

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part B

Faculty of Engineering and Information
Sciences

2017

Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks

Reza Rasti

Isfahan University of Medical Sciences

Mohammad Teshnehlab

K. N. Toosi University of Technology

Son Lam Phung

University of Wollongong, phung@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers1>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Rasti, Reza; Teshnehlab, Mohammad; and Phung, Son Lam, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks" (2017). *Faculty of Engineering and Information Sciences - Papers: Part B*. 1496.

<https://ro.uow.edu.au/eispapers1/1496>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks

Abstract

This work addresses a novel computer-aided diagnosis (CAD) system in breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). The CAD system is designed based on a mixture ensemble of convolutional neural networks (ME-CNN) to discriminate between benign and malignant breast tumors. The ME-CNN is a modular and image-based ensemble, which can stochastically partition the high-dimensional image space through simultaneous and competitive learning of its modules. The proposed system was assessed on our database of 112 DCE-MRI studies including solid breast masses, using a wide range of classification measures. The ME-CNN model composed of three CNN experts and one convolutional gating network achieves an accuracy of 96.39%, a sensitivity of 97.73% and a specificity of 94.87%. The experimental results also show that it has competitive classification performances compared to three existing single-classifier methods and two convolutional ensemble methods. The proposed ME-CNN model could provide an effective tool for radiologists to analyse breast DCE-MRI images.

Disciplines

Engineering | Science and Technology Studies

Publication Details

R. Rasti, M. Teshnehlab & S. Lam. Phung, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognition*, vol. 72, pp. 381-390, 2017.

Breast Cancer Diagnosis in DCE-MRI using Mixture Ensemble of Convolutional Neural Networks

Reza Rasti^{a,b}, Mohammad Teshnehlab^c, Son Lam Phung^d

^aArtificial Intelligent Lab, Department of Electrical Engineering of K. N. Toosi University of Technology, Tehran, Iran

^bDepartment of Biomedical Engineering, Isfahan University of Medical Sciences, Isfahan, Iran

^cDepartment of Electrical Engineering of K. N. Toosi University of Technology, Tehran, Iran

^dSchool of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Australia

Abstract

This work addresses a novel computer-aided diagnosis (CAD) system in breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). The CAD system is designed based on a mixture ensemble of convolutional neural networks (ME-CNN) to discriminate between benign and malignant breast tumors. The ME-CNN is a modular and image-based ensemble, which can stochastically partition the high-dimensional image space through simultaneous and competitive learning of its modules. The proposed system was assessed on our database of 112 DCE-MRI studies including solid breast masses, using a wide range of classification measures. The ME-CNN model composed of three CNN experts and one convolutional gating network achieves an accuracy of 96.39%, a sensitivity of 97.73% and a specificity of 94.87%. Experimental results also show that it has competitive classification performances compared to three existing single-classifier methods and two convolutional ensemble methods. The proposed ME-CNN model could provide an effective tool for radiologists to analyse breast DCE-MRI images.

Keywords: Breast cancer, DCE-MRI, convolutional neural networks, mixture ensemble of experts, CAD systems.

1. Introduction

Breast cancer is the second leading cause worldwide of cancer-related deaths in women, exceeded only by lung cancer [1]. According to the American Cancer Society, in the US alone an estimate of 231,840 new cases of invasive breast cancer and 40,290 breast cancer deaths occurred in 2015. On a more positive note, death rates from breast cancer have been declining since 1989, with larger decreases in women under the age of 50. One reason for this decline is attributed to earlier detection of breast masses via screening, which is an important step in the treatment of this disease [2].

In breast image analysis for cancer diagnosis, the multiplicity and complexity of the lesions may occur with dense tissue interactions, so it will be difficult for radiologists to detect and analyze masses precisely. Therefore, many imaging techniques and computer aided systems

Email addresses: mr.r.rasti@ieee.org (Reza Rasti), teshnehlab@eetd.kntu.ac.ir (Mohammad Teshnehlab), phung@uow.edu.au (Son Lam Phung)

Preprint submitted to Pattern Recognition

June 24, 2017

have been developed to assist radiologists. These developments have a major contribution to the identification and treatment of breast cancer.

It has been shown that DCE-MRI between different clinical imaging is a valuable diagnostic tool for breast cancer due to the high sensitivity and high resolution in dense breast tissues [3]. Compared to mammography and ultrasound, MR imaging is the most sensitive imaging modality for detecting breast cancer among high-risk populations [4]. Correct interpretation of breast DCE-MRI images depends significantly on the visualization quality, operator experience, and time available for data analysis. Because manual analysis of MR slices is time-consuming and error-prone, many dedicated systems have been developed to assist radiologists in the localization or diagnosis of breast lesions. Although some systems are currently used in clinical situations, fully automatic detection or diagnosis of breast lesions is still an open problem in DCE-MRI [5, 6].

This study proposes a novel CAD system for breast DCE-MRI. The system has two major stages: i) tumor candidate segmentation based on the intensity and morphological information of the masses in the image; and ii) tumor classification based on a new deep learning ensemble of convolutional neural networks (CNNs). The performances of the proposed segmentation algorithm and the diagnostic ensemble model are assessed in a real dataset constituted of 112 patients (women with high or intermediate risk).

The paper is structured as follows: Section 2 presents the related works in the breast cancer detection and diagnostic systems, and recent trends in convolutional ensemble methods. Section 3 first describes the image acquisition, pre-processing, ROI selection processes, and CNN model and a batch training algorithm based on resilient propagation inspired by [7]. It then presents the proposed mixture of CNN experts (ME-CNN), and a learning method for an ensemble of six-layer CNNs. Section 4 presents the experimental results of the proposed ensemble method and discusses its comparative performances with several existing methods. Section 5 gives the research conclusion.

2. Related Work

Several CAD systems utilizing different breast imaging techniques have been developed for the detection and diagnosis of breast masses. However, there are only a few publications on CAD systems using DCE-MRI, which are summarized in Table 1. In general, the existing approaches use either hand-crafted features [6, 8] or automatically-learned features [9]. For hand-crafted features, traditional steps are applied to extract features from the input image, and the focus is on feature engineering. For automatically learned features, the focus is on learning directly the mapping from the input image to the classification label.

