SAFNet: a deep spatial attention network with classifier fusion for breast cancer detection

Si-Yuan Lu[1], Shui-Hua Wang[1,*], Yu-Dong Zhang[1,*],

1 School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK

* Correspondence should be addressed to Shui-Hua Wang & Yu-Dong Zhang
Email: Si-Yuan Lu (sl672@le.ac.uk), Shui-Hua Wang (shuihuawang@ieee.org), Yu-Dong Zhang (yudongzhang@ieee.org)

Abstract: Breast cancer is a top dangerous killer for women. An accurate early diagnosis of breast cancer is the primary step for treatment. A novel breast cancer detection model called SAFNet is proposed based on ultrasound images and deep learning. We employ a pre-trained ResNet-18 embedded with the spatial attention mechanism as the backbone model. Three randomized network models are trained for prediction in the SAFNet, which are fused by majority voting to produce more accurate results. A public ultrasound image dataset is utilized to evaluate the generalization ability of our SAFNet using 5-fold cross-validation. The simulation experiments reveal that the SAFNet can produce higher classification results compared with four existing breast cancer classification methods. Therefore, our SAFNet is an accurate tool to detect breast cancer that can be applied in clinical diagnosis.

Keywords: breast cancer; ultrasound image; ResNet; randomized neural network; randomized vector functional-link; computer-aided diagnosis

1. Introduction

Various factors can cause breast cancer, including radiation, family history, etc. There are millions of women diagnosed with breast cancer annually, and approximately fifty percent of them would eventually lose their lives because of late diagnosis [1]. Therefore, accurate detection of breast cancer at its initial stage plays a crucial part in treating this deadly disease. Ultrasound imaging is one of the most prevailing imaging modalities in clinical diagnosis that can generate clear images of the organs inside bodies [2]. Hence, ultrasound images (USIs) can be used for breast cancer diagnosis.

Nevertheless, manual interpretation of the rich information in USIs suffers from low reproducibility and low efficiency, so developing automatic USI analysis systems is necessary, which can make accurate predictions based on the USIs. Due to the unprecedented development of deep learning and computer vision, computer-aided diagnosis (CAD) has gained significant improvements, such as classification and segmentation [3-6]. Researchers and practitioners have proposed a bunch of breast cancer detection models in the past decade.

