maximum performance of 91.52% and A_z value of 0.9226 for 100 selected 3D-LESH features.

It is further evident from comparative results in Table 4 that 3D-LESH is a good choice for detecting the various cancer stages. In this case, ELM outperformed every other classifier by distinguishing the different stages with a classification accuracy of 95.45% for stage I, 98.18% for stage II and 98.18% for stage III, when all features were selected. The SVM linear classifier was found to be capable of distinguishing between the different stages of breast cancer with a maximum accuracy of 96.09% for stage I, 86.36% for stage II and 91.09% for stage III (Table 4). The rest of the classifiers showed a lower overall performance.

In all cases, multiclass classification was performed on the basis of the 'one-vs-all' scheme to distinguish between different stages of abnormalities. The application of the proposed feature set in conjunction with machine learning classifiers affirmed the resilience of 3D-LESH features.

In summary, the results reported in Tables 3 and 4 emphasize the superiority of 3D-LESH features for distinguishing between different stages of cancer, as well as for detecting the existence of cancer in the medical images. Furthermore, it is evident that selecting a subset of all available features does not deteriorate the classification performance. The ROC curve for different classifiers, depicted in Fig. 8, provides a comparison of the overall performance of classifiers.

Table 3
3D-LESH-based classification accuracy % (Abnormal/Normal) for MRI data set.

Features Selected	SVM with RBF		SVM with Linear		SVM with Polynomial		ESN		ELM	
	Accuracy	Az	Accuracy	Az	Accuracy	Az	Accuracy	Az	Accuracy	Az
50	98.95%	0.9909	98.94%	0.9954	98.44%	0.9859	90.65%	0.8946	97.89%	0.9758
100	99.47%	0.9954	99.47%	0.9954	98.42%	0.9909	91.52%	0.9226	98.92%	0.9894
All	98.43%	0.9843	99.48%	1.0000	98.45%	0.9864	85.83%	0.8990	99.47%	0.9861

Table 4
Selected 3D-LESH features (50, 100 & All) with multiclass (One-vs-All) SVM. Results show classifier performance accuracy % for the MRI data set.

Stage	SVM with Linear			SVM with RBF			SVM with Polynomial			ESN			ELM		
	50	100	All	50	100	All	50	100	All	50	100	All	50	100	All
I	88.09	71.72	96.09	72.72	85.37	89.36	60.90	83.86	93.09	86.87	80.89	89.99	97.27	95.45	95.45
II	73.45	72.72	86.36	77.27	53.63	84.46	67.27	77.72	81.81	73.63	81.78	86.79	97.27	95.45	98.18
III	78.18	70.91	91.09	77.73	68.18	86.54	81.18	83.36	83.36	87.90	84.45	92.34	96.36	95.45	98.18

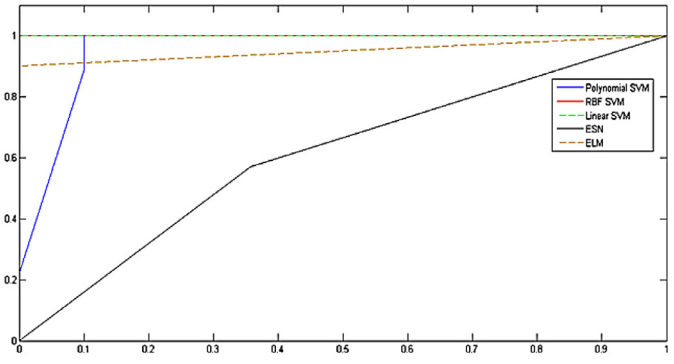

Fig. 8. ROC curve for classification of abnormal/normal cases in the MRI data set (linear SVM and RBF SVM curves are seen to be overlapping, whereas the polynomial SVM curve is partially overlapping, with part of it invisible). It can be seen that the performance of linear SVM and ELM is higher than other classifiers whereas ESN shows notably inferior performance.