Convolutional neural network (CNN) is one of the most powerful techniques in the automatic feature-learning approach. It has been used successfully for numerous applications, including object recognition and image classification [14], handwritten digit recognition [15], character recognition [16], face detection [17], speech recognition [18]. In medical image analysis area, CNNs have been applied for many tasks and imaging modalities, such as pulmonary nodules detection in chest CT scans [19], colonic polyps detection in CT colonography [20], nuclei detection and classification in histopathological images [21], brain lesion segmentation in MRI [22], and hemorrhage detection in color fundus images [23].

To improve the classification accuracy, ensemble methods have been developed, where several CNN outputs are fused using an additional network [24], or a softmax layer [25, 26]. Table

Table 1: Summary of recent computer-aided systems for breast DCE-MRI.

Author	Year	Database	Problem	Method	Results
Chang et al. [10]	2012	132 solid masses (63 benign, 69 malignant)	Diagnosis	Analysis of the characteristic kinetic curve of MR masses derived from fuzzy c-means clustering (FCM) + binary logistic regression classification.	Acc=86.36%, Se=85.51%, Sp=87.30%, AUC=0.9154
Huang et al. [8]	2013	95 solid masses (44 benign, 51 malignant), 82 patients	Diagnosis	3-D MR morphological features extracted using texture-based gray-level co-occurrence matrix (GLCM), tumor shape + ellipsoid fitting classification.	Acc=88.42%, Se=88.24%, Sp=88.64%, AUC=0.8926
Soares et al. [11]	2013	35 lesions (20 benign, 15 malignant)	Diagnosis	3D lacunarity multi-fractal analysis to characterize spatial complexity of DCE-MR breast tumors at multiple scales + SVM classifier.	AUC=0.96
Chang et al. [12]	2014	95 biopsy confirmed lesions (28 benign, 67 malignant), 54 patients	Detection	Motion registration, detection of focal tumor breast lesions using kinetic features extracted from pixel-based time-signal intensity curve (TIC), morphological features of detected lesions.	Detection rate of 92.63% of all tumor lesions, with 6.15 FP/case.
Hassanien et al. [13]	2014	25 MR images, 90 train samples, 46 test samples	Diagnosis	Hybrid approach combining type-II fuzzy sets, adaptive ant-based clustering, multilayer perceptron (MLP) classifier, statistical-based feature extraction	Acc=95.1%
Rasti et al. [9]	2015	30 MR images, 120 samples each for benign and malignant	Diagnosis	A semi-automatic algorithm for lesion segmentation and ROI selection ROI classification by CNNs with ave-pooling	Acc=98.7%
Gubern et al. [6]	2015	209 lesions: 114 benign, 105 malignant (55 mass, 50 non-mass)	Detection	Lesion detection using features of blobs and enhanced voxels, Estimation of malignancy score using RF classifier and region-based morphological and kinetic features.	Acc=89% (91% for mass-like, 86% for non-mass-like malignant lesions) at 4 FP/normal case.

Acc=accuracy; Se=sensitivity; Sp=specificity; AUC=area under the ROC curve (See Section 4.1).

2 summarizes the recent works on medical applications based on an ensemble of CNNs. More theoretical basis for mixture ensemble is given in [27].

Table 2: A summary of convolutional ensemble methods in medical applications.

Author	Year	Application	Ensemble Method	Notes
Ciresan et al. [28]	2013	Mitosis detection in breast cancer histology images	Output map averaging of 2 CNNs	Fusion of 2 independently trained CNNs
Wang et al. [29]	2014	Mitosis detection in breast cancer histology images	Cascaded ensemble of a learned CNN and a set of hand-crafted features + Random forests classifiers + a weighted average of the outputs	Features are extracted independently.
Brebisson et al. [30]	2015	Anatomical brain segmentation in MRI	8 simple CNN branches fused at fully connected layers according to 8 different inputs + 2 FC layers + a softmax layer output	The model is a deep neural network architecture with end-to-end learning for segmentation.
Song et al. [31]	2015	Cervical cytoplasm and nuclei segmentation	3 branches of CNNs fused at fully connected layers and a softmax layer output	Each branch has a different scaled input - a multi-scale CNN with end-to-end learning.
Maji et al. [32]	2016	Retinal vessels detection in fundus images	Output map averaging of 12 CNNs	Each CNN is trained independently on patches randomly selected from the training images.
Chen et al. [33]	2016	Mitosis detection in breast cancer histology images	Output map averaging of 3 CNNs	Each CNNs is a different CaffeNet [34] trained independently with transfer learning method.

In this article, we explore the capability of CNN and convolutional ensemble methods for breast cancer diagnosis in DCE-MRI. It is an extended version of our work in [9] to include a significantly larger database, an improved segmentation method, and a more accurate diagnostic classifier.

3. Material and Methods

In this section, we first describe the image data acquisition for our DCE-MRI dataset, and present several pre-processing techniques to segment and select the region of interest from an input breast MR image. We then describe the architecture and mathematical model of convolutional neural network. Finally, we present the proposed mixture ensemble of CNNs for the classification of benign versus malignant breast tumors.

3.1. Image Data Acquisition

The data used in this work were collected at the Imaging Center of Milad Hospital in Tehran. The screening protocol was a 1.5 Tesla Siemens Healthcare MRI system, which has a dedicated

bipolar phased-array breast coil. This provided T1-weighted axial image data over the whole breast, obtained with a radio frequency spoiled gradient-recalled sequence with the image size of 512×512 pixels. Imaging was performed prior to and subsequent to a bolus injection of 0.2 mmol/kg of Gd-DTPA in both manual and intravenous ways. The injection was followed by the 15 cc normal saline. Then, 26 breast MR image acquisitions in each case, including one pre-contrast and other post-contrast series, were acquired. After agent injection, the first acquisition was at the 90th second, and subsequent acquisitions were obtained every 100 seconds approximately. Each axial acquisition series included 128 axial slices on average.

Overall, 112 DCE-MRI breast examinations from high- or intermediate-risk patients were obtained. These examinations were histopathologically proven to be malignant (53 cases) or benign (59 cases). The patients' age ranged from 37 to 71 years (the mean age: 48 years). The effective lesion radius had a range of $3.8\text{-}31.4 \text{ mm}$ and an average of 12.3 mm .

3.2. Pre-processing, Segmentation, and ROI Selection

A set of breast axial slices was selected and annotated by an expert radiologist. The set presented the best appearance of the lesions among other image slices. The slices were labeled according to patients' IDs. For the first selected post-contrast slices, the subtraction image set was obtained by subtracting the corresponding pre-contrast ones.