Rouhi, Jafari, Kasaei and Keshavarzian (2015) [7] proposed a segmentation and recognition system for breast cancer detection based on mammograms. A cellular neural network was trained to segment the mammograms. Then, they extracted a group of handcrafted image features and trained several classical artificial neural networks for classification. Amrane, Oukid, Gagaoua and Ensari (2018) [1] used the *k*-nearest neighbors (*k*-NN) algorithm and naïve Bayesian classifier (NBC) to classify breast cancer from healthy samples using clinical data. They discovered that the *k*-NN performed better than the NBC. Gao, Wu, Li, Zheng, Ruan, Shang and Patel (2018) [8] presented a CNN-based model for breast cancer classification. They designed a shallow CNN model to learn latent representations from the medical images. Aslan, Celik, Sabanci and Durdu (2018) [9] employed four different classification models to diagnose breast cancer with blood statistics, including *k*-NN, extreme learning machine (ELM), artificial neural network (ANN), and support vector machine (SVM). They also utilized an optimization method to find the optimal hyper-parameters in the four classification models. Dai, Chen, Zhu and Zhang (2018) [10] trained a random forest to diagnose breast cancer in clinical data. The Wisconsin Diagnostic Breast Cancer dataset was utilized in the experiments to evaluate the classification ability of their method. Ghasemzadeh, Sarbazi Azad and Esmaeili (2018) [11] put forward a breast cancer classification approach using mammograms. A Gabor wavelet transform was employed to generate the image representations from the mammograms, and several machine learning (ML) algorithms were trained for classification, including SVM, ANN, decision tree (DT), etc. A public mammogram dataset (DDSM) was employed in their evaluation experiments. Gupta and Gupta (2018) [12] also trained a group of heterogeneous ML classifiers to detect breast cancer using the Wisconsin Diagnostic Breast Cancer dataset. Through their experiments, they found that the multi-layer perceptron was better than other classification models. Heidari, Khuzani, Hollingsworth, Danala, Mirniaharikandehei, Qiu, Liu and Zheng (2018) [13] presented a locally preserving projection to reduce the dimension of image feature vectors from the mammograms, which were extracted based on the bilateral asymmetry of the breasts. Two classical machine learning algorithms were selected as the classifiers, including SVM and *k*-NN. The leave-one-out validation method was employed to obtain the testing results of their model. Hussain, Aziz, Saeed, Rathore and Rafique (2018) [14] proposed to calculate a group of different features using the mammograms based on the texture analysis, Fourier transform, scale-invariant feature transform (SIFT), etc. Then, combined features were used to train an SVM and a DT using 10-fold cross-validation. Wang, Li, Wang, Jiang, Yao, Zhang and Xin (2019) [15] put forward a mass detection algorithm based on the deep CNN and ELM. They extracted the deep features and fused them with classical handcrafted features to form the feature vectors. Finally, an ELM was trained to distinguish the benign and malignant samples. Islam, Haque, Iqbal, Hasan, Hasan and Kabir (2020) [16] conducted a comparison of the classification results of the ANN and SVM on the Wisconsin Breast Cancer dataset for breast cancer detection. They found that the ANN outperformed the SVM marginally. Lahoura, Singh, Aggarwal, Sharma, Mohammed, Damasevicius, Kadry and Cengiz (2021) [17] provided a breast cancer classification system based on cloud computing. A gain ratio feature selection algorithm was used to eliminate the excessive features, and an ELM served as the classifier for remote breast cancer detection. The Wisconsin Diagnostic Breast Cancer was employed in their experiments. Zuluaga-Gomez, Al Masry, Benaggoune, Meraghni and Zerhouni (2021) [18] presented a breast cancer diagnosis framework using ResNet and thermal images. The trained model can achieve good classification performance in their experiments. Rehman, Zhuang, Muhamed Ali, Ibrahim and Li (2019) [19] attempted to find out the correlation between microRNAs and breast cancer. They proposed to utilize a bunch of feature selection algorithms to obtain the refined microRNA set which were the most related to breast cancer. A random forest and an SVM were trained for breast cancer classification with the refined microRNAs as the inputs. Singh and Singh (2020) [20] presented a systematic review on the breast cancer diagnosis using thermography. Stark, Hart, Nartowt and Deng (2019) [21] used personal healthcare data to predict breast cancer risk. The input features included age, ethnicity, etc. For classification, multiple machine learning models were trained, including ANN, SVM, NBC, etc. However, their accuracies were relatively low. Tapak, Shirmohammadi-Khorram, Amini, Alafchi, Hamidi and Poorolajal (2019) [22] implemented the survival prediction of breast cancer using clinical data and classical machine learning classifiers, including SVM, ANN, random forest, etc. however, the sensitivity of their best model was only 36%. Zheng, Lin, Gao, Wang, He and Fan (2020) [23] developed a breast tumor classification model using a convolutional neural network (CNN) and Adaboost algorithms. They gathered images of different modalities, including mammograms, USIs, and magnetic resonance images (MRIs). Their model can accurately classify the samples as normal, benign, and malignant after training. Khuriwal and Mishra (2018) [24] firstly exploited watershed segmentation to remove the background in the mammograms. Then, they extracted 12 texture and statistical features from the segmented images, including entropy, energy, mean value, standard deviation, etc. Finally, an ANN model was trained to classify breast cancer from normal controls. Kadam, Jadhav and Vijayakumar (2019) [25] employed a sparse autoencoder to identify malignant breast cancer from benign breast cancer. Mercan, Mehta, Bartlett, Shapiro, Weaver and Elmore (2019) [26] proposed a breast cancer classification framework based on biopsy images. They generated small patches from the images and trained a CNN model to generate patch-level tissue labels. Then, the image-level tissue labels can be obtained based on the fusion of patch-level tissue labels. Finally, an SVM was trained to identify the breast cancer subtypes. Turkki, Byckhov, Lundin, Isola, Nordling, Kovanen, Verrill, von Smitten, Joensuu, Lundin and Linder (2019) [27] employed a CNN to generate features from the tissue microarray images. Then, the principal component analysis (PCA) and Fisher vector were used for feature reduction. An SVM was trained to predict the input as high breast cancer risk and low breast cancer risk. Zeebaree, Haron, Abdulazeez and Zebari (2019) [28] put forward a segmentation method using USI. They separated the USIs into small patches and extracted several features from the patches to train a back propagation neural network (BPNN), which was used to classify the patches as region of interest (ROI) and non-ROI. Finally, a region growing algorithm was used to refine the segmented results. Sharma and Mehra (2020) [29] developed a new method to recognize breast cancer in histopathological images. They employed two different approaches to extract image features. One is the classical handcrafted features, including Hu moment, texture analysis. The other is the features based on deep CNN models, including VGG-16, VGG-19, and ResNet-50. In the classification stage, the SVM, linear discriminant analysis (LDA), random forest were trained to distinguish benign cancer from malignant ones. Zuluaga-Gomez, Al Masry, Benaggoune, Meraghni and Zerhouni (2020) [30] trained a deep CNN model to detect breast cancer in thermal images. In experiments, they tested a bunch of different backbone models. The best accuracy of their model was 92%. Mahmood, Arsalan, Owais, Lee and Park (2020) [31] tried to implement breast cancer detection by mitotic cell counting. A faster region CNN (FR-CNN) was developed to detect mitotic cells in the images based on ResNet-50 and DenseNet-201. The FR-CNN achieved an F1-score of 85.8% on a public dataset.

The above analysis shows that CAD systems for breast cancer detection using either clinical data or medical images are becoming more and more effective. However, we believe that the performance of automatic breast cancer detection classification can be further improved because most current methods are developed based on either deep CNN models or classical machine learning models with handcrafted features. This paper presented a novel and simple breast cancer detection model called spatial attention fusion network (SAFNet) based on USIs. The main contributions are three-fold:

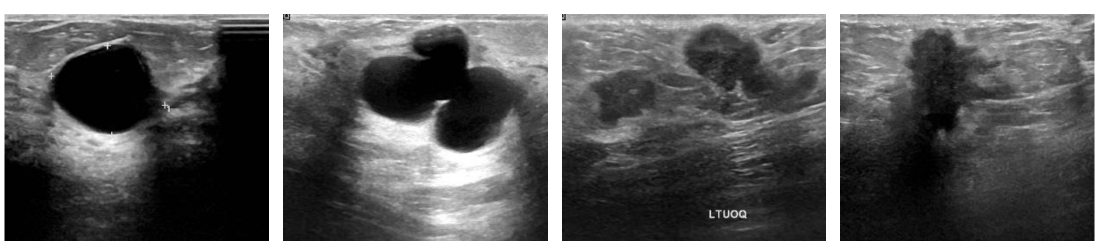
1) A spatial attention deep CNN model was designed based on the spatial attention module and the ResNet-18 backbone, pre-trained on the ImageNet dataset, and fine-tuned with the USIs for image feature extraction.
2) Three randomized neural networks (RNNs) were employed as the classifiers to prevent the overfitting problem because they are all shallow networks with simple structures and are easy and fast to train.
3) We proposed a late fusion mechanism to fuse the output labels using majority voting to stabilize the classification results of our breast cancer classification system.