4.1. Comparative analysis of 3D-LESH and wavelet-based feature extraction for classification performance

We compared the performance of 3D-LESH with a benchmark wavelet feature extraction technique (Ferreira & Borges, 2003; Mousa, Munib, & Moussa, 2005). Ferreira and Borges (2003) applied the Daubechies wavelet transform (Daubechies, 1992) to two-dimensional mammograms by first decomposing the image into four levels, as demonstrated by Mallat (1989). Subsequently, only low-level frequency coefficients were selected as features, as they demonstrated the capability to enhance classification accuracy among multiple abnormality types, and to detect abnormal cases in mammograms.

In this paper, we extended the wavelet technique (Ferreira & Borges, 2003) to 3D MRI images by selecting 100 largest wavelet coefficients, which were experimentally determined to be the most appropriate choice for reducing the effect of curse of dimensionality, without compromising performance. Ferreira and Borges (2003) also experimented with different numbers of selected, high-degree wavelet coefficients and applied SVM to classify these. They found 100 largest selected features to be the most appropriate choice, which concurred with our findings.

The results for 3-D LESH and wavelet-based feature extraction techniques are reported in Tables 3, 4 and 5 respectively. The comparative results confirm the superior performance of 3D-LESH features in both binary and multiclass classification cases.

4.2. Statistical analysis of results using the t*-test*

We compared performance of the 3D-LESH feature extraction technique with the benchmark wavelet texture feature extraction technique, using the *t-test* at a significance level of 0.05. For this experiment, we selected the ELM classifier since it resulted in a comparatively higher performance in cases of binary as well as multiclass classifications. Other classifiers' performance may also be evaluated, which is proposed for future study. Let μ_L and μ_W be mean performance accuracies for 3D-LESH and wavelet-based methods respectively (with the mean calculated from ten-fold cross-validation results). We tested the following hypothesis:

$$H_0 : \mu_L - \mu_w = 0 \quad (LESH\ performs\ like\ wavelets)\ \text{against} \tag{25}$$

$$H_0 : \mu_L - \mu_w = 0 \quad (LESH\ performs\ better\ than\ wavelets) \tag{26}$$

The results are reported in Table 6. In both binary and multistage cases, the p-value is seen to be higher than the *0.05* confidence level, which suggests the difference in the classification accuracy is significant. Thus, it can be stated that our proposed 3D-LESH feature extraction technique performs significantly better than the wavelet-based approach. In future, other state-of-the-art feature extraction methods may also be compared with our technique.

The *t*-test results (Table 7) testify to the significance of our proposed algorithm since, regardless of the choice of classifier, the 3D-LESH outperforms wavelet feature descriptors. This is due to the fact that the wavelets fail to represent discontinuities along curves and edges. The wavelet feature descriptors primarily focus on a specific type of abnormality and try to detect its malignancy, whereas the 3D-LESH-based methodology has the advantage of detecting the malignancy of any type of abnormality and can differentiate among different stages of malignancy with considerable efficiency.

Our proposed 3D-LESH feature descriptors are computed as a histogram of local energy; the highest degree 3D-LESH feature coefficients thus represent the most prominent set of features in the local area within an image. We experimented with different subsets of 3D-LESH features and observed their effect on the overall system performance (Tables 3 & 4). Our experiments showed that a subset of these features can be selected with minimal compromise on classification accuracy, thus reducing dimensionality.

4.3. Visual evaluation of the 3D-LESH technique

For further evaluation of our proposed technique, we visualized 3D-LESH features for the MRI dataset with different classification

Table 5
Wavelet (Daubechies)- based feature extraction classification results for binary classification of normal and abnormal cases for the MRI data set.

Measure	SVM with RBF	SVM with Linear	SVM with Polynomial	ESN	ELM
Accuracy %	71.36	72.42	73.47	50.50	93.28
A_z	0.7460	0.7568	0.7659	0.5443	0.9249

Table 6
Wavelet-based feature extraction for multiclass (One-vs-All) classification performance accuracy % for the MRI data set.

