ResNet-SCDA-50 for Breast Abnormality Classification

Xiang Yu, Cheng Kang, David S. Guttery [ID], Seifedine Kadry [ID], Yang Chen [ID], and Yu-Dong Zhang [ID]

Abstract—*(Aim)* Breast cancer is the most common cancer in women and the second most common cancer worldwide. With the rapid advancement of deep learning, the early stages of breast cancer development can be accurately detected by radiologists with the help of artificial intelligence systems. *(Method)* Based on mammographic imaging, a mainstream clinical breast screening technique, we present a diagnostic system for accurate classification of breast abnormalities based on ResNet-50. To improve the proposed model, we created a new data augmentation framework called SCDA (Scaling and Contrast limited adaptive histogram equalization Data Augmentation). In its procedure, we first conduct the scaling operation to the original training set, followed by applying contrast limited adaptive histogram equalisation (CLAHE) to the scaled training set. By stacking the training set after SCDA with the original training set, we formed a new training set. The network trained by the augmented training set, was coined as ResNet-SCDA-50. Our system, which aims at a binary classification on mammographic images acquired from INbreast and MINI-MIAS, classifies masses, microcalcification as "abnormal", while normal regions are classified as "normal". *(Results)* We present the first attempt to use the image contrast enhancement method as the data augmentation method, resulting in an averaged 98.55 percent specificity and 92.83 percent sensitivity, which gives our best model an overall accuracy of 95.74 percent. *(Conclusion)* Our proposed method is effective in classifying breast abnormality.

Index Terms—Breast cancer, ResNet-50, contrast limited adaptive histogram equalization, classification

---

1 INTRODUCTION

As one of the most aggressive cancers worldwide, breast cancer caused more than 2 million new cases in 2018 [1], [2]. Specifically, the incidence of breast cancer in the UK had risen to 46,109 in 2017, accounting for 15.1 percent of all cancer cases and was the most common cancer diagnosed in that year [3]. It is widely recognized that prevention and early diagnosis significantly reduce cancer mortality. While the primary prevention strategies for breast cancer are currently under exploration, early detection considerably improves prognosis. For early detection of breast cancer, mammography is one of the most commonly used screening techniques, reported to be responsible for a 20-40 percent reduction in fatalities [4]. Furthermore, it can provide radiologists with

- X. Yu and C. Kang are with the Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K. E-mail: {xy144, ck254}@le.ac.uk.
- D. S. Guttery is with the Leicester Cancer Research Centre, University of Leicester, Leicester LE2 7LX, U.K. E-mail: dsg6@leicester.ac.uk.
- S. Kadry is with the Department of mathematics and computer science, Faculty of science, Beirut Arab University, Beirut, Lebanon. E-mail: s.kadry@bau.edu.lb.
- Y. Chen is with the Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China also with the School of Computer Science and Engineering, Southeast University, Nanjing 210096, China also with the School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China also with the Key Laboratory of Computer Network and Information Integration (Ministry of Education), Southeast University, Nanjing 210096, China. E-mail: chenyang.list@seu.edu.cn.
- Y.-D. Zhang is with the Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K. and also with the Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia. E-mail: yudongzhang@ieee.org.

an image upon on which they can make diagnostic decisions accordingly. However, the time-consuming interpretation and complexity of mammograms can result in a high false-positive rate and more importantly, misdiagnosis (i.e., false-negatives). Therefore, efficient and accurate computerised auxiliary diagnostic systems are becoming increasingly important both for radiologists and clinical practice.

The advent of deep learning has come to the fore as a method for concurrent computer-aided (CAD) systems while the performance of conventional CAD systems being faraway from satisfaction. Unlike traditional breast cancer CAD systems that heavily rely on manually designed components for specific purposes and are hindered by a lack of generality, modern CAD systems incorporating deep learning have been improved both on accuracy and efficiency [5], [6], [7], [8]. Compared with the multiple-step traditional CAD systems, CAD systems based on deep CNN generally solely consist of candidate detection and classification components.

To achieve an end-to-end detection system of the lesion, Lotter *et al.* proposed to train deep CNN with patch-based lesion regions in the DDSM dataset [9]. Subsequently, candidate areas, determined by a scanning model, are fed to the pre-trained classifier [10]. To minimise the intervention of the user while using the detection system, an automated CAD system that integrates detection, segmentation, and classification of breast masses has been proposed in [6]. In that study, cascaded deep learning methods for detection were applied to select hypotheses, which were subsequently refined by Bayesian optimization. Furthermore, the segmentation breaks down into two steps: gross segmentation is obtained by deep structured output learning and then improved by a level set method [11], [12]. Classification is realized by a deep learning classifier pre-trained with hand-crafted features and

fine-tuned on annotated breast mass classification datasets. The approach was tested on the CBIS-DDSM dataset and achieved a true positive rate (TPR) of 0.98 ± 0.02 at 1.67 false positives per mammogram on testing data. In another study, 45,000 mammogram images from a private database were used to compare the performance of traditional CAD systems with CNN-based approaches [13]. Usually, the supervised deep CNN models have to be trained with a large cohort of images before eliminating overfitting and achieving high-level performance. However, large numbers of annotated images are sometimes inaccessible, primarily due to the high potential expenditure on collection and maintenance. Transfer learning, therefore, is widely employed in various studies because it allows the knowledge learned from one domain to be transferred into another. Depending on the specific scenarios of the utilisation of transfer learning, pre-trained deep CNN models can be repurposed for detection, classification tasks by fine-tuning or can purely be used as feature extractors. As justified by Tajbakhsh *et al.,* the pre-trained deep CNNs, after sufficient fine-tuning, performed no worse than deep CNNs trained from scratch. Also, pre-trained deep CNNs showed higher robustness to the size of the dataset for fine-tuning [14]. However, deep tuning and shallow tuning are difficult as no standard criteria exist to determine which one is superior but depends on specific scenarios. Furthermore, Azizpour *et al.* pointed out that the high similarity between databases for pre-training and targeted databases gives rise to the success of the transfer learning approach. For transfer learning in the field of breast cancer CAD systems, there are numerous works that lead to successful conclusions. A seven-layer CNN model, four convolutional layers and three fully connected layers, achieved an AUC of 0.81 in a breast tomosynthesis (DBT) training set after being trained by regions of interest (ROIs) extracted DBT dataset. However, the AUC increased to 0.90 after transferring with DBT [15]. In another study, the capability of different CNNs for tumour feature extraction when used as feature extractors was compared. The performance of support vector machines trained on CNN-extracted features and computer-extracted features showed that transfer learning improves the performance of computer-aided-diagnosis (CADx) systems [16].