The rest of this paper is arranged as follows. The introduction of the dataset for evaluation experiments is given in Section 2. The detailed presentation of the proposed SAFNet is demonstrated in Section 3. Section 4 presents the experiment results, and the discussion is demonstrated in Section 5. Finally, the conclusion of this study is provided in Section 6.

2. Materials

We utilized a public USI dataset to evaluate the generalization ability of our SAFNet [32]. The USI dataset can be downloaded from Kaggle (https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset). We finally obtained 437 benign USIs, 210 malignant USIs, and 166 normal USIs, which were approximately in size of 500 × 500 pixels. The ages of the women in the USIs ranged from 25 to 75. We focused on breast cancer detection in this study. We regarded both the benign and malignant samples as breast cancer USIs. Some samples in the USI dataset are shown in Figure 1.

Breast cancer



Normal

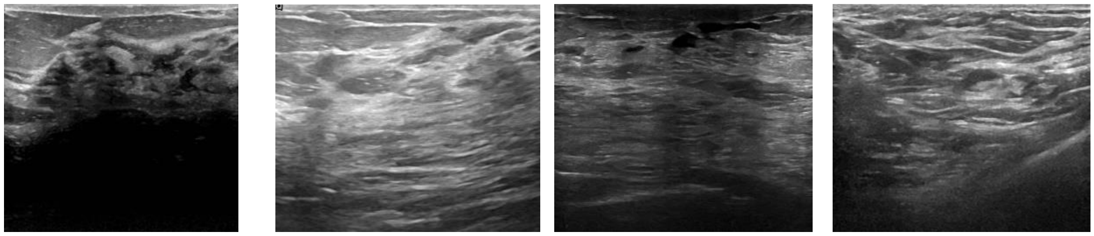


Figure 1: Samples in the USI dataset (The four USIs in the first row are breast cancer, and the four USIs in the second row are normal.)

3. Methodology

We proposed a spatial attention fusion network (SAFNet) to diagnose breast cancer in USIs. The SAFNet was designed using both deep learning and classical machine learning. The overview of the SAFNet is presented in Figure 2. Firstly, the backbone model for the SAFNet was a pre-trained ResNet-18, which was modified with a spatial attention module and fine-tuned on the USI dataset. Afterward, the fine-tuned backbone generated image representations using the training and testing USIs. Afterward, three RNNs were trained using the image representations and the labels, including ELM, random vector functional-link net (RVFL), and Schmidt neural network (SNN). Finally, the prediction labels of the SAFNet were calculated by the majority voting-based fusion of the predictions from the three RNNs. The spatial attention module can improve the image representation learning ability of the pre-trained ResNet-18 backbone. The RNNs were simple three-layered networks, which can effectively avoid the overfitting problem. The late fusion of the RNNs was designed to handle the negative effects of randomly initialized parameters in the RNNs and improve the classification performance of the SAFNet. 5-fold cross-validation was employed in our evaluation experiments.