Abnormality type	SVM with RBF	SVM with linear	SVM with polynomial	ESN	ELM
Stage I	70.90	66.36	68.18	66.67	85.45
Stage II	57.27	83.63	55.45	49.80	84.54
Stage III	88.18	93.63	93.63	75.00	85.98

Table 7
The results of the *t*-test at level of significance = 0.05 for binary classification.

Method	Alternate Hypothesis Ha	P-value	T-value	Null Hypothesis H_0
SVM (Linear)	$\mu_L - \mu_w > 0$	0.00004	9.39	Reject
SVM (Polynomial)	$\mu_L - \mu_w > 0$	0.000019	7.45	Reject
SVM (RBF)	$\mu_L - \mu_w > 0$	0.000002	9.43	Reject
ESN	$\mu_L - \mu_w > 0$	0.000004	8.94	Reject
ELM	$\mu_L - \mu_w > 0$	0.038	2.002	Reject

techniques, using the method suggested by Levman et al. (2008). The latter is based on computation of principal component analysis (PCA) for dimensionality reduction of the feature vectors. Other popular methods found in the literature include self-organizing maps for feature set visualization of lesions in MRI data (Nattkemper & Wismüller, 2005), an SVM-based method (Komura, Nakamura, Tsutsumi, Aburatani, & Ihara, 2004) and a relative distance map approach (Somorjai et al., 2004).

In our case, 3D-LESH feature vectors are of 128 length and hence challenging to visualize. We performed PCA to reduce their dimensionality. PCA is a powerful tool to measure variation in observations while discovering a useful relationship between them. It rotates the feature vectors in such a way that the resultant orthogonal axes are aligned to the maximum variance in them (Levman et al., 2008). Hence to visualize the capability of our 3D-LESH features to discriminate normal and abnormal cases, we projected the 3D-LESH space to two and three dimensional spaces by application of PCA, plotting only two or three PCA components, since they preserve the highest Eigen values. Fig. 9 shows the 3D-LESH features (100 largest selected coefficients) in a three-dimensional PCA component space. We can see that these features exhibit the potential to help discriminate abnormal cases from normal ones. Fig. 10 depicts a visual representation of the feature vectors in a two-dimensional PCA component space (for PCA components one and two) for the three selected classifiers. It is clear that the discriminating power of 3D-LESH enhances the classification performance.

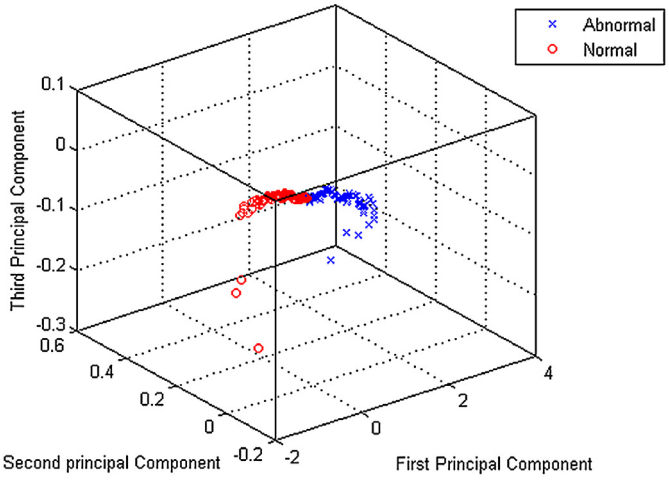

Fig. 9. Abnormal and normal cases plotted in three principal component space account for 85% of total variance in 3D-LESH feature vectors (comprising the 100 largest 3D-LESH coefficients).