In this paper, we presented a new breast abnormality diagnostic system that achieved high accuracy on MINI-MIAS and INbreast datasets [17], [18]. In these two datasets, abnormalities in mammograms (mainly microcalcification and mass) are annotated by experienced radiologists and extracted according to location information. The backbone of our model is ResNet-SCDA-50, which is based on ResNet-50 trained by augmented data given by our SCDA, which applies scaling operation and CLAHE to form the augmented training dataset. CLAHE was performed for two reasons: 1) CLAHE is widely used for medical image enhancement [19], [20] and the contrast of breast images is greatly enhanced after applying CLAHE. Therefore, higher quality images for the subsequent components in our system can be obtained. and 2) it can be repurposed as a data augmenter by stacking the enhanced images with original images. Scaling is involved because we found that our CNN models trained by dataset being scaled and enhanced by CLAHE showed better performance. The classifiers designed in this work were based on a transfer learning approach that utilizes the CNN models as feature extractors with new classifier layer concatenated. We show here that ResNet-50 is superior as a feature extractor towards classification. Therefore, after cascading all of the optimized blocks, we found our ResNet-SCDA-50 achieved at 95.71 percent, which outperformed the state-of-the-art by a big margin.

2 BACKGROUND

The development of image processing technique can be divided into pre-deep-learning era and deep learning era. In pre-deep learning era, image processing tasks like segmentation, clustering, enhancement are solved by constructing specific models [21], [22], [23], which are later on termed as classical methods. Initially designed for large scale image recognition [24], [25], [26], [27], [28], [29], [30], deep CNNs have been widely utilized in a range of fields including natural language processing [31], [32], speech recognition [33], segmentation [34], [35], [36] and object localization [37], [38] thanks to the rapid development of firmware and software. Also, recent years witnessed the great advancement of deep CNNs including the depth, connections between layers and even more advanced training methods [27], [28], [29], [39], [40], [41].

It is difficult to tell which contributes most to the success of the advanced deep CNNs but it is of certain that connections between layers would be too profound to be neglected. In GoogLeNet [27], a novel architecture termed inception, which increases the depth and width of the network but keeps the computational cost low at the same time, was developed. InceptionV3 [39], an advanced version of GoogLeNet, improves inception model by replacing big kernels of filters with smaller ones and introducing new connections. Inspired by Inception block, Xception architecture [40], which forms of a depthwise convolution and the following pointwise convolution, was proposed. Compare to canonical convolution, depthwise and pointwise convolution is more parameter efficient. Given the size of feature-maps $M \times N$ by W channels, the size of kernels $D_F \times D_F$, the size of features-maps after convolution $M' \times N'$ by W' channels, then the total number of parameters of canonical convolution is $D_F \times D_F \times W \times W'$ with bias term neglected. For depthwise convolution, where the group of filters is 1, the number of parameters is $D_F \times D_F \times W$; For pointwise convolution, where the size of kernel is 1×1, the number of parameters is $W \times W'$, which leads to a total number of parameters of $(D_F \times D_F + W') \times W$. Therefore the rate of parameters between depthwise and pointwise convolution and canonical convolution is:

$$
\begin{aligned}
R_D &= \frac{(D_F \times D_F + W') \times W}{D_F \times D_F \times W \times W'} \\
&= \frac{1}{W'} + \frac{1}{D_F{}^2},
\end{aligned}
\tag{1}
$$

As can be seen in Eq. (1), the reduction rate $1 - R_D$ could be meaningful if provided proper W' and D_F. To solve the problem with training deep CNNs, residual learning, a new short-cut connection method, was presented in [28]. In residual learning, given the identity X and the mapping of $F(X)$, suppose that the underlying mapping of features $H(X)$ after a sequence of stacked layers is related to $F(X)$ and x by:
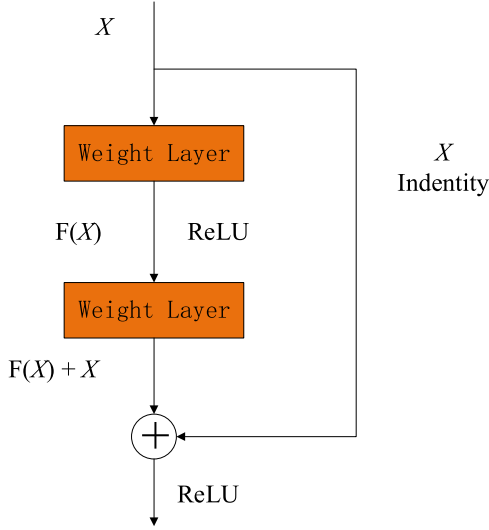
Fig. 1. Residual block.

Fig. 2. Dense connection block.

$$F(X) = H(X) - X. \tag{2}$$

Then $H(X)$ can be reformulated as:

$$H(X) = F(X) + X. \tag{3}$$

It is assumed the optimization of original $H(X)$ would be easier if it can be optimized by the identity mapping $F(X) + X$. The identity mapping was realized in the residual block by shortcut connections which skip one or more layers as presented in Fig. 1. ReLU is the Rectified Linear Unit [42].