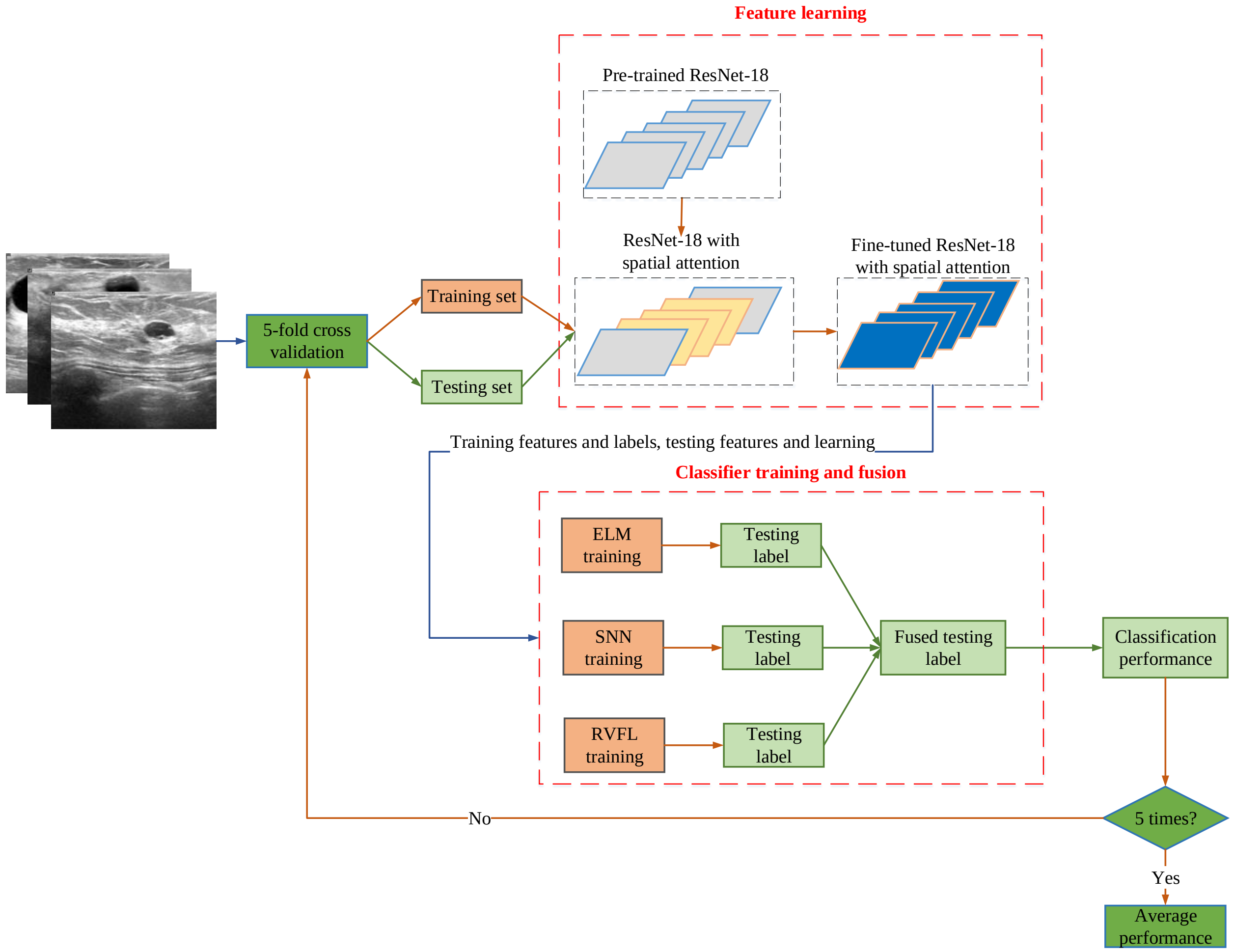


Figure 2: Overview of the proposed SAFNet for breast cancer detection (A pre-trained ResNet-18 was combined with the spatial attention module for representation learning, three RNNs were trained for classification, and the final output of the SAFNet was obtained by ensemble learning.)

3.1. Feature learning based on spatial attention network

During the recent ten years, deep CNN architectures can be the first choice for computer vision tasks. Among the various CNN models, ResNet can be one of the most significant inventions designed with residual connections to make it easy to approximate identity mappings [33]. The residual connection can directly link two layers that are not adjacent, which skips the hidden layers between the two layers, as is shown in Figure 3. Given $\mathbf{X}$ as the activation maps of the previous layer, the training target was $\mathrm{G}(\mathbf{X})$. with the residual connection, the target now becomes

$$\mathbf{F}(\mathbf{X}) := \mathbf{G}(\mathbf{X}) - \mathbf{X} \tag{1}$$

In this way, the original learning target can be expressed as

$$\mathbf{G}(\mathbf{X}) = \mathbf{F}(\mathbf{X}) + \mathbf{X} \tag{2}$$

The function F(**X**) is the residual between the training target and identity mapping. Instead of learning the target mapping directly, the model was trained to approximate the residual function so that the identity mappings could be trained more effectively.

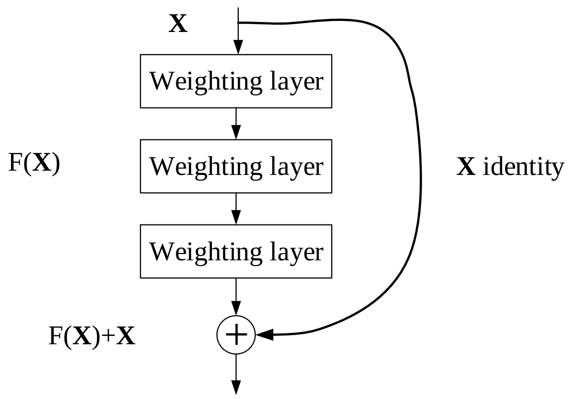


Figure 3: An example of a residual connection (A shortcut connection is added to skip the middle layers.)

The attention modules in the recent networks were inspired by human vision, which often focuses on specific regions of images. Researchers discovered that the attention mechanism could be embedded in current CNN models to boost their classification ability [34], and the implementation of attention modules can be plug-and-play [35]. Therefore, we proposed to embed the spatial attention module with the pre-trained ResNet-18 for feature learning from the breast USIs. A detailed presentation of the spatial attention module with ResNet-18 is illustrated in Figure 4. The spatial attention module was inserted between the rectified linear unit (ReLU) layer and the multiplication layer. In the spatial attention module, an average pooling layer and a max pooling layer were employed to generate two sets of feature maps. Then, the two groups of feature mappings were concatenated and passed to a convolution layer to fuse them together. Finally, a ReLU served as the activation function. The feature mappings of the spatial attention module were multiplied with the activation matrix of the top ReLU layer. We also modified the densely connected layers before the final layer of the ResNet-18 according to our breast USI dataset. The 'Fully connected layer 256' contained 256 nodes, which served as the feature layer in this study. As the breast cancer diagnosis is a binary classification problem, the dimension of the output layer was set as 2.

The pre-trained ResNet-18 with spatial attention mechanism was fine-tuned on the USI dataset for only four total epochs before the feature extraction from the 'Fully connected layer 256'.