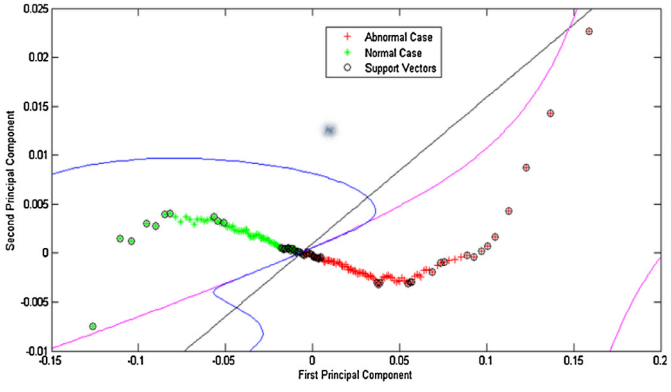

Fig. 10. Abnormal and normal cases plotted in two principal component spaces account for 64% of total variance in 3D-LESH feature vectors (comprising the 100 largest 3D-LESH coefficients). The decision boundary lines are shown for: Linear kernel SVM (black), RBF kernel-based SVM (blue) and Polynomial kernel SVM (pink). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. Clinical significance of the proposed CDSS

Research in clinical decision support systems for health care is continuously evolving, leading to the emergence of novel scientific challenges at the intersection of medical science, patient care and information technology.

MRI is known to be the most sensitive modality for breast cancer detection, one that efficiently highlights abnormalities in the breast region, and is thus recommended for women at risk of breast cancer.

Patient survival is proportionate to the stage of the cancer at diagnosis. According to Nover et al, "It is 98% for a10-year survival rate for patients with stages 0 and I of the disease whereas

it is 65% for a10-year survival rate for patients with stage III disease" (Nover et al., 2009). It is thus vital to identify the disease at an early stage. Cancer stages in MRI images are labelled using TNM representation by the American Joint Committee on Cancers (AJCC), where T(tumour) corresponds to the size and location of the tumour, N(node) corresponds to where the cancer has spread in the lymph node and M(Metastasis) corresponds to whether it has reached other parts of the body (Nover et al., 2009).

By diagnosing cancer, as well as its stage, physicians can plan appropriate treatments, namely, the type of surgery, chemotherapy or radiation therapy, in addition to better predicting the chance of recovery and the likely recurrence of cancer (Edge & Compton, 2010).

Our proposed 3D-LESH-based CDSS is capable of providing enhanced patient care through early diagnosis of breast cancer in women via MRI analysis. It can also offer a more accurate prognosis of the disease, enabling appropriate laboratory tests to be performed and suitable therapy and drug dosage to be determined.

5. Conclusions and future work

Intelligent CDSS or expert systems for diagnosing chronic diseases are vital. As highlighted in in this paper, research carried out in this area has mainly involved diverse experiments using benchmark medical datasets, to improve the chances of early diagnosis and greater survival rates. For example, in other related work, Yang and Chen suggested the application of data-mining techniques, such as decision tree classification and rule-based methods, for detecting lung cancer stages, using patient data extracted from pathological reports accessed from the TCGA website (Yang & Chen, 2015). Neves et al. proposed another system, combining multi-scale linearity with SVM and polynomial classification techniques to detect prostate cancer. They reported encouraging classification performance with an AUC value of 0.906 (Neves & Nascimento, 2009). Lee and Phoebe suggested a cloud computing framework to analyse medical images for detection of diverse kinds of cancer, namely, breast, skin, prostate and lung cancers (Lee and Phoebe, 2015).

This article is another contribution to the growing field of CDSS applications. It provides an effective hybrid framework for assessing breast cancer based on an analysis of volumetric medical images. The analysis is carried out by applying novel image processing and machine-learning techniques in multiple stages, as part of our proposed CDSS framework, to identify the existence of potentially cancerous cells and various cancer stages. This can enable greater accuracy in cancer prognosis leading to timely interventions and treatments.

Comparative experimental results reported in this paper demonstrate that, when 3D-LESH features are combined with SVM, ESN and ELM machine learning classifiers, the overall accuracy attained by our proposed feature extraction technique is higher, compared to many of the techniques found in literature.

It is also evident from the results that a subset of these features can be selected without compromising classification accuracy, hence reducing the dimensionality. We experimented with different numbers of highest degree 3D-LESH coefficients and found that $N = 100$ is most appropriate for improving the classification accuracy, while reducing the effect of curse of dimensionality (Wajid & Hussain, 2015).