DenseNet [29] was another state-of-the-art network which allowed subsequent layers access to the features extracted by previous layers by concatenating those feature-maps. Suppose that lth layer takes all of feature-maps $x_0, \ldots, x_{l-1}$ produced in previous layers as input, then feature-maps x_l in lth can be obtained by:

$$x_l = H_l([x_0, \ldots, x_{l-1}]), \tag{4}$$

where $[x_0, \ldots, x_{l-1}]$ denotes the concatenation operation of feature-maps in previous layers. $H_l(\cdot)$ is a composite function comprised of three sequenced operations: batch normalization (BN), rectified linear unit (ReLU), and convolution with the kernel size 3 by 3. The illustration of the dense connection used in DenseNet is given in Fig. 2.

In Fig. 2, W_i, H_i is the width and height of feature maps, M denotes the channels of feature maps. k is the growth rate to control the number of new feature maps produced in each layer. In summary, the performance of deep CNNs can be improved by solely stacking more convolutional layers to a certain depth, architectural innovations such as shortcut connection, however, remains to be most desirable in advancing deep CNNs. Also, the performance of the state-of-the-art CNNs is closely related to the practical application scenarios. We then explored the most appropriate structure of CNNs by transferring different CNNs into our classification tasks here.
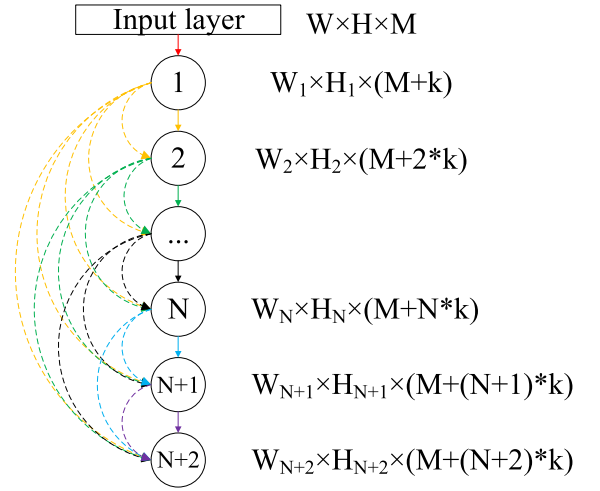
3 METHODOLOGY

3.1 Dataset

This section provides with a brief introduction to dataset MINI-MIAS and INbreast. MINI-MIAS is a widely used digitalized mammogram dataset that contains 322 grayscale images from 161 patients, with both left and right breasts. Each image is 1024 pixels × 1024 pixels. According to the categories given, there are mainly four types of abnormalities presented: microcalcification, mass (well defined, spiculated, ill-defined), architectural distortion, and asymmetry. According to the severity of abnormality, all 322 images can be divided into three categories that are normal without any lesions found, benign, and malignant. The corresponding number of images for the three categories as mentioned above were 207, 64, and 51, respectively. As the location information of lesions in each mammogram image is provided, we selected the patches of interest from the original mammogram images directly. For mammogram images without lesions, we chose the patches by following the criteria that each patch is filled with breast covering more than 80 percent area of the entire patch. Also, for mammogram images containing two or more lesions, each lesion is centred and cropped out of the image. By doing so, 330 patches are acquired where 207, 69, and 54 of them are classified as normal, benign, and malignant, respectively. The details of the dataset are presented in Table 1. Similarly, we acquired 2292 abnormal patches and 12171 normal patches from INbreast dataset. To mitigate the problem with the imbalance between normal samples and abnormal samples, we randomly selected 2292 patches from 12171 normal samples. The information about MINI-MIAS is presented in Table 1 while the details of the merged dataset are shown in Table 2.

3.2 Contrast Limited Adaptive Histogram Equalization

In medical images, one of the challenges for automated systems is the low contrast, which leads to failures of algorithms in processing or analyzing. To address the issue of low contrast, digitized image contrast enhancement methods were proposed.

TABLE 1
Details of MINI-MIAS

Categories	Images	Patches
Normality	207	207
Benign	64	69
Malignant	51	54
Overall	322	330

Ordinary histogram equalization was performed in the following way. Given a discrete grayscale image [X], let N_i be the frequency for grayscale i to appear in the image having N pixels in total, then the possibility of pixel x belongs to grayscale i in range of 0 to L is:

$$p_x(i) = p(x = i)$$
$$= \frac{N_i}{N}(0 \le i \le L), \tag{5}$$

where L is the highest intensity over the image [X]. Therefore, the $p_x(i)$ is the normalized probability that ranges from 0 to 1 with respect to grayscale i. Hence, the cumulative distribution function (CDF) can be expressed as:

$$f_{cd\_x}(i) = \sum_{j=0}^{i} p_x(j). \tag{6}$$

Conversion of the CDF into an equalized one is to map $f_{cd\_x}(i)$ into a new function by multiplying $f_{cd\_x}(i)$ with the desired constant pixel value, which is generally set to be 255. However, for regions that are significantly lighter or darker than most regions in the image, sufficient contrast enhancement is unlikely to be obtained. To improve the details in local regions of the image, adaptive histogram equalization turns out to be the alternative method. Unlike histogram equalization that transforms pixels according to image histograms globally, adaptive histogram equalization transforms each pixel by using a transformation function derived from a local region. However, the slow computational speed and tendency to overamplify noises in the relatively homogeneous area are two main shortcomings to be solved. As a variant of adaptive histogram equalization, CLAHE has shown to overcome these problems efficiently [19].