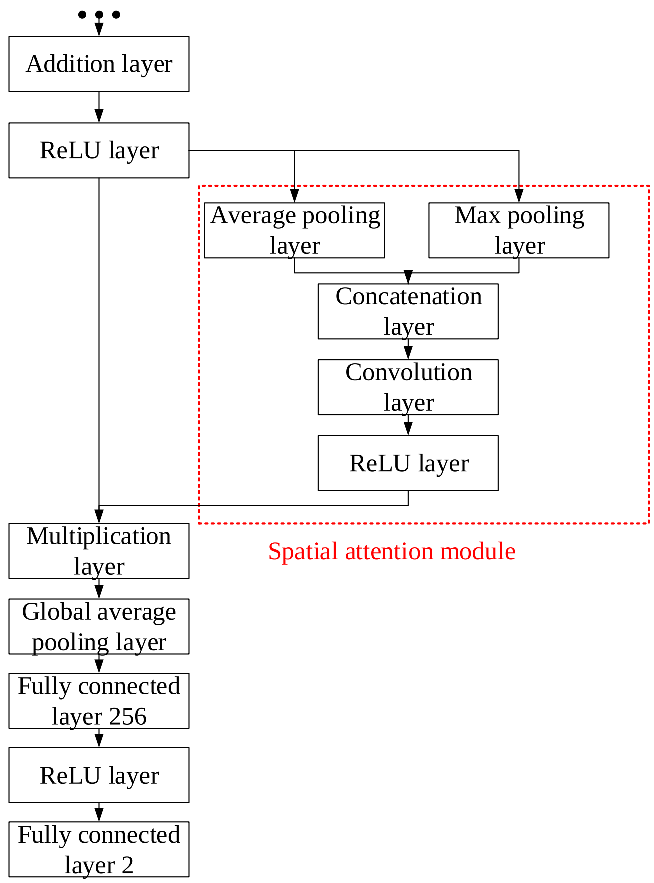


Figure 4: The spatial attention module with ResNet-18 (The spatial attention module was inserted into the ResNet-18. In The spatial attention module, the activations were processed with average pooling and max pooling, respectively. Then, the two activation maps were concatenated and sent into a convolution layer with ReLU activation function.)

3.2. Training and fusion of classifiers

As for the classifier in our SAFNet, we utilized three RNNs, including ELM [36], random vector functional-link (RVFL) [37], and Schmidt neural network (SNN) [38]. The three RNNs are all shallow networks with merely three layers, so the overfitting problem can be alleviated effectively, which often occurs when training deep CNN architectures with a small training set. Meanwhile, the training of RNNs is considerably faster compared with BPNNs benefiting from the randomized weights and biases in the hidden layer. Though RNNs can be trained fast, their classification performance is promising. Therefore, RNNs have been employed in a variety of machine learning tasks [39-41].

The architecture of an ELM is demonstrated in Figure 5. The parameters to be trained include the hidden weights $\boldsymbol{w}_i$, hidden biases b_i, and output weights $\boldsymbol{\beta}_i$. The training algorithm of an ELM consists of only three steps, as is given in Algorithm 1. Suppose we get the training representation from the fine-tuned backbone and their labels as

$$\boldsymbol{S_{train}} = [(\boldsymbol{f_1}, \boldsymbol{t_1}), (\boldsymbol{f_2}, \boldsymbol{t_2}), (\boldsymbol{f_3}, \boldsymbol{t_3}), \dots, (\boldsymbol{f_N}, \boldsymbol{t_N})] \tag{3}$$

where the $\boldsymbol{f_j} = (f_{j1}, f_{j2}, f_{j3}, \dots, f_{jn})^{\mathrm{T}}$ denotes the *j*-th image feature vector, the $\boldsymbol{t_j} = (t_{j1}, t_{j2}, t_{j3}, \dots, t_{jm})^{\mathrm{T}}$ represents the label of the *j*-th USI, and *N* stands for the entire number of training USIs. Then, the activation matrix of the hidden $\mathbf{H}_{\mathbf{act}}$ can be calculated as

$$\mathbf{H}_{\boldsymbol{act}} = \sum_{i=1}^{\widehat{N}} g\left(\boldsymbol{w_i} \boldsymbol{f_j} + b_i\right) = \begin{bmatrix} g(\boldsymbol{w_1}\boldsymbol{f_1} + b_1) & \cdots & g(\boldsymbol{w_{\widehat{N}}}\boldsymbol{f_1} + b_{\widehat{N}}) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w_1}\boldsymbol{f_N} + b_1) & \cdots & g(\boldsymbol{w_{\widehat{N}}}\boldsymbol{f_N} + b_{\widehat{N}}) \end{bmatrix}_{N \times \widehat{N}}, j = 1, \dots, N \tag{4}$$

in which the $g(x)$ denotes the activation function, and $\widehat{N}$ stands for the dimension of the hidden space. Finally, the prediction of the ELM is

$$\mathbf{O} = \mathbf{H}_{\boldsymbol{act}} \boldsymbol{\upbeta} \tag{5}$$

where $\boldsymbol{\upbeta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}, \dots, \boldsymbol{\beta_{\widehat{N}}})^{\mathrm{T}}$ is the output weights, and the $\mathbf{O} = (\boldsymbol{o_1}, \boldsymbol{o_2}, \boldsymbol{o_3}, \dots, \boldsymbol{o}_{\mathrm{N}})^{\mathrm{T}}$ denotes the prediction matrix of the ELM. The learning purpose is to make the predictions of the ELM equal to the ground-truth labels, so we can have

$$\mathbf{H}_{\boldsymbol{act}} \boldsymbol{\upbeta} = \mathbf{T} \tag{6}$$

where $\mathbf{T} = (\boldsymbol{t_1}, \boldsymbol{t_2}, \boldsymbol{t_3}, \dots, \boldsymbol{t_N})^{\mathrm{T}}$ means the ground-truth labels. In this way, the output weights $\boldsymbol{\upbeta}$ can be obtained based on the Moore-Penrose pseudo-inverse as

$$\boldsymbol{\upbeta} = \mathbf{H}_{\boldsymbol{act}}^{\dagger} \mathbf{T} \tag{7}$$

where the $\mathbf{H}_{\boldsymbol{act}}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{H}_{\boldsymbol{act}}$. With random initialization and pseudo-inverse, all the parameters in the ELM are decided within three steps. ELM training is much faster than vanilla BPNN, which is trained with iterations. Yet, the classification performance of the ELM is usually promising because it is likely to obtain smaller weight norms [42].