In the future, more extensive evaluation and clinical validation will be performed with additional 3D clinical data sets in conjunction with recently reported feature selection and classification methods. We also intend to employ a new type of a hybrid approach to optimize an ensemble feature set adaptively, on the basis of a combination of 3D-LESH and other state-of-the-art feature extraction techniques, while selecting the most significant features to optimise classification performance. Furthermore, an intensive study of the clinical significance of 3D-LESH features when applied to a real clinical data set may lead to potentially useful clinical insights.

Finally, we plan to experiment with both deep and reenforcement learning algorithms (Mahmud et al., 2018; Zhong G. et al., 2018) as well as SVM variants such as the arbitrary norm SVM (Huang et al., 2010), the sparsely connected SVM (Huang, Zheng, King, & Lyu, 2009) and the multi-layered ESN (Malik, Hussain, & Wu, 2017), to further enhance and optimise classification of volumetric medical images. We will also exploit sensitivity analysis (SA) to identify 3D-LESH features that have minimal effect on classification results. SA can faciltate the ranking of features according to their effect on the model output, helping reduce imperfections related to dimensionality reduction (Boulila, Bouatay, & Farah, 2014; Ferchichi, Boulila, & Farah, 2015a, and 2015b).

Acknowledgements

Professor A. Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/M026981/1, and also by the Royal Society of Edinburgh (RSE) and National Natural Science Foundation of China (NNSFC) joint-project grant no. 61411130162. We also wish to thank the anonymous reviewers who helped improve the quality of the manuscript. The work was also partially supported by the National Natural Science Foundation of China (NSFC) under grant no. 61473236, Suzhou Science and Technology Program under grant no. SYG201712, SZS201613, and Key Program Special Fund in XJTLU under grant no. KSF-A-01.

References

Agliozzo, S., De Luca, M., Bracco, C., Vignati, A., Giannini, V., Martincich, L., et al. (2012). Computer-aided diagnosis for dynamic contrast-enhanced breast MRI of mass-like lesions using a multiparametric model combining a selection of morphological, kinetic, and spatiotemporal features. *Medical Physics, 39*(4), 1704–1715. https://doi.org/10.1118/1.3691178.

Aghaei, F., Tan, M., Hollingsworth, A. B., Qian, W., Liu, H., & Zheng, B. (2015). Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy. *Medical Physics, 42*(11), 6520–6528. https://doi.org/10.1118/1.4933198.

Agner, S. C., Xu, J., Fatakdawala, H., Ganesan, S., Madabhushi, A., Englander, S., et al. (2009). Segmentation and classification of triple negative breast cancers using DCE-MRI. In *2009 IEEE international symposium on biomedical imaging: From nano to macro* (pp. 1227–1230).

Ali, F., & Ray, S. (2013). SAR Analysis for handheld mobile phone using DICOM-based voxel model. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications, 12*(2), 363–375.

American Cancer Society. (2016). *Breast cancer facts & figures 2015-2016*. Atlanta: American Cancer Society, Inc.

Aminololama-Shakeri, S., & Khatri, V. P. (2014). Emerging modalities in breast cancer imaging. *Surgical Oncology Clinics of North America, 23*(4), 735–749.

Arbach, L., Stolpen, A., & Reinhardt, J. M. (2004). Classification of breast MRI lesions using a backpropagation neural network (BNN). In *2004 2nd IEEE international symposium on biomedical imaging: Macro to nano (IEEE Cat No. 04EX821): 2* (pp. 253–256).

Boulila, W., Bouatay, A., & Farah, I. R. (2014). A probabilistic collocation method for the imperfection propagation: Application to land cover change prediction. *Journal of Multimedia Processing and Technologies, 5*(1). Retrieved from http://www.dline.info/jmpt/fulltext/v5n1/2.pdf.

Chen, W., Giger, M. L., Bick, U., & Newstead, G. M. (2006). Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. *Medical Physics, 33*(8), 2878–2887.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging, 26*(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7.