The enhancement procedure of CLAHE can be divided into two phases: histogram equalization with contrast limit applied and bilinear interpolation algorithm. Histogram equalization first performs to provide an adjusted histogram while bilinear interpolation algorithm reduces the computational costs on forming a new image. Given an image I, the M by N subregions by dividing I, the process of histogram equalization can be divided into the following steps:

TABLE 2
Overall Composition of the Merged Dataset

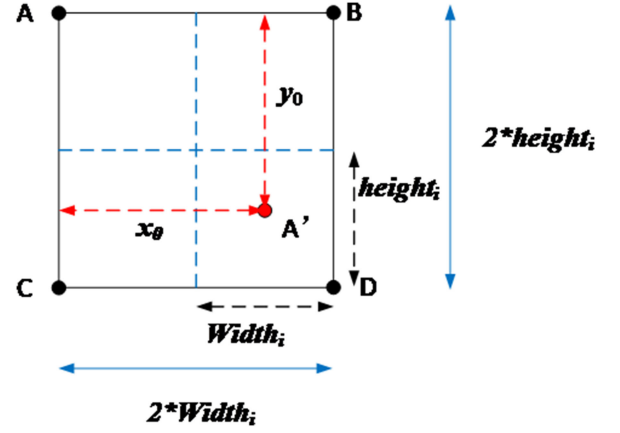| Categories	MINI-MIAS	INbreast	Overall
Normality	207	2292	2499
Abnormality	123	2292	2415
Overall	330	4584	4914

Fig. 3. Bilinear interpolation.

$$f_x^n(i) = \sum_{j=0}^{i} p_x^n(j) \tag{7}$$
$$(0 \le i \le 255, 1 \le n \le M * N),$$

where p_x^n is as the same form as Eq. (5) but was the possibility of pixel x in subregion n. Initially, the summation of pixel intensity that is above the pre-determined clip limit CL is acquired, normalized and can be noted as T^n, which is later evenly distributed to all intensity levels by the average increase AI^n. Therefore AI^n can be simply denoted as:

$$AI^n = \frac{T^n}{256}. \tag{8}$$

Then the adjusted histogram can be expressed in the following way:

$$p_x^n(i)' = \begin{cases} \text{CL} & if \quad p_x^n(i) \ge \text{CL} - AI^n \\ p_x^n(i) + AI^n & if \quad p_x^n(i) < \text{CL} - AI^n \end{cases}. \tag{9}$$

After the adjustment of image histogram, block linearly interpolation was applied to reduce the computational cost. As shown in Fig. 3, for pixel A' in subregion i, the value is determined by the four pixels in neighbourhood subregions and can be calculated by:

$$P_{A'} = x_0 * y_0 * P_A + (1 - x_0) * y_0 * P_B$$
$$+ x_0 * (1 - y_0) * P_C + (1 - x_0) * (1 - y_0) * P_D, \tag{10}$$

where P_A, P_B, P_C, P_D, and $P_{A'}$ denotes the pixel value of pixel A, B, C, D, and A' respectively. x_0 and y_0 are normalized by 2*Width.

A pair of breast patches before enhancement and after is shown in Fig. 4.

We used mean squared error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Absolute Mean Brightness Error (AMBE) [43] to measure the quality changes of the image before and after enhancement. Let I and J be the image before and after enhancement respectively, the size of which is m rows by n columns. Then MSE can be expressed as:

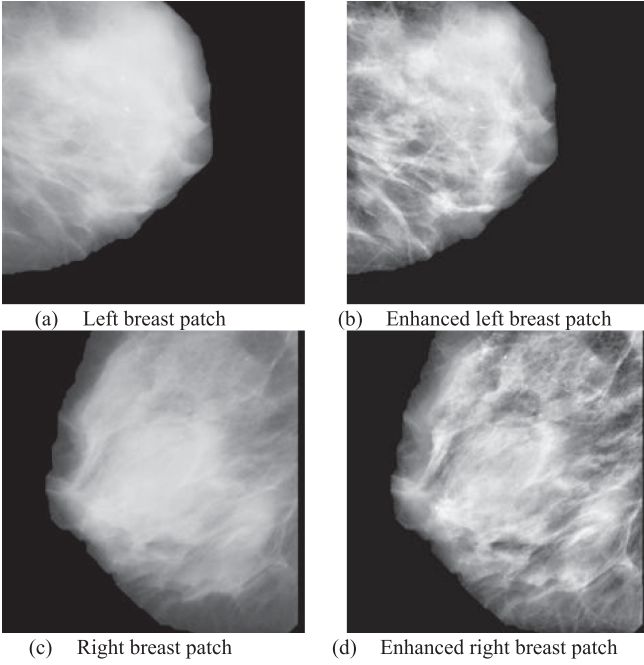$$MSE = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (J(i,j) - I(i,j))^2. \tag{11}$$

(a)   Left breast patch   (b)   Enhanced left breast patch

(c)   Right breast patch   (d)   Enhanced right breast patch

Fig. 4. Breast patches and corresponding enhanced patches.

TABLE 4
Contrast of Patches Before and After Enhancement

ID OF PATCHES	ORIGINAL	CDA-ENHANCED	DIFFERENCE
1	6.61	8.11	1.50
2	8.11	80.99	72.88
3	4.55	55.80	51.25
. . .	. . .	. . .	. . .
4914	73.69	131.60	57.91
	10.42 (mean)	60.59 (mean)	50.17 (mean)

reflecting the ratio between signal and noise, implies the better depression of noise.

According to the above equations, we were able to calculate corresponding values with respect to MSE, PSNR, and AMBE. We received MSE with an average value of 218.67 while the maintained PSNR to be desirable. Moreover, mean AMBE is 8.87, which is the increase in intensity. Detailed results are presented in Table 3. As contrast is another valuable indicator of image quality, we compared our enhanced image with the original image and found the enhanced image is superior to the original image, as can be seen from Table 4.