The general steps of the automatic ROI selection algorithm are shown in Fig. 1. The first step was background reduction via first post-contrast subtraction, followed by contrast enhancement and breast regions cropping. The aim of breast regions cropping step was to locate approximately the breast tissues and reduce the disturbing effects of other anatomical structures such as the chest. The size and position of the rectangular cropping was determined empirically based on the image acquisition settings.

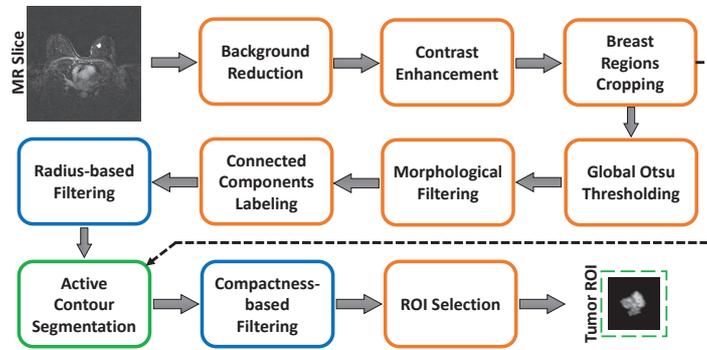


Figure 1: The main steps for ROI selection in breast DCE-MRI.

Next, the global Otsu thresholding [35] and the morphological top-hat filtering [36] were performed to remove non-lesion structures. After connected components labeling, pseudo lesions were isolated for further processing. Radius-based filtering was then applied to remove regions with radius outside the range $[r_{\min}, r_{\max}]$. In our case, $r_{\min} = 3.8 \text{ mm}$, and $r_{\max} = 31.4 \text{ mm}$.

Then, localized active contour (LAC) segmentation was applied to each remaining lesion of interest. Here, we used the Chan-Vese active contour model [37], with parameters with $Max\text{-}iter = 60$, $Mu = 0.1$, and $Sigma = 4$. The LAC segmentation step was useful because it recovered tumor pixels that were removed by morphological operations.

To reduce false positives, compactness-based filtering was applied next. The *compactness* C of a region is defined as $C = \frac{s_x s_y}{\max(s_x, s_y)^2}$, where s_x and s_y are width and height of the bounding box, respectively [38]. A region was removed if its compactness C was less than 0.2. Fig. 2 shows example outputs of the preprocessing, segmentation and ROI selection, described in this section. Overall, 562 ROIs (244 malignant and 318 benign) were extracted with the ROI selection algorithm over the whole set slices.

To conclude Section 3.2, the techniques presented here are effective for the real DCE-MRI dataset we have to deal with. However, they can be made more general and less ad-hoc by devising methods to determine the thresholds adaptively.

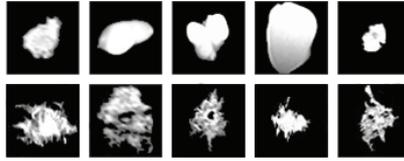


Figure 2: Example of automatic MR tumor segmentation and ROI generation. *Top row:* benign tumors; *Bottom row:* malignant tumors.

3.3. Convolutional Neural Network

A brief description of convolutional neural network is warranted because it is a key tool in this paper. CNN, originally proposed by LeCun [39], is a neural network model with three main architectural ideas: local receptive fields, weight sharing, and sub-sampling in the spatial domain. The network was initially designed for the recognition of two-dimensional image patterns, and has become one of the major architectures in deep learning [40]. CNN has many strengths. First, feature extraction and classification are integrated into one structure and are fully adaptive. Second, the network extracts 2-D image features at increasing dyadic scales. Third, it is relatively invariant to image noise and local geometric distortions.

A CNN consists of three main types of layers: (i) convolution layers, (ii) sub-sampling (or max-pooling) layers, and (iii) an output layer. Network layers are arranged in a feed-forward structure: each convolution layer is followed by a sub-sampling layer, and the last convolution layer is followed by the output layer. The convolution and sub-sampling layers are considered as 2-D layers, whereas the output layer is considered as a 1-D layer. In CNN, each 2-D layer has several planes. A plane consists of neurons that are arranged in a 2-D array. The output of a plane is called a feature map.

Mathematical model: Consider a CNN shown in Fig. 3 with 6 layers: convolution C1, sub-sampling S2, convolution C3, subsampling S4, convolution C5, and output O6. For convolutional layer l (where $l = 1, 3, 5, \dots$), let f_l be the activation function of layer l . Let $W_{m,n}^l$ be the 2-D convolution kernel for the connection from feature map m in layer $(l-1)$ to feature map n in layer l . Let b_n^l be the bias term associated with feature map n . Let p_n^l denote the set of all planes in the layer $(l-1)$ that are connected to feature map n . The output feature map n of convolution layer l is defined as

$$o_n^l = f_l \left(\sum_{m \in p_n^l} o_m^{l-1} \otimes W_{m,n}^l + b_n^l \right) , \quad (1)$$

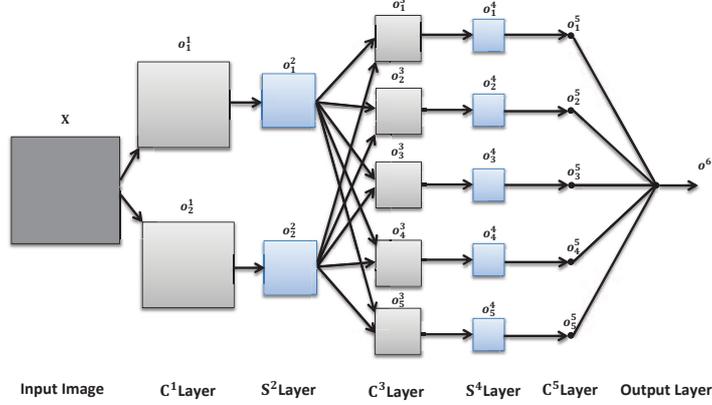


Figure 3: A convolutional neural network with 6 layers for processing input image of size 32×32 pixels. From left to right, the number of feature maps is 2, 2, 5, 5, 1, and the size of weight kernels is 5×5 , 1×1 , 3×3 , 1×1 , 6×6 , 1×1 .

where \otimes denotes the 2-D convolution, and o_m^{l-1} is output feature map of layer $l-1$.