Daubechies, I. (1992). *Ten lectures on wavelets*. PA: SIAM.

Dosil, R., Pardo, X. M., & Fdez-Vidal, X. R. (2005). Decomposition of three-dimensional medical images into visual patterns. *IEEE Transactions on Biomedical Engineering, 52*(12), 2115–2118.

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of Surgical Oncology, 17*(6), 1471–1474.

ElNawasany, A. M., Ali, A. F., & Waheed, M. E. (2014). *A novel hybrid perceptron neural network algorithm for classifying breast MRI tumors* (pp. 357–366). Cham: Springer.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Ferchichi, A., Boulila, W., & Farah, I. R. (2015a). An intelligent possibilistic approach to reduce the effect of the imperfection propagation on land cover change prediction (pp. 520–529).

Ferchichi, A., Boulila, W., & Farah, I. R. (2015b). Using evidence theory in land cover change prediction to model imperfection propagation with correlated inputs parameters. *2015 7th International joint conference on computational intelligence (IJCCI)*.

Ferrari, R. J., Allaire, S., Hope, A., Kim, J., Jaffray, D., & Pekar, V. (2011). Detection of point landmarks in 3D medical images via phase congruency model. *Journal of the Brazilian Computer Society, 17*(2), 117–132. https://doi.org/10.1007/s13173-011-0032-8.

Ferreira, C. B. R., & Borges, D. L. (2003). Analysis of mammogram classification using a wavelet transform decomposition. *Pattern Recognition Letters, 24*(7), 973–982.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.) ISBN 0-201-18075-80.

Guo, T., Zhang, L., & Tan, X. (2017). Neuron pruning-based discriminative extreme learning machine for pattern classification. *Cognitive Computation, 9*(4), 581–595. doi:10.1007/s12559-017-9474-4.

Hassanien, A. E., & Kim, T. (2012). Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks. *Journal of Applied Logic, 10*(4), 277–284.

Hassanien, A. E., Moftah, H. M., Azar, A. T., & Shoman, M. (2014). MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. *Applied Soft Computing, 14*, 62–71. https://doi.org/10.1016/j.asoc.2013.08.011.

Honda, E., Nakayama, R., Koyama, H., & Yamashita, A. (2016). Computer-aided diagnosis scheme for distinguishing between benign and malignant masses in breast DCE-MRI. *Journal of Digital Imaging, 29*(3), 388–393.

Huang, G.-B. (2015). What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cognitive Computation, 7*(3), 263–278.

Huang, K., Zheng, D., King, I., & Lyu, M. R. (2009). Arbitrary norm support vector machines. *Neural Computation, 21*(2), 560–582. https://doi.org/10.1162/neco.2008.12-07-667.

Huang, K., Zheng, D., Sun, J., Hotta, Y., Fujimoto, K., & Naoi, S. (2010). Sparse learning for support vector classification. *Pattern Recognition Letters, 31*(13), 1944–1951.

Huang, Y.-H., Chang, Y.-C., Huang, C.-S., Wu, T.-J., Chen, J.-H., & Chang, R.-F. (2013). Computer-aided diagnosis of mass-like lesion in breast MRI: Differential analysis of the 3-D morphology between benign and malignant tumors. *Computer Methods and Programs in Biomedicine, 112*(3), 508–517. https://doi.org/10.1016/j.cmpb.2013.08.016.

Jaeger, H. (2001). The echo state approach to analyzing and training recurrent neural networks - with an Erratum note. *GMD Report 148*: German National Research Center for Information Technology.

Keyvanfard, F., Shoorehdeli, M. A., Teshnehlab, M., Nie, K., & Su, M.-Y. (2013). Specificity enhancement in classification of breast MRI lesion based on multi-classifier. *Neural Computing and Applications, 22*(S1), 35–45.

Komura, D., Nakamura, H., Tsutsumi, S., Aburatani, H., & Ihara, S. (2004). Multidimensional support vector machines for visualization of gene expression data, 175–179.