3.3 Architecture

3.3.1 Architecture of CNN

However, in practice, object categories in classification tasks are generally no more than hundreds while training a network with a large number of parameters from scratch is strenuous and time-consuming, especially when the size of training data is small. Therefore, transfer learning, an effective way to adapt classifiers trained in one domain to another, can be more feasible though slight changes in architecture generally need to be made. There are two ways of transferring a pre-trained CNN to a new task including combining the pre-trained CNN as a feature extractor with a corresponding new classifier or fine-tuning the pre-trained CNNs by adjusting the architectures. When used as feature extractors, the parameters in the pre-trained CNNs remain unchanged whilst features extracted in a certain depth of CNN were taken as the input of the concatenated classifiers. When pre-trained CNNs are fine-tuned, the architecture of base CNNs has to be changed slightly. For example, the number of outputs in CNNs designed for large scale image classification is 1,000. However, when applying CNNs to binary classification, the last 1000-neuron fully connected layer should be replaced with a 2-neuron fully connected layer, which turns out to be the new classifier. Then the number of parameters in the overall CNN need to be updated outnumbers that of the network combining CNN-based feature extractor with adjoining classifier. Thus it takes longer time to train the classifiers in the former form compared to the one in the latter form.

To reduce the computational costs of our system in this work, we used pre-trained CNNs as feature extractor and therefore designed new classifiers that take the extracted features as input. The dimensions of features output by CNNs after fully-connected layer are generally more than thousands. Directly feeding the new binary-classification classifier
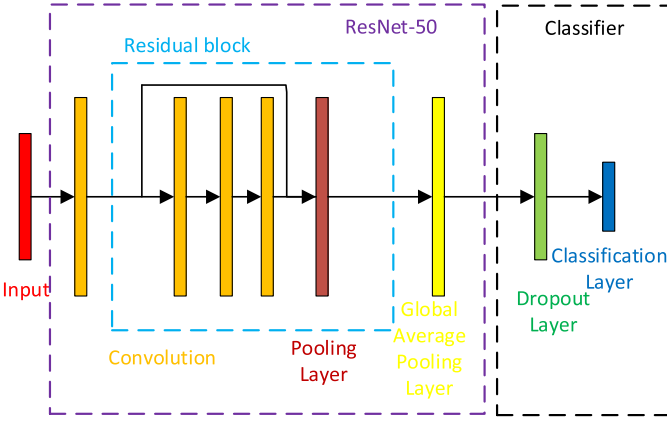
The calculation of PSNR was based on MSE and can be written as:

$$PSNR = 10 \times \log_{10}\left(\frac{(L-1)^2}{MSE}\right), \qquad (12)$$

where L stands for the value of the highest grey level in the original image. AMBE, which is the absolute value between the expectation of enhanced image and the expectation of the original image, can then be presented in the following form:

$$AMBE = |E[Y] - E[X]|, \qquad (13)$$

where $|\bullet|$ indicates absolute operation while the E[X] and E[Y] are calculated by:

$$E[\,X] = \frac{1}{m \times n}\sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j) \qquad (14)$$

$$E[\,Y] = \frac{1}{m \times n}\sum_{i=1}^{m}\sum_{j=1}^{n} J(i,j). \qquad (15)$$

The bigger MSE means the more significant difference between the image before and after enhancement, which signals the effectiveness of enhancement. The greater PSNR,

TABLE 3
Measurements of Enhancement

ID OF PATCHES	MSE	PSNR	AMBE
1	306.79	16.93	13.62
2	242.81	20.39	9.23
3	266.10	18.03	12.07
. . .	. . .	. . .	. . .
4914	513.87	20.39	10.80
	218.67 (mean)	20.21 (mean)	8.87 (mean)

Fig. 5. Architectures of proposed classifiers.

(a) Original image   (b) Scaled image, Scaling factor S_f = 1.5

(c) Cropped image of (b)

Fig. 6. Image scaling.

with those features would lead to overfitting, which can be effectively mitigated by introducing dropout layer between the fully-connected layers and the final classification layer. The gross architectures of combined classifiers are presented in Fig. 5. Thereafter, to avoid confusion, when the state-of-the-art networks in the remainder of this paper are mentioned, they are actually referring to the networks experienced the same adjustment in Fig. 5.

3.3.2 Architecture of SCDA

SCDA is a sequenced framework that consists of augmentation and concatenation. For augmentation, image scaling is first applied while CLAHE is applied to the scaled image subsequently. In this research, we chose the scale factor that is greater than 1, which aims at providing refined images. Scaling is implemented by resizing the images in the training set to S_f times of original size, followed by cropping operation that crops out the patches having the same size as original images from the centre of the scaled images (Fig. 6), where blue circular item denotes the lesion region in the original image. Scaling factor S_f specifies the ratio between the original width and the width of the scaled image. After the acquisition of the cropped image, CLAHE is applied then. The detailed flows of SCDA are presented below. Given the original training set X, the augmented training set X' can be expressed as:

$$X' = H[X, T(X)], \qquad (16)$$

where $H[\cdot]$ remains as concatenation, $T(\mathrm{X})$ is the transformation of X and can be further extended as:

$$T(X) = CLA(S(X)). \qquad (17)$$

$CLA(\cdot)$ is the operation of CLAHE while $S(\mathrm{X})$ stands for the scaling transformations.