For the sub-sampling layer l (where $l = 2, 4, \dots$), output feature map n is defined as

$$o_n^l = f_l(w_n^l \mathcal{P}(o_n^{l-1}) + b_n^l), \quad (2)$$

where w_n^l and b_n^l are the scalar weights, and the bias term. Here, $\mathcal{P}(\cdot)$ denote the pooling operation (e.g. max-pooling or average pooling).

For output layer l , each neuron is fully connected to all feature maps of the previous layer. The response of an output neuron is given as

$$o^l = f_l \left(\sum_{m \in p^l} o_m^{l-1} w_m^l + b^l \right). \quad (3)$$

Here, p^l denotes the set of feature maps in layer C^{l-1} that are connected to the output neuron.

CNN training algorithm: In this paper, we use a batch training algorithm based on the resilient back-propagation or RPROP, proposed in [41]. In the RPROP algorithm, the learning step is increased or decreased, depending only on the sign of the error gradient. This algorithm has been shown to converge faster compared to the standard gradient descent algorithm [7]. RPROP works well even when the gradient has very small magnitudes. Furthermore, it is computationally efficient because only the first-order derivative of the error function is required. Detail implementation of RPROP training algorithm for CNN is given in our technical report [42].

3.4. Proposed Mixture Ensemble of Convolutional Experts

We propose a mixture ensemble of convolutional neural networks (or ME-CNN) for classification of an ROI image as benign or malignant. An overview of the mixture ensemble model is shown in Fig. 4. There are L experts and one gating network, which share the same input. Each expert could be specialized in one region of the high-dimensional input space. The gating network is trained to produce input-adaptive weights that are used to fuse the outputs of the experts. In the

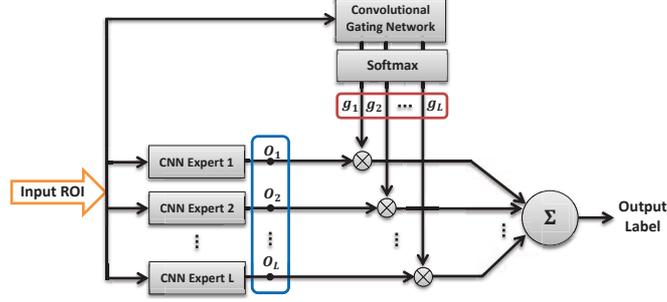


Figure 4: Schematic diagram of the proposed mixture ensemble of convolutional experts.

proposed model, CNNs are used as the experts and the gating network. Note that all components (experts and gating network) are trained simultaneously via an end-to-end optimization process.

Mathematical formulation: For a given input sample (ROI), let O_1, O_2, \dots, O_L be the outputs of L convolutional experts. Each output depends on the input sample, the network adjustable parameters (weights and biases), and the activation function. The convolutional gating network (CGN) is trained to produce a weight g_i for each expert output O_i , where $i = 1, 2, \dots, L$. Let G_1, G_2, \dots, G_L be the non-probabilistic outputs of the last layer of the CGN. The softmax function is applied on these non-probabilistic outputs $\{G_i\}$ to obtain the fusion weights $\{g_i\}$:

$$g_i = \frac{\exp(G_i)}{\sum_{j=1}^L \exp(G_j)}. \quad (4)$$

Clearly, the fusion weights satisfy the condition:

$$\sum_{i=1}^L g_i = 1. \quad (5)$$

The final output of the mixture ensemble of convolutional experts is given as

$$O = \sum_{i=1}^L g_i O_i. \quad (6)$$

ME-CNN training: The use of CNNs in the mixture ensemble requires significant modification to the training algorithm. As mentioned previously, CNN experts and convolutional gating network are trained simultaneously in our approach. Consider a training set of K samples, where \mathbf{x}_k is the k -th input sample and \mathbf{d}^k the desired output vector.

For the CNN experts, we define, similarly to [27], their error function for the k -th input sample as

$$E_{\text{expert}}^k = -\ln \left(\sum_{i=1}^L g_i^k \exp \left(-\frac{1}{2} \|\mathbf{d}^k - \mathbf{O}_i^k\|^2 \right) \right), \quad (7)$$

where \mathbf{O}_i^k is the output vector of CNN expert i . For CNN expert i , the effective error signal for all training samples (a component in the derivative of E_{expert}^k) can be written as

$$e_i = \sum_{k=1}^K h_i^k (\mathbf{d}^k - \mathbf{O}_i^k), \quad (8)$$

where

$$h_i^k = \frac{g_i^k \exp\left(-\frac{1}{2} \|\mathbf{d}^k - \mathbf{O}_i^k\|^2\right)}{\sum_{j=1}^L g_j^k \exp\left(-\frac{1}{2} \|\mathbf{d}^k - \mathbf{O}_j^k\|^2\right)}. \quad (9)$$

Here, h_i^k can be considered as a posterior-probability estimate produced by expert i for input sample \mathbf{x}_k . Conceptually, it means that h_i^k is high if: 1) output \mathbf{O}_i^k of expert i for input sample \mathbf{x}_k is close to the desired output \mathbf{d}^k ; and 2) weight g_i^k given by the CGN to expert i is high.

For the convolutional gating network, we define its total error function as

$$E_{\text{gate}} = \frac{1}{2} \sum_{k=1}^K \|\mathbf{h}^k - \mathbf{g}^k\|^2, \quad (10)$$

where $\mathbf{h}^k \triangleq [h_1^k, h_2^k, \dots, h_L^k]^T$ is a vector of the experts' posterior probability estimates, and $\mathbf{g}^k \triangleq [g_1^k, g_2^k, \dots, g_L^k]^T$ is the CGN output vector.

Once the error signal for CNN experts is defined in (8) and the error function for the CGN is defined in (10), we can compute the gradient w.r.t the free parameters of the networks via the error back-propagation algorithm, in a similar way as [42]. Subsequently, the RPROP algorithm can be used to determine simultaneously the free parameters of the CNNs and CGN. In this way, each CNN expert is trained to specialize in a region of the high-dimensional input space, whereas the convolutional gating network is trained to produce optimized fusion weights g_i .

4. Results and Discussion

In this section, we first describe the performance measurements and the experimental method (Subsection 4.1). We then analyze the classification performance of the CNN in comparison with three other single-classifier methods (Subsection 4.2). Finally, we discuss the performance of the proposed ME-CNN method in relation to two other convolutional ensemble methods in Subsection 4.3.

4.1. Performance Measures and Experimental Method

Classification performance in this binary problem (benign versus malignant breast tumor) was assessed based on the following basic measures:

- True positives (TP) is the number of test samples in the malignant class that are correctly classified.
- True negatives (TN) is the number of test samples in the benign class that are correctly classified.