Kovesi, P. D. (1999). Image features from phase congruency. Videre: Journal of Computer Vision Research, *MIT Press 1, 3*, 1–26.

Kovesi, P. D. (2000). Phase congruency: A low-level image invariant. *Psychological Research, 64*, 136–148.

Kovesi, P. (2003). Phase congruency detects corners and edges. In The Australian pattern recognition society conference: DICTA 2003, (pp. 309–318).

Kumar, G., & Bhatia, P. K. (2014). A detailed review of feature extraction in image processing systems. In *2014 Fourth international conference on advanced computing & communication technologies* (pp. 5–12).

Lee, S. H., Kim, J. H., Cho, N., Park, J. S., Yang, Z., Jung, Y. S., et al. (2010). Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. *Medical Physics, 37*(8), 3940–3956.

Lee, Howard, & Phoebe, Yi-Ping (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications, 42*(12), 5356–5365 ISSN 0957-4174.

Levman, J., Leung, T., Causer, P., Plewes, D., & Martel, A. L. (2008). Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines. *IEEE Transactions on Medical Imaging, 27*(5), 688–696. https://doi.org/10.1109/TMI.2008.916959.

Li, H., Zhu, Y., Burnside, E. S., Huang, E., Drukker, K., Hoadley, K. A., et al. (2016). Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer, 2*(1), 16012.

Liu, N., Sakamoto, J. T., Cao, J., Koh, Z. X., Ho, A. F. W., Lin, Z., & Ong, M. E. H. (2017). Ensemble-Based Risk Scoring with Extreme Learning Machine for Prediction of Adverse Cardiac Events. *Cognitive Computation, 9*(4), 545–554. https://doi.org/10.1007/s12559-017-9455-7.

Løkse, S., Bianchi, F. M., & Jenssen, R. (2017). Training Echo State Networks with Regularization Through Dimensionality Reduction. *Cognitive Computation, 9*(3), 364–378. Available at https://doi.org/10.1007/s12559-017-9450-z .

Malik, Z. K., Hussain, A., & Wu, Q. J. (2017). Multilayered echo state machine: A novel architecture and algorithm. *IEEE Transactions on Cybernetics, 47*(4), 946–959. doi:10.1109/TCYB.2016.2533545.

Mahmud, M., Kaiser, M., Hussain, A., & Vassanelli, S. (2018). Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Transactions on Neural Networks and Learning Systems*, 1–17.

Mahrooghy, M. (2013). In K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Eds.), *Heterogeneity wavelet kinetics from DCE-MRI for classifying gene expression based breast cancer recurrence risk BT – Medical image computing and computer-assisted intervention – MICCAI 2013: 16th international Conference, Nagoya, Japan, September 22–26* (pp. 295–302). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intelligence, 11*(7), 674–693.

Meinel, L. A., Stolpen, A. H., Berbaum, K. S., Fajardo, L. L., & Reinhardt, J. M. (2007). Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *Journal of Magnetic Resonance Imaging, 25*(1), 89–95.

Menezes, G. L. G., Knuttel, F. M., Stehouwer, B. L., Pijnappel, R. M., & van den Bosch, M. A. A. J. (2014b). Magnetic resonance imaging in breast cancer: A literature review and future perspectives. *World Journal of Clinical Oncology, 5*(2), 61–70. https://doi.org/10.5306/wjco.v5.i2.61.

Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters, 6*(5), 303–313. https://doi.org/10.1016/0167-8655(87)90013-4.

Mousa, R., Munib, Q., & Moussa, A. (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems with Applications, 28*(4), 713–723.

Nattkemper, T. W., & Wismüller, A. (2005). Tumor feature visualization with unsupervised learning. *Medical Image Analysis, 9*(4), 344–351. https://doi.org/10.1016/j.media.2005.01.004.

Neves, L. A., & Nascimento, M. Z. (2009). Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer. *In Expert Systems with Applications, 41*(11), 5017–5029 ISSN 0957-4174.