As a type of randomized neural network, ELM leverages the random projection to project the input features into the random hidden space, and only the output weights are trained. This random feature projection has also been used in deep networks in recent years [43]. ELM is a popular classification model in the last decade, which has been used in a variety of machine learning tasks, such as geography [44], big data analysis [45], clustering [46], chemistry [47], food safety [48], etc.

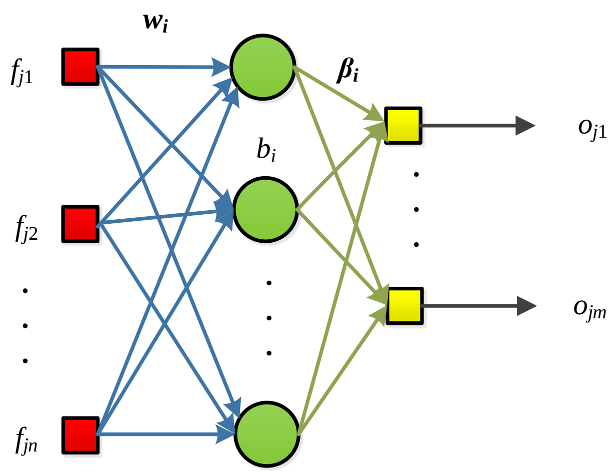


Figure 5: The architecture of an ELM

Algorithm 1: ELM training steps

Step 1: assign the hidden weights $\boldsymbol{w_i}$, hidden biases b_i with randomized values.

Step 2: generate the activation matrix of the hidden nodes using the training set.

Step 3: calculate the output weights $\boldsymbol{\beta_i}$ using the pseudo-inverse.

The architecture of an RVFL is demonstrated in Figure 6. There is an obvious difference between the architecture of RVFL and ELM. There is a shortcut connection between the input nodes and the output space [49]. In an RVFL, the input representations are randomly mapped into the hidden space (the green nodes), and the randomly mapped features are concatenated with the input feature vectors [50]. This extra shortcut connection from the input to the output layer can effectively stabilize the classification performance of the RVFL and improve the robustness of the system [51, 52].

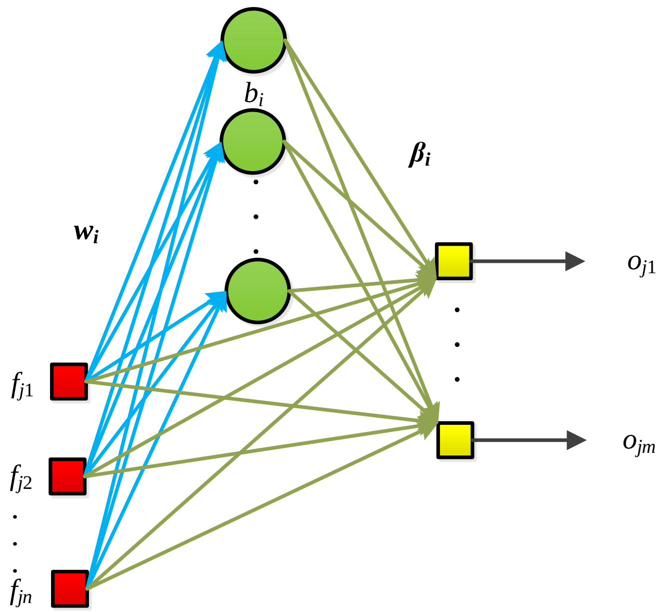


Figure 6: The structure of a RVFL

The training algorithm of RVFL is summarized in Algorithm 2. The activation matrix of the hidden layer with $\widehat{N}$ nodes can be formulated the same as equation (4). Then, the activation matrix of the hidden space is concatenated with the input features to generate the combined features **F**:

$$\mathbf{F} = concat(\mathbf{H}_{\boldsymbol{act}}, \mathbf{F}_{\boldsymbol{train}}) \tag{8}$$

where $\mathbf{F}_{\boldsymbol{train}} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \ldots, \boldsymbol{f}_\mathrm{N})^\mathrm{T}$ stands for the features from the feature layer in the backbone, and *concat*() is the matrix concatenation function. The training purpose is to achieve that the predictions of the RVFL are equal to the ground-truth labels:

$$\mathbf{O} = \mathbf{T} \tag{9}$$

where $\mathbf{O} = (\boldsymbol{o}_1, \boldsymbol{o}_2, \boldsymbol{o}_3, \ldots, \boldsymbol{o}_\mathrm{N})^\mathrm{T}$ is the predictions of the RVFL, and $\mathbf{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \ldots, \boldsymbol{t}_{\boldsymbol{N}})^\mathrm{T}$ means the ground-truth labels of the training set. Therefore, we have

$$\mathbf{F}\boldsymbol{\beta} = \mathbf{T} \tag{10}$$

Consequently, the output weights can be determined as:

$$\boldsymbol{\beta} = \mathbf{F}^\dagger\mathbf{T} \tag{11}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \ldots, \boldsymbol{\beta}_{\widehat{N}})^\mathrm{T}$ is the output weights, and $\mathbf{F}^\dagger$ represents the pseudo-inverse of the **F**.

Algorithm 2: RVFL training steps

Step 1: assign the hidden weights $\boldsymbol{w_i}$, hidden biases b_i with randomized values.

Step 2: compute the activation matrix of the hidden layer using the training set.

Step 3: concatenate the input features with the activation matrix of the hidden layer.

Step 4: determine the output weights $\boldsymbol{\beta_i}$ using the pseudo-inverse.