- False positives (FP) is the number of test samples in the benign class that are incorrectly classified.
- False negatives (FN) is the number of test samples in the malignant class that are incorrectly classified.

Subsequently, other measures including *accuracy* (Acc), *sensitivity* (Se), *specificity* (Sp), *positive predictive value* (PPV), *negative predictive value* (NPV) are computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

$$Sensitivity = \frac{TP}{TP + FN}, \quad (12)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (13)$$

$$PPV = \frac{TP}{TP + FP}, \quad (14)$$

$$NPV = \frac{TN}{TN + FN}, \quad (15)$$

Other performance indicators include the *Receiver Operating Characteristic* (ROC) curve and *Area Under the ROC Curve* (AUC).

In addition to the above measures, statistical significance of performance difference between two methods was determined by the bootstrap method. For this purpose, 10,000 bootstraps were applied to compute the 95% confidence interval for the AUC value [43, 44]. The p -values of pairwise comparison of ROC curves were calculated. If the p -value is less than 0.05, the conclusion is that the two methods have statistically different performances.

To compute classification measures, 5-fold cross-validation was applied at the patient level. The data set was divided into five partitions of approximately equal sizes. Four partitions were used for training and the remaining partition was used for testing. This step was repeated five times until all different choices for the test set were evaluated. The classification measures were averaged over the five folds. Furthermore, to increase the amount of training data, each sample in the training partitions was rotated by 90° , 180° , and 270° , and flipped horizontally to give 4 additional training samples.

4.2. Comparative Evaluation of the CNN

In this experiment, we analyzed the performance of CNN in classifying benign versus malignant breast MR tumors, and compared it with other feature extraction and classification methods. Following this purpose, for all extracted ROIs, 10 scalar features were extracted according to [3], [13], [45] and [46]. They included five shape features (circulatory, spiculation, roughness, NRL-mean, and NRL-entropy) and five texture features (angular second moment (ASM), correlation, sum average, sum variance and sum entropy).

In the benchmark analysis, we investigated four different classifiers: multilayer perceptron, support vector machine, random forests classifier, convolutional neural network.

- **Multi-layer perceptron** (MLP): Several fully-connected MLPs with one hidden layer and sigmoid neurons were considered. The number of neurons in a hidden layer were varied

from 10 to 30 in a step of 2. The number of neurons in the output layer was 2. Four different training algorithms were applied: gradient descent with momentum (GDM), resilient propagation (RPROP), scaled conjugate gradient (SCG), and Levenberg-Marquardt (LM). The training parameters are: $max\ epoch = 100$, and $MSE\ goal = 10^{-5}$.

- **Support vector machine (SVM)**: Radial basis function was used as the kernel; the kernel coefficient was set to $\gamma = 0.1$. Penalty parameter C was set to 2^i , with i varied between -4 and 4 with a step of 2 .
- **Random forests (RF)**: The maximum number of trees was varied among 50 , 100 , 200 , and 500 . The max-depth of the tree was equal to the number of features ($n = 10$) [47]. RF classifier is a state-of-the-art method for breast cancer detection in DCR-MRI [6].
- **Convolutional neural network (CNN)**: The CNN was applied to process directly each ROI of size 32×32 pixels. We evaluated five different CNNs, each having six active layers. The activation function for C-layers was the *sigmoid*, which was proposed by LeCun [48]: $f(x) = 1.7159 \tanh(2x/3)$. The activation function for S-layers was the linear function. The activation function for the output layer was the *tanh*. Three sub-sampling methods were considered for 2×2 pixel blocks: ave-pooling, sum-pooling, and max-pooling. Tables 3 and 4 summarize the structural information of these networks: feature map sizes, kernel sizes, number of feature maps per layer, connection types, and the total number of free parameters.

Training was done in batch mode, and the batch size varied from 32 , 64 , 128 , 256 samples to the full-size (100% of the training samples). The RPROP algorithm [41] was used for training the CNNs, where the step size for weight update was dynamically changed by an increment factor η^+ or a decrement factor η^- , depending on the sign of the gradient, see [7]. Different values for η^+ and η^- were evaluated on the training set in order to find suitable ones: η^+ from 1.01 to 1.10 , and η^- from 0.90 to 0.99 with a step of 0.01 . A good property of the RPROP algorithm is that it works for many values of η^+ and η^- , as long as they are close to 1 . The maximum number of training epochs is $max_epoch = 300$, and the training target is $MSE\ goal = 10^{-2}$.

We implemented the CNNs in MATLAB; this library is available online¹. The experiment was run on a computer with Core-i7 CPU at 4GHz (AMD FX-8350: 4.7M), 16 GB of RAM, and Windows-7 64-bit operating system. Training CNN5 took 3.8s per ROI on average. Testing CNN5 took 0.003s per ROI on average (classification time).

Table 3: Size of feature maps and kernels for the implemented CNNs with 6 active layers.

Layer	C1	S2	C3	S4	C5	O6
Feature map size	28×28	14×14	12×12	6×6	1×1	1
Kernel size	5×5	2×2	3×3	2×2	6×6	1

Table 5 shows comparative performances of the individual MLP, SVM, RF and CNN classifiers. A number of observations can be drawn. First, if only hand-crafted features were used,

¹URL: www.uow.edu.au/phung/docs/cnn-matlab.zip

Table 4: Other structural information of the implemented CNNs.

Structure	C1 FM	C3 FM	C5 FM	O6 FM	C3 connections	C1, S2, S4, C5 connections	Free parameters
CNN1	2	2	2	1	full	1-to-1	175
CNN2	2	4	4	1	full	1-to-1	293
CNN3	3	6	6	1	full	1-to-1	493
CNN4	4	8	8	1	full	1-to-1	729
CNN5	5	10	10	1	full	1-to-1	1001

FM: number of feature maps

Table 5: Classification performance of individual MLP, SVM, RF and CNN classifier.