Nover, A. B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W. Y., Shih, W.-H., et al. (2009). Modern breast cancer detection: A technological review. *International Journal of Biomedical Imaging*.

Ojeda-Fournier, H., de Guzman, J., & Hylton, N. (2013). Breast magnetic resonance imaging for monitoring response to therapy. *Magnetic Resonance Imaging Clinics of North America, 21*(3), 533.

Palestrant, S., Comstock, C. E., & Moy, L. (2014). Approach to breast magnetic resonance imaging interpretation. *Radiologic Clinics of North America, 52*(3), 563–583.

Sarfraz, M. S., & Hellwich, O. (2008a). Head pose estimation in face recognition across pose scenarios. *International Conference on Computer Vision Theory and Applications VISAPP, Portugal, 1*, 235–242.

Sarfraz, M. S., & Hellwich, O. (2008b). An efficient front-end facial pose estimation system for face recognition. *International Journal of Pattern Recognition and Image Analysis, Springer, 18*, 434–441.

Sarfraz, M. S., & Hellwich, O. (2009). On head pose estimation in face recognition. *Computer Vision and Computer Graphics. Theory and Applications, Lecture Notes CCIS, 24*, 162–175.

Seely, J. M. (2012). Management of breast magnetic resonance imaging-detected lesions. *Canadian Association of Radiologists Journal, 63*(3), 192–206. https://doi.org/10.1016/j.carj.2010.11.003.

Shulman, L., Willett, W., Sievers, A., & Knaul, F. (2010). Breast cancer in developing countries: Opportunities for improved survival. *Journal of Oncology, 2010*, 1–6.

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA: ACancer Journal for Clinicians, 66*, 7–30. doi:10.3322/caac.21332.

Somorjai, R. L., Dolenko, B., Demko, A., Mandelzweig, M., Nikulin, A. E., Baumgartner, R., et al. (2004). Mapping high-dimensional data onto a relative distance plane—an exact method for visualizing and characterizing high-dimensional patterns. *Journal of Biomedical Informatics, 37*(5), 366–379.

Sundaram, M., Ramar, K., Arumugam, N., & Prabin, G. (2011). Histogram-based contrast enhancement for mammogram images. In *2011 International conference on signal processing, communication, computing and networking technologies* (pp. 842–846).

Szab, B., Wiberg, M., Bon, B., & Aspelin, P. (2004). Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast. *European Radiology, 14*(7), 1217–1225.

Vapnik, V. N. (2000). *The nature of statistical learning theory*. New York, NY: Springer New York.

Wajid, S. K., Hussain, A., & Luo, B. (2014). An efficient Computer-Aided Decision Support System for breast cancer diagnosis using Echo State Network classifier. In *2014 IEEE symposium on computational intelligence in healthcare and e-health (CICARE)* (pp. 17–24). IEEE.

Wajid, S. K., & Hussain, A. (2015). Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert Systems with Applications, 42*(20), 6990–6999.

Wajid, S. K., Hussain, A., Huang, K., & Boulila, W. (2016). Lung cancer detection using Local Energy-based Shape Histogram (LESH) feature extraction and cognitive machine learning techniques. In *2016 IEEE 15th International conference on cognitive informatics & cognitive computing (ICCI*CC)* (pp. 359–366). IEEE.

Yang, H., & Phoebe Chen, Yi-Ping (2015). Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information. In *Expert systems with applications: Vol. 42* (pp. 6168–6176). Issues 15–16.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage, 31*(3), 1116–1128.

Zhong, G., Yan, S., & Huang, K. (2018). *Cogn Comput, 10*, 179. https://doi.org/10.1007/s12559-017-9515-z.

Zhong, S., Xie, X., Lin, L., & Wang, F. (2017). Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing, 238*, 191–204.

Zakir, U., Zafar, I., & Edirisinghe, A. E. (2011). Road sign detection and recognition by using local energy-based shape histogram (LESH). *International Journal of Image Processing, 4*, 566–582.