We also employ an SNN as the classifier in our SAFNet. SNN is a feedforward neural network proposed by Schmidt, Kraaijveld and Duin (1992) [38] thirty years ago. The structure of the SNN is almost the same as the ELM, which is shown in Figure 7. There are learnable output biases in the output layer, which is the only difference between the architecture of SNN and ELM. Therefore, the training algorithm of SNN is the same as ELM, which is illustrated in Algorithm 3. With the training features and labels, the prediction matrix of the SNN with $\widehat{N}$ nodes in the hidden layer is

$$\mathbf{O} = \left[\sum_{i=1}^{\widehat{N}} g\left(\boldsymbol{w_i f_j} + b_i\right)\right] \times concat(\boldsymbol{\beta}, \boldsymbol{C}) \tag{12}$$

where $\boldsymbol{C} = (c_1, c_2, c_3, \ldots, c_m)^{\mathrm{T}}$ is the output biases. The output weights and biases can be generated using close-formed solutions similar to the ELM and RVFL. In the original SNN, the standard numerical methods were employed to calculate the output weights and biases.

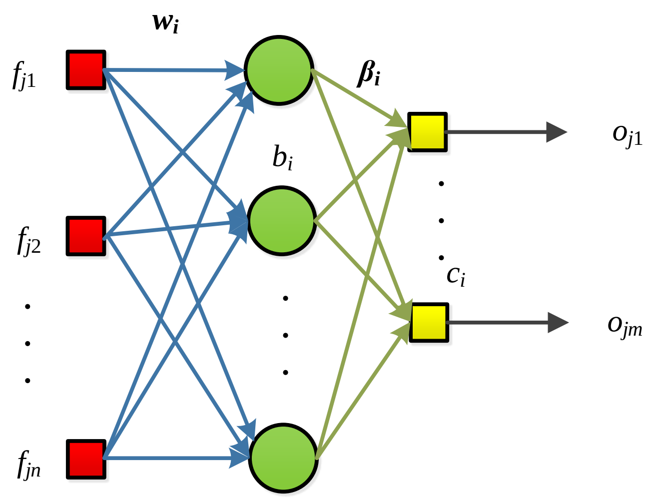


Figure 7: The structure of an SNN

Algorithm 3: SNN training steps

Step 1: assign the hidden weights $\boldsymbol{w_i}$, hidden biases b_i with randomized values.

Step 2: calculate the activation matrix of the hidden layer using the training set.

Step 3: determine the output weights $\boldsymbol{\beta_i}$ and biases c_i using the pseudo-inverse.

Now, we have presented three classifiers for breast cancer detection. The randomly initialized parameters in the RNNs can reduce the training time of our models, but ill-conditioned input weights and biases can worsen the classification performance of our breast cancer detection model, which can cause serious results in real-world applications. Therefore, we proposed to utilize late fusion to further improve our system's generalization performance and robustness, which directly fuses the predictions of the three RNNs by majority voting. Compared with swarm optimization methods, majority voting works more efficiently. The final model with this majority voting-based fusion is termed as SAFNet. The classifiers: ELM, RVFL, and SNN are heterogeneous RNNs, so the late fusion of their predictions can receive diversified information so that the fused results can be better.

4. Experimental results

Our SAFNet was developed using MATLAB 2021a. The training and testing experiments are run on a laptop with CPU i7 7700HQ, 16 GB memory, and GPU GTX1060 (6GB). The trained models can be deployed in other environments with appropriate configurations.

4.1. Hyper-parameter settings

The hyper-parameters in the proposed SAFNet are illustrated in Table 1. We set the mini-batch size as only 64 according to the capability of our GPU. The learning rate was set as 1×10^{-4}, which is a conventional value. The optimizer for fine-tuning the spatial attention backbone was Adam. We fine-tuned the backbone for only four epochs in order to prevent overfitting. The dimension of the hidden space in the three RNNs was 1000, as the feature dimension was 256. The random mapping from the input space to the hidden space of high dimension is beneficial to improve the classification performance.

Table 1: Hyper-parameters in the proposed SAFNet

Hyper-parameter	Value
Mini-batch size	64
Learning rate	1×10^{-4}
Optimizer	Adam
Max epochs	4
Number of hidden neurons in the three RNNs	1000

4.2. Classification performance

The classification performance of the proposed SAFNet is listed in Table 2. We employed sensitivity, precision, F1-score, and accuracy as the performance metrics in the evaluation experiments.

Table 2: Classification results of the SAFNet (F: fold)

	Sensitivity	Precision	F1-score	Accuracy
F1	93.38%	98.45%	95.85%	92.95%
F2	95.59%	100.00%	97.74%	96.15%
F3	93.94%	96.12%	95.02%	91.67%
F4	94.78%	98.45%	96.58%	94.23%
F5	96.95%	97.69%	97.32%	95.51%
Average	94.93%	98.14%	96.50%	94.10%

4.3. Effects of the spatial attention mechanism

To obtain the effects of the spatial attention mechanism, we tested the performance of the proposed SAFNet with and without the spatial attention module using 5-fold cross-validation. The backbone for the SAFNet without the spatial attention mechanism was the vanilla pre-trained ResNet-18. The results are shown in Table 3 and Figure 8.