3.4 Experiment design

To avoid the unexpected experimental results due to the inappropriate partition of the dataset, we utilised 5-fold cross-validation to examine our proposed data augmentation method and the classification models. At the beginning of the experiment, 80 percent, namely 4 out of 5 folds, of the dataset is partitioned into training set while the ratios between negative samples and positive samples in training
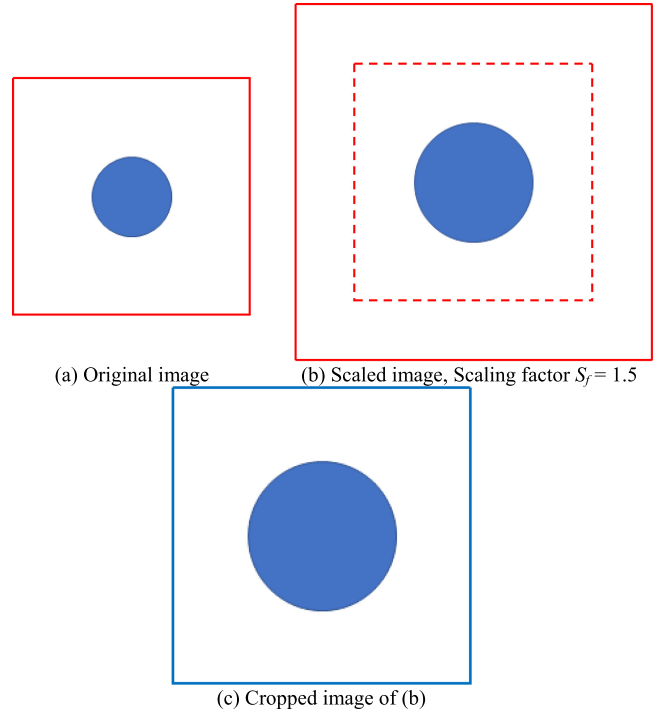
set and testing set are roughly close to 1:1. CLAHE was then applied to the training set while testing set remained unchanged. As a result, a new training set that combines the original training set and the enhanced training set is formed. Later on, different CNN models were trained by training datasets before and after enhancement. Note that the size of the input image is correspondingly adjusted to meet the input requirement of different models.

3.4.1 Configurations

All of the experiments were carried out on a personal computer with a single Geforce GTX 1060 GPU and a memory of 8G. The framework used in this research was the deep learning toolbox provided by Matlab with pre-trained deep CNN models. With no specification, the following parameters were applied: 0.5 was set as the dropout probability of the fully connected layer, the optimization algorithm was adaptive moment estimation, 64 for mini-batch size. By trial and error, we set the number of maximum epochs to be 20 and the initial learning rate to be 0.01. The learning rate dropped to 0.1 of the previous one every 10 epochs.

3.4.2 Model Evaluations

In this paper, we followed the standardised accuracy calculation criteria. The accuracy comes from two parts, which are specificity and sensitivity that can be defined as:

$$R_{TP} = \frac{TP}{TP + FN} \qquad (18)$$

$$R_{TN} = \frac{TN}{TN + FP}, \qquad (19)$$

where TP, FN, TN, FP stands for true positive, false negative, true negative, and false positive, respectively. R_{TP} and

TABLE 5
Averaged Performance of Classifier Before
and After Enhancement

Models	CDA	Sensitivity (%)	Specificity (%)	Accuracy (%)
GoogLeNet	N	88.41	96.73	92.64
	Y	**90.84**	**98.10**	**94.53**
ResNet-50	N	89.66	97.41	93.61
	Y	**92.39**	**98.50**	**95.50**
DenseNet201	N	87.35	96.53	92.02
	Y	**90.72**	**97.95**	**94.40**
Xception	N	88.85	96.65	92.82
	Y	**92.38**	**97.56**	**95.01**
Inceptionv3	N	88.81	96.57	92.76
	Y	**92.80**	**97.53**	**95.21**

R_{TN} correspond to true positive rate (the sensitivity), and true negative rate (the specificity), respectively. Specificity and sensitivity, therefore, lead to the accuracy by:

$$R_{AC} = \frac{TN + TP}{TN + FN + TP + FP}, \qquad (20)$$

To examine the performance of CNN models, the sensitivity, specificity and accuracy are calculated for 5 folds testing sets respectively and then be averaged to eliminate the fluctuation of performance due to the random partition of datasets.

4 EXPERIMENT RESULTS

4.1 Performance of CNN Models Before and After Contrast Enhancement

As the enhancement experiment showed, the contrast, as well as the quality of enhanced images, is greatly improved. To validate the effectiveness of utilizing CLAHE as the data augmentation method, we trained the different CNN models by original training set and augmented training dataset and the results are shown in Table 5.

CDA stands for whether CLAHE was applied as a data augmentation method (**Y**) or not (**N**). The bold font denotes the more preferable results. As can be seen from Table 5, all of the models have higher specificity than the sensitivity while many of them showed similar performance in terms of final accuracy when these models are trained by the original training set. However, ResNet-50 achieves the exceptional accuracy given the fact that it was the mean accuracy of 5 folds. It seems that residual learning is more effective in giving more useful representations of high-level features in this classification scenario. Moreover, the specificity and sensitivity of CNN models trained by the augmented training set are consistently higher than that of models trained by the original training set. As a result, the higher specificity and the sensitivity contribute to a nearly 2.5 percent accuracy increase, which shows the feasibility of the proposed data augmentation method.

4.2 Performance of SCDA

To validate the performance of SCDA, we compared SCDA with other traditional data augmentation methods on ResNet-50. Results are shown in Table 6.