Classifier	Configuration	Classification Measures						
		Acc (%)	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC	MSE
MLP	LM training, 26 hidden neurons, $\mu=0.0015$ $\mu^+ = 10, \mu^- = 0.1,$	87.95	88.64	87.18	88.64	87.18	0.914	0.337
SVM	Kernel: RBF, $C = 4, \gamma = 0.1$	89.16	86.36	92.31	92.68	85.71	0.930	0.318
RF	Trees=100, depth=10	90.36	93.18	87.18	89.13	91.89	0.937	0.288
CNN	CNN5, max-pooling full-batch RPROP, $\eta^+ = 1.02, \eta^- = 0.98$	92.77	93.18	92.31	93.18	92.31	0.954	0.272

Acc=accuracy; Se=sensitivity; Sp=specificity; PPV=positive predictive value; NPV=negative predictive value.

the random forests classifier performed better than the SVM and the MLP. Second, the CNN classifier (CNN5 with 1001 free parameters) achieved a classification rate of 92.77%, which is statistically higher than the RF, SVM, and MLP classifiers. Using the bootstrapping method, the p -value between the ROC curves of the RF and CNN classifier is 0.041 ($p < 0.05$). This result show that the features automatically learnt by the CNN are more discriminative than the existing hand-crafted features.

The classification results also indicated that the pooling method and the numbers of 2-D feature maps in C1 and C3 layers played an important role in the diagnostic performance. Max pooling performed better than average pooling or sum pooling in this classification problem (benign-versus-malignant breast tumor). For fixed training data, an increase in the number of feature maps leads to an increase in the number of free parameters in the network and causes overfitting. At present, there is no theoretical framework to decide the optimal number of feature maps. In our diagnostic problem, this decision was made based on empirical experimentation.

4.3. Comparative Evaluation of the Proposed ME-CNN

In this experiment, we evaluated the classification performance of the proposed ME-CNN. The simple CNN1 structure, described in Tables 3 and 4, was used as CNN expert. Max-pooling was selected for sub-sampling the convolutional feature maps. The number of CNN experts were varied between 2, 3, 4, and 5. A convolutional gating network (CGN) as shown in Table 6 was used. Note that the number of output neurons in the CGN is equal to the number of CNN experts. The CGN used the *sigmoid* activation function for all layers.

For training the ME-CNN, we used the RPROP algorithm in the full batch mode. The MSE goal was 0.01, and the maximum number of epochs was 300. Using a grid search, the following learning parameters were selected: i) $\eta_{exp}^+ = 1.02$, and $\eta_{exp}^- = 0.98$ for CNN experts; ii) $\eta_g^+ =$

Table 6: Structural information of the convolutional gating network (CGN).

Structure	C1 FM	C3 FM	C5 FM	O6 FM	C3 connections	C1,S2,S4,C5 connections	Total parameters
CGN	2	2	2	2, 3, 4, or 5	full	1-to-1	178, 181, 184, or 187

FM: number of feature maps

1.0003, and $\eta_g^- = 0.9997$ for the CGN. Figure 5 shows the 2-D feature maps that were extracted by the C1 layer of different modules (CNN experts and CGN) in a ME-CNN after training.

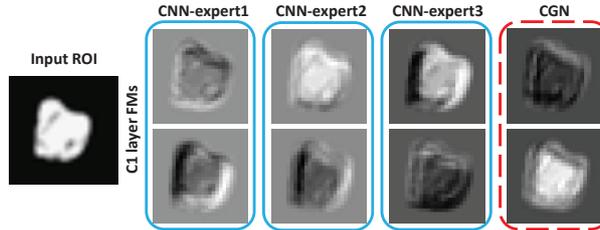


Figure 5: 2-D feature maps by layer C1 of different modules in the ME-CNN.

To gain insights on the comparative performances of the proposed ME-CNN, we also evaluated two other convolutional ensemble methods: ave-ensemble and soft-ensemble.

- **Ave-ensemble** [28, 32, 33]: This method trains several CNNs from the available ground-truth labels. For a given test image, the outputs of the CNNs are averaged to generate the final output of the ensemble. We experimented with ensembles of 2, 3, 4, and 5 CNNs with structures as shown in Table 4.
- **Soft-ensemble** [30, 31]: This method uses several parallel CNN branches. At the end of each convolutional branch, feature maps are concatenated. All features are then processed by a fully-connected layer, followed by a softmax layer for final output. In other words, this method does not use a convolutional gating network for fusion. In our experiment, the number of CNN experts were varied between 2, 3, 4, and 5.

Note that for both ave-ensemble and soft-ensemble methods, the standard MSE cost function was used for training. The size of 2-D feature maps and kernels was the same as in Tables 3.

Computation Efficiency: Training the ME-CNN took 4.6s per ROI on average. In comparison, training the Ave-ensemble and the Soft-ensemble took on average 7.9s and 4.2s per ROI, respectively. Testing the ME-CNN took 0.004s per ROI on average. Testing the Ave-ensemble and the Soft-ensemble took on average 0.007s and 0.004s per ROI, respectively.

Classification Accuracy: Table 7 shows the classification performances, estimated via the 5-fold cross validation, of three methods: ME-CNN, ave-ensemble, and soft-ensemble. To provide a baseline for comparison, the classification measures for the single-expert methods (CNN5 and RF) are also included in the table. The diagnostic accuracy for different ensemble methods were: ME-CNN = 96.39%, Soft-ensemble = 95.18%, Ave-ensemble = 93.98%.

Table 7: Classification performances of different ensemble methods.

Ensemble method	Configuration	Parameters	Classification Measures						
			Acc (%)	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC	MSE
Ave-ensemble	fusion of CNN4 and CNN5	1750	93.98	95.46	92.31	93.33	94.38	0.969	0.198
Soft-ensemble	3 branches of CNN1 with a softmax output layer	1046	95.18	95.45	94.87	95.45	94.87	0.976	0.169
Proposed ME-CNN	3 CNN1 experts + CGN	703	96.39	97.73	94.87	95.56	97.37	0.991	0.152
None	Single CNN5	1001	92.77	93.18	92.31	93.18	92.31	0.954	0.272
None	Single RF	-	90.36	93.18	87.18	89.13	91.89	0.937	0.288

Acc=accuracy; *Se*=sensitivity; *Sp*=specificity; *PPV*=positive predictive value; *NPV*=negative predictive value.

Table 8: Pair-wise p -values obtained via bootstrapping for comparison of 5 methods.