TABLE 6
Results of ResNet-50 Trained by the Augmented Training Set

Augmentation method	Sensitivity (%)	Specificity (%)	Accuracy (%)
Scaling	**93.24**	98.10	95.71
Vertical shift	92.78	98.33	95.60
Horizontal shift	92.52	98.48	95.55
Rotation	92.17	98.61	95.45
Vertical Flip	92.44	**98.70**	95.62
Horizontal Flip	92.80	98.44	95.67
SCDA (our method)	92.83	98.55	**95.74**

TABLE 7
Method Comparison

Models	Method	Averaged Accuracy (%)	Best Accuracy (%)
Nguyen [44]	HMI + FNN	73.50 ± 1.35	-
Gorgel Sertbas [45]	SWT-SVM	-	90.10
Liu [46]	WFRFT + PCA + SVM	92.16 ± 3.60	-
Wu [47]	FRFE + CAR-BBO	-	92.52
Our method	ResNet-SCDA-50	**95.74 ± 0.24**	**96.11**

The scaling factor is 1.2. Vertical shift and horizontal shift translate the images by 20 pixels downward and rightward respectively. For rotation, we choose the random rotation angle ranging from 0 to 360. Vertical flip and horizontal flip generate new images by flipping images vertically and horizontally. In our method, we set the scaling rate to be 1.2 because scaling (when the scaling factor is 1.2) not only gives best-trained ResNet-50 amongst all of the traditional methods but also shows high sensitivity. The possible explanation could be fine-grained structures in scaled images turn out to be more clear and recognizable. However, the reason why vertical flip provides the highest specificity remains to be explored. Nevertheless, our method performed best in terms of accuracy though sensitivity and specificity are continuing to be improved.

4.3 Comparison With State-of-the-Art Approaches

In this section, we showed the comparison results in our best model with state-of-the-art methods. The best performance of our model achieved an accuracy of 96.11 while the averaged accuracy of our model on 5 runs of cross-validation is 95.74 with a deviation of 0.24. As can be seen, our model outperforms all of the approaches by a significant margin in terms of accuracy (Table 7).

5 DISCUSSION

Due to the powerful performance of deep CNNs on feature extraction, some traditional image preprocessing techniques are becoming unnecessary for CNN based diagnostic systems. However, we deployed CLAHE image contrast enhancement as a new data augmentation method and found that the performance of the system greatly improved. From the experimental results of CLAHE, all of the metrics showed significant increases, which means the quality and the contrast of image are improved. The purposes of data augmentation are mainly two-fold. One is to mitigate the problem brought by lack of data; another is to solve the problem of overfitting. As can be seen from the results, CLAHE perfectly

fulfils these two objectives and results in an averaged increase in accuracy of all of the models. Therefore, CLAHE can be used as a new data augmentation method besides an image contrast enhancement method.

Based on this, the proposed SCDA method, which combines traditional image scaling and CLAHE, performs even better on data augmentation than sole CDA or scaling as can be proved by experimental results. Also, there is of great potential in exploring multiple combinations, which integrates two or more traditional data augmentation methods with our proposed method.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed a new data augmentation method termed as SCDA and developed a diagnostic system for accurate classification of breast abnormalities. Before inputting the patches acquired from original mammogram images to our CNN network, we improved data preprocessing and data augmentation by applying the CLAHE contrast enhancement method on patches with the enlarged ROIs by scaling. Measurement of image contrast shows that the quality of image patches has been considerably improved. To specify the CNN models showing best performance on the binary classification task, we explored the models with state-of-the-art connection methods including inception (GoogLeNet, Inception v3), residual learning (ResNet), dense connection (DenseNet), depthwise and pointwise convolution (Xception). The experimental results show that ResNet-50 gives the best performance amongst all of those models. Therefore, we propose to provide ResNet-50 with the augmented training set formed by SCDA method and the performance of ResNet-50 achieved averaged accuracy of 95.74 percent after the 5-fold cross-validation. Therefore, we believe our system has great potential in the field of diagnosing breast abnormalities. But there are some aspects of our system to be enhanced. As can be seen from the experiment results, the sensitivity, which is an important reference to the performance of the CAD system, could be further improved. This problem could be partly solved by further exploration of the combination of classical data augmentation and our proposed one, which will be one of the future works. Moreover, the performance of the diagnosis system relies highly on the performance of the detection system. Therefore, we will focus on developing automatic detection systems for breast abnormalities by integrating our improved diagnositic system, and therefore complete an end-to-end CAD system for the full mammographic image. Also, the size of the dataset indirectly determined the performance of CAD systems. So, it's more reasonable for us to work on a larger dataset, based on which can be meaningful to improve the performance of the current system.

REFERENCES

[1] W. C. R. Fund Breast Cancer Statistics, May 12, 2018. [Online]. Available: [Online]. Available: *https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics*

[2] WHO, Cancer, May 12, 2019. [Online]. Available: *https://www.who.int/news-room/fact-sheets/detail/cancer*

[3] J. B. Sarah Caul, Cancer registration statistics, England: 2017, May 26, 2019 [Online]. Available: *https://www.ons.gov.uk/peoplepopulationand community/healthandsocialcare/conditionsanddiseases/bulletins/cancer registrationstatisticsengland/2017*

[4] A. C. Society. May 12, 2016. [Online]. Available: *https://www.cancer. org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2015-2016.pdf*.

[5] K. P. Michiel Kallenberg, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1322–1331, May 2016.

[6] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," *Med. Image Anal.*, vol. 37, pp. 114–128, Apr. 2017.

[7] L. Shen, "End-to-end training for whole image breast cancer diagnosis using an all convolutional design," 2017, *arXiv:1711.05775*.

[8] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks," *J. Medical Imag.*, vol. 6, p. 031409, 2019.

[9] M Heath, K. Bowyer, D Kopans, R Moore, "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digit. Mammography*, 2000, pp. 212–218.

[10] W. Lotter, G. Sorensen and C. David, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. ed., Berlin, Germany: Springer, 2017, pp. 169–177.

[11] R. Fedkiw and S. Osher, *Level Set Methods and Dynamic Implicit Surfaces*, Berlin, Germany: Springer, vol. 153, 2006.

[12] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[13] T. Kooi *et al.*, "Large scale deep learning for computer aided detection of mammographic lesions," *Med. Image Anal.*, vol. 35, pp. 303–312, Jan. 2017.