Method	RF	CNN5	Ave-ensemble	Soft-ensemble	ME-CNN
RF	-	0.041	0.011	0.008	0.003
CNN5	0.041	-	0.092	0.043	0.039
Ave-ensemble	0.011	0.092	-	0.188	0.041
Soft-ensemble	0.008	0.043	0.188	-	0.064
ME-CNN	0.003	0.039	0.041	0.064	-

Table 8 shows the p -values, which were obtained via bootstrapping, for pair-wise comparison between the RF, CNN5, Ave-ensemble, Soft-ensemble, and ME-CNN. By combining Tables 7 and 8, we can conclude that the ME-CNN had significantly higher accuracy than the Ave-ensemble and the single-classifier methods (CNN5 and RF); the corresponding pair-wise p -values were all less than 0.05 (for 95% confidence level). Furthermore, the ME-CNN has a statistically similar accuracy as the Soft-ensemble; the p -value between the two methods was 0.064 (greater than 0.05).

Roles of CNN experts and CGN: The above improvement in classification accuracy of the proposed ME-CNN can be attributed to the fact that CNN experts are trained simultaneously in a competitive manner. The convolutional gating network is trained to play the adaptive weighting role. Over time, the CGN can partition the high-dimensional input space into subspaces depending on individual expert’s performance.

To analyze the CNN experts’ ability to partition the input space, we computed disagreement factors and the correlation coefficients. For a pair of CNN experts, the *disagreement factor* is the ratio between the number of samples on which they disagree versus the total number of samples [49]. The *correlation coefficient* is computed over the two sets of real outputs that the two CNNs produced for the test set. For the trained ME-CNN in our experiment, the average disagreement factor between the three experts was relatively high (0.61), and the average correlation coefficient was very low (0.03).

ME-CNN versus Soft-ensemble: Although the classification accuracy of the ME-CNN and Soft-ensemble was similar, the ME-CNN had several advantages. Firstly, the ME-CNN used significantly fewer free parameters than the Soft-ensemble (703 versus 1046 parameters), and therefore was less prone to overfitting. This is useful for medical diagnosis applications where the number of training samples is typically small, due to the limited number of patients. Secondly, the ME-CNN took fewer training iterations (epochs) converge than the Soft-ensemble, as

shown in Fig. 6(a). Thirdly, in our experiment as the number of experts or branches increased, the classification accuracy of the ME-CNN did not deteriorate as quickly as the Soft-ensemble, see Fig. 6(b). At present, there is no analytical method to determine the optimum number of convolutional experts. For practical applications, a possible method for estimating a suitable number of experts is through empirical evaluation (e.g. cross-validation) on the training data.

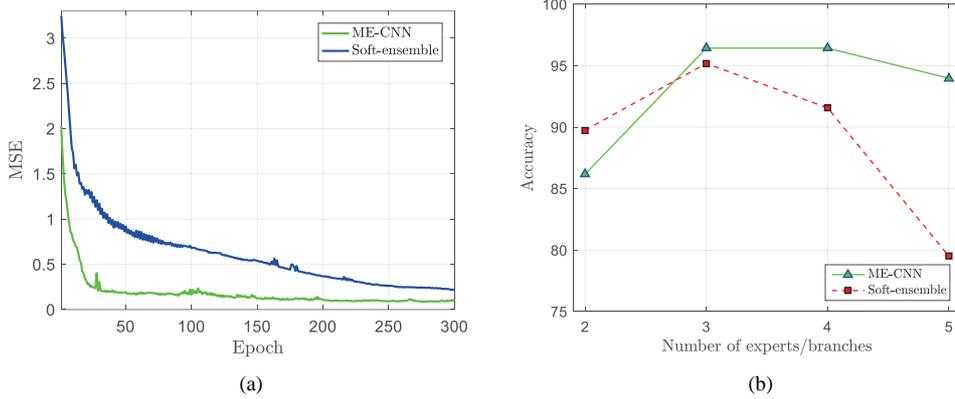


Figure 6: Comparison of the ME-CNN and Soft-ensemble: (a) Average training MSE versus training epoch; (b) Classification rate on the test sets versus the number of experts/branches.

5. Conclusion

This article presents a CAD system for the diagnosis of breast cancer in DCE-MRI. For the localization stage, we present an automatic segmentation method based on local active contours for finding the region of interest from a breast MR image. For the diagnosis stage, we propose a new model for mixture ensemble of convolutional neural networks, called ME-CNN, for classification of benign versus malignant breast tumors. In the proposed model, the experts and gating network, all based on the CNNs, are trained simultaneously in an end-to-end optimization approach. Experimental results on a dataset of 112 DCE-MRI cases have shown that the proposed ME-CNN achieves competitive classification performances, compared with existing classifiers and ensemble methods. It also has the advantages of fast execution time in both training and testing, and a compact structure with small number of free parameters. By expanding the dataset to include non-mass lesions and more cases (patients), the proposed diagnostic approach has the potential to support radiologists in breast MR data analysis. Among several promising directions, one could extend the ME-CNN approach to the pre-processing stage, by combining with recent advances in fully convolutional networks for semantic segmentation.

Acknowledgment

The authors would like to thank the anonymous reviewers and Dr. Hossein Rabbani for their constructive comments, which helped improving the paper.