[14] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: full training or fine tuning?," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May, 2016.

[15] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Med. Phys.*, vol. 43, Dec. 2016, Art. no. 6654.

[16] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *J. Med. Imag.*, vol. 3, 2016, Art. no. 034501.

[17] A. G. Gale, "The mammographic image analysis society digital mammogram database," in *Proc. 2nd Int. Workshop Digit. Mammography*, 1994, vol. 1069, pp. 375–378.

[18] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, pp. 236–248, Feb. 2012.

[19] Z. S. Pisano *et al.*, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imag.*, vol. 11, pp. 193–200, Nov. 1998.

[20] N. M. Sasi and V. Jayasree, "Contrast limited adaptive histogram equalization for qualitative enhancement of myocardial perfusion images," *Engineering*, vol. 5, 2013, Art. no. 326.

[21] Y. Jiang *et al.*, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Trans. Cybern.*, vol. 45, pp. 688–701, 2014.

[22] Y. Jiang, J. Zheng, X. Gu, J. Xue, and P. Qian, "A novel synthetic CT generation method using multitask maximum entropy clustering," *IEEE Access*, vol. 7, pp. 119644–119653, 2019.

[23] P. Qian *et al.*, "mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 819–832, Apr. 2020.

[24] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei *ILSVRC-2012*, 2012. [Online]. Available: http://www.image-net. org/challenges/LSVRC/2012/

[25] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[27] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[28] K. He, X. Zhang, R. Shaoqing, and S. Jian, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[29] H. Gao, L. Zhuang, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[30] P. Qian *et al.*, "SSC-EKE: Semi-supervised classification with extensive knowledge exploitation," *Inf. Sci.*, vol. 422, pp. 51–76, 2018.

[31] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[32] A. Bordes, S. Chopra, and J. Weston, "Question answering with subgraph embeddings," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2014, pp. +615–620.

[33] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6. pp. 82–97, Nov. 2012.

[34] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.

[35] K.-J. Xia, H.-S. Yin, and Y.-D. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-flow algorithm," *J. Med. Syst.*, vol. 43, 2019, Art. no. 2.

[36] P. Qian *et al.*, "Cross-domain, soft-partition clustering with diversity measure and knowledge reference," *Pattern Recognit.*, vol. 50, pp. 155–177, 2016.

[37] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[38] S. Ren, K. He, R. Girshick, and S. Jian, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[40] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.

[41] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[42] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[43] R. Porwal and S. Gupta, "Appropriate contrast enhancement measures for brain and breast cancer images," *Int. J. Biomed. Imag.*, vol. 2016, 2016, Art. no. 4710842.

[44] X. Zhang, J. Yang, and E. Nguyen, "Breast cancer detection via Hu moment invariant and feedforward neural network," in *Proc. AIP Conf.*, 2018, Art. no. 030014.

[45] P. Görgel, A. Sertbas, and O. Uçan, "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines," *Expert Syst.*, vol. 32, pp. 155–164, 2015.

[46] G. Liu, "Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform," *Adv. Mech. Eng.*, vol. 8, Feb. 2016, Art. no. 11.

[47] X. Wu, "Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization," *Simulation*, vol. 92, pp. 873–885, Sep. 12, 2016.

Xiang Yu received the bachelor's and master's degrees from Huanggang Normal University and Xiamen University, P.R. China, in 2014 and 2018, respectively. Currently, he is Working toward the PhD degree in the Department of Informatics, University of Leicester, U.K. Also, he was sponsored by CSC and by the University of Leicester as a graduate teaching assistant (GTA). His research interests include medical image segmentation, machine learning, and deep learning.

Cheng Kang received the master's degree from Shenzhen University, he is currently working toward the PhD degree from the Department of Informatics, University of Leicester, U.K. His research interests focus on EEG signal processing, deep learning. Currently, he is currently working on the early detection of breast cancer by artificial intelligence.

David Guttery He is currently a co-investigator on an integrated, collaborative programme of clinical and translational research between the University and Leicester and Imperial College funded by Cancer Research U.K. his research interests are intertwined with those of professor Jacqui Shaw (see Professor Jacqui Shaw's webpage), which focus on the utility of circulating nucleic acids and other circulating biomarkers for early detection and monitoring of cancer. He is currently a co-investigator on an integrated, collaborative programme of clinical and translational research between the University and Leicester and Imperial College funded by Cancer Research UK.

Seifedine Kadry (Senior Member, IEEE) received the bachelor's degree in applied mathematics, from Lebanese University, in 1999, the MS degree in computation from Reims University, France, and EPFL (Lausanne), in 2002, and the PhD degree in applied statistics, in 2007 from Blaise Pascal University, France, HDR degree, in 2017 from Rouen University. At present his research focuses on education using technology, system prognostics, stochastic systems, and probability and reliability analysis. He is ABET program evaluator.

Yang Chen (Senior Member, IEEE) received the MS and PhD degrees in biomedical engineering from First Military Medical University, China, in 2004 and 2007, respectively. Since 2008, he has been a faculty member with the Department of Computer Science and Engineering, Southeast University, China. His recent work concentrates on the medical image reconstruction, image analysis, pattern recognition, and computerized-aid diagnosis.

Yu-Dong Zhang (Senior Member, IEEE) received the PhD degree from Southeast University, in 2010. He worked as a postdoctoral from 2010 to 2012 in Columbia University, USA, and as an assistant research scientist from 2012 to 2013 at Research Foundation of Mental Hygiene (RFMH), USA. He served as a full professor from 2013 to 2017 in Nanjing Normal University, where he was the director and founder of Advanced Medical Image Processing Group in NJNU. Now he serves as professor with the Department of Informatics, University of Leicester, U.K.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.