References

- [1] C. E. DeSantis, S. A. Fedewa, A. Goding Sauer, J. L. Kramer, R. A. Smith, A. Jemal, Breast cancer statistics, 2015: Convergence of incidence rates between black and white women, *CA: a cancer journal for clinicians*.
- [2] D. B. Kopans, Breast imaging, 3rd Edition, Boston, Massachusetts: Lippincott Williams and Wilkins: Harvard Medical School, 2007.
- [3] K. Nie, J.-H. Chen, J. Y. Hon, Y. Chu, O. Nalcioglu, M.-Y. Su, Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI, *Academic Radiology* 15 (12) (2008) 1513–1525.
- [4] E. Warner, D. Plewes, R. Shumak, G. Catzavelos, L. Di Prospero, M. Yaffe, V. Goel, E. Ramsay, P. Chart, D. Cole, et al., Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer, *Journal of Clinical Oncology* 19 (15) (2001) 3524–3531.
- [5] K. Yamaguchi, D. Schacht, G. M. Newstead, A. R. Bradbury, M. S. Verp, O. I. Olopade, H. Abe, Breast cancer detected on an incident (second or subsequent) round of screening MRI: MRI features of false-negative cases, *American Journal of Roentgenology* 201 (5) (2013) 1155–1163.
- [6] A. Gubern-Mérida, R. Martí, J. Melendez, J. L. Hauth, R. M. Mann, N. Karssemeijer, B. Platel, Automated localization of breast cancer in DCE-MRI, *Medical Image Analysis* 20 (1) (2015) 265–274.
- [7] S. L. Phung, A. Bouzerdoum, A pyramidal neural network for visual pattern recognition, *IEEE Transactions on Neural Networks* 18 (2) (2007) 329–343.
- [8] Y.-H. Huang, Y.-C. Chang, C.-S. Huang, T.-J. Wu, J.-H. Chen, R.-F. Chang, 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) (2013) 508–517.
- [9] R. Rasti, M. Teshnehlab, R. Jafari, A CAD system for identification and classification of breast cancer tumors in DCE-MR images based on hierarchical convolutional neural networks, *Computational Intelligence in Electrical Engineering* 6 (1) (2015) 1–14.
- [10] Y.-C. Chang, Y.-H. Huang, C.-S. Huang, P.-K. Chang, J.-H. Chen, R.-F. Chang, Classification of breast mass lesions using model-based analysis of the characteristic kinetic curve derived from fuzzy c-means clustering, *Magnetic Resonance Imaging* 30 (3) (2012) 312–322.
- [11] F. Soares, F. Janela, M. Pereira, J. Seabra, M. M. Freire, 3D lacunarity in multifractal analysis of breast tumor lesions in dynamic contrast-enhanced magnetic resonance imaging, *IEEE Transactions on Image Processing* 22 (11) (2013) 4422–4435.
- [12] Y.-C. Chang, Y.-H. Huang, C.-S. Huang, J.-H. Chen, R.-F. Chang, Computerized breast lesions detection using kinetic and morphologic analysis for dynamic contrast-enhanced MRI, *Magnetic Resonance Imaging* 32 (5) (2014) 514–522.
- [13] A. E. Hassanien, H. M. Moftah, A. T. Azar, M. Shoman, MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier, *Applied Soft Computing* 14 (2014) 62–71.
- [14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [15] X.-X. Niu, C. Y. Suen, A novel hybrid CNN–SVM classifier for recognizing handwritten digits, *Pattern Recognition* 45 (4) (2012) 1318–1325.
- [16] W. Yang, L. Jin, D. Tao, Z. Xie, Z. Feng, Dropsample: A new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition, *Pattern Recognition* 58 (2016) 190–203.
- [17] C. Garcia, M. Delakis, Convolutional face finder: A neural architecture for fast and robust face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (11) (2004) 1408–1423.
- [18] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing* 22 (10) (2014) 1533–1545.
- [19] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, B. van Ginneken, Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1160–1169.
- [20] H. Roth, et al., Improving computer-aided detection using convolutional neural networks and random view aggregation, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1170–1181.
- [21] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, N. M. Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1196–1206.
- [22] S. Pereira, A. Pinto, V. Alves, C. A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1240–1251.
- [23] M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, C. I. Sánchez, Fast convolutional neural net-

- work training using selective data sampling: Application to hemorrhage detection in color fundus images, *IEEE Transactions on Medical Imaging* 35 (5) (2016) 1273–1284.
- [24] M. Oberweger, P. Wohlhart, V. Lepetit, Training a feedback loop for hand pose estimation, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3316–3324.
- [25] J. Guo, S. Gould, Deep CNN ensemble with data augmentation for object detection, *arXiv preprint arXiv:1506.07224*.
- [26] G. Antipov, S.-A. Berrani, J.-L. Dugelay, Minimalistic CNN-based ensemble model for gender prediction from face images, *Pattern Recognition Letters* 70 (2016) 59–65.
- [27] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton, Adaptive mixtures of local experts, *Neural Computation* 3 (1) (1991) 79–87.
- [28] D. C. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, 2013, pp. 411–418.
- [29] H. Wang, A. Cruz-Roa, A. Basavanahally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, A. Madabhushi, Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection, in: *SPIE Medical Imaging, International Society for Optics and Photonics*, 2014, pp. 90410B–90410B.
- [30] A. de Brebisson, G. Montana, Deep neural networks for anatomical brain segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 20–28.
- [31] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, T. Wang, Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning, *IEEE Transactions on Biomedical Engineering* 62 (10) (2015) 2421–2433.
- [32] D. Maji, A. Santara, P. Mitra, D. Sheet, Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images, *arXiv preprint arXiv:1603.04833*.
- [33] H. Chen, Q. Dou, X. Wang, J. Qin, P.-A. Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI Press, 2016, pp. 1160–1166.
- [34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, 2014, pp. 675–678.
- [35] N. Otsu, A threshold selection method from gray-level histograms, *Automatica* 11 (285-296) (1975) 23–27.
- [36] E. R. Dougherty, R. A. Lotufo, *Hands-on morphological image processing*, Vol. 59, SPIE press, 2003.
- [37] T. F. Chan, L. A. Vese, Active contours without edges, *IEEE Transactions on Image Processing* 10 (2) (2001) 266–277.
- [38] H. Han, L. Li, F. Han, B. Song, W. Moore, Z. Liang, Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme, *IEEE Journal of Biomedical and Health Informatics* 19 (2) (2015) 648–659.
- [39] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86 (11) (1998) 2278–2324.
- [40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [41] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in: *International Conference on Neural Networks*, IEEE, 1993, pp. 586–591.
- [42] S. L. Phung, A. Bouzerdoum, Matlab library for convolutional neural networks, Tech. rep., University of Wollongong, Technical Report, <http://www.uow.edu.au/~phung/download.html> (2009).
- [43] F. W. Samuelson, N. Petrick, Comparing image detection algorithms using resampling, in: *IEEE International Symposium on Biomedical Imaging*, IEEE, 2006, pp. 1312–1315.
- [44] F. W. Samuelson, N. Petrick, S. Paquerault, Advantages and examples of resampling for CAD evaluation, in: *IEEE International Symposium on Biomedical Imaging*, IEEE, 2007, pp. 492–495.
- [45] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, J. M. Reinhardt, 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) (2007) 89–95.
- [46] R. M. Haralick, K. Shanmugam, I. H. Dinstein, Textural features for image classification, *IEEE Transactions on Systems, Man and Cybernetics* (6) (1973) 610–621.
- [47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [48] Y. A. LeCun, L. Bottou, G. B. Orr, K.-R. Müller, Efficient backprop, in: *Neural Networks: Tricks of the trade*, Springer, 2012, pp. 9–48.
- [49] M. Javadi, S. A. A. Arani, A. Sajedin, R. Ebrahimpour, Classification of ECG arrhythmia by a modular neural network based on mixture of experts and negatively correlated learning, *Biomedical Signal Processing and Control*

8 (3) (2013) 289–296.