Table 3: Effects of the spatial attention mechanism in the SAFNet

Model	Sensitivity	Precision	F1-score	Accuracy
SAFNet (without spatial attention)	94.68%	97.84%	96.22%	93.59%
SAFNet (with spatial attention)	94.93%	98.14%	96.50%	94.10%

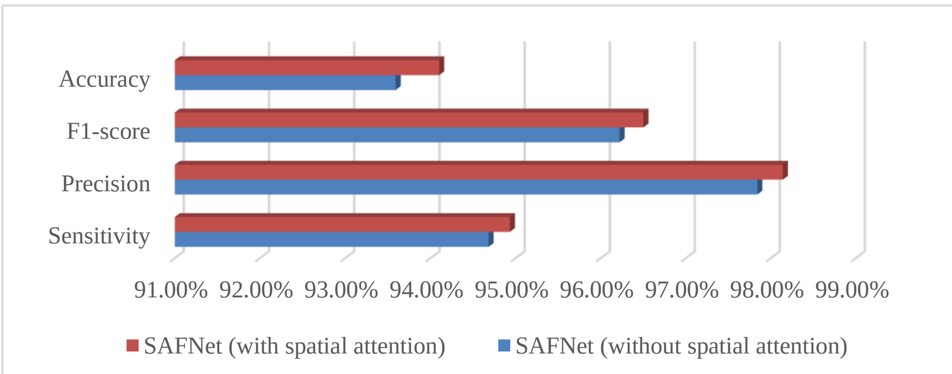


Figure 8: Effects of the spatial attention mechanism in the SAFNet

4.4. Effects of the late fusion

We also tested the effects of the late fusion of classifiers using 5-fold cross-validation. The classification results of the SAFNet with the ELM classifier, RVFL classifier, SNN classifier, and the fusion of the three RNNs are given in Table 4 and Figure 9.

Table 4: Effects of the late fusion of classifiers

Model	Sensitivity	Precision	F1-score	Accuracy
SAFNet (ELM)	94.24%	**98.30%**	96.22%	93.59%
SAFNet (SNN)	94.76%	96.76%	95.73%	92.82%
SAFNet (RVFL)	94.50%	97.99%	96.20%	93.59%
SAFNet (fusion of the three RNNs)	**94.93%**	98.14%	**96.50%**	**94.10%**

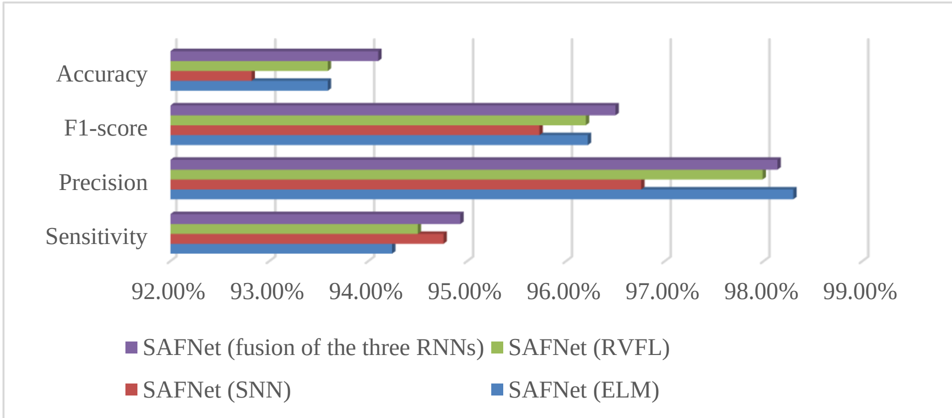


Figure 9: Effects of the late fusion of classifiers

4.5. Visual explanation

We utilized the Gradient-weighted class activation mapping (Grad-CAM) [53] for a visual explanation of the proposed SAFNet. Some Grad-CAMs of the breast cancer USIs based on the SAFNet were presented in Figure 10. The regions in red and orange are considered the most significant for the predictions by the SAFNet.

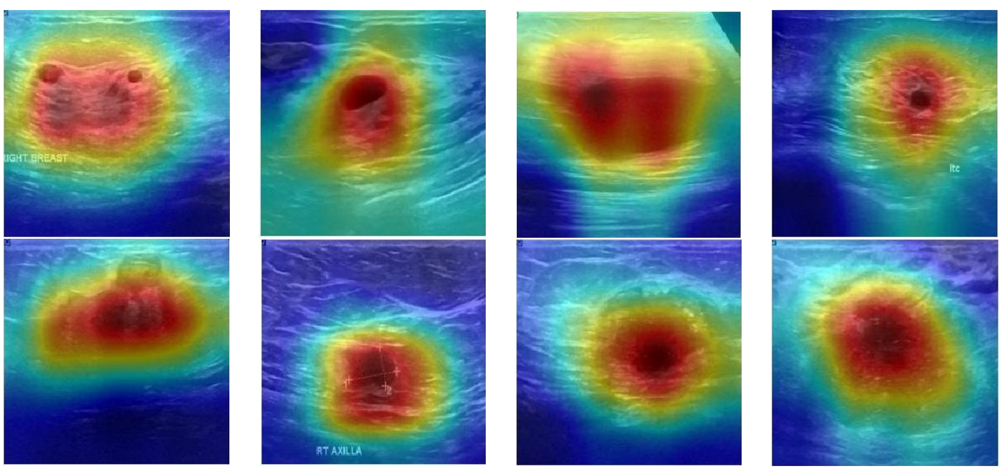


Figure 10: Grad-CAMs of the breast cancer USIs based on the SAFNet

4.6. Comparison with state-of-the-art methods

We compared our SAFNet with other existing breast cancer detection methods, and the statistics were presented in Table 5 and Figure 11. Our SAFNet produced better classification results than four existing breast cancer classification methods in terms of sensitivity, precision, F1-score, and accuracy.

Table 5: Comparison with other state-of-the-art breast cancer detection models

Model	Sensitivity	Precision	F1-score	Accuracy
CNN-GTD [15]	85.10%	-	-	86.50%
SeResNet18 [18]	90.00%	91.00%	91.00%	90.00%
FR-CNN [31]	-	87.60%	-	-
SD-CNN [8]	83.00%	-	-	90.00%
SAFNet (ours)	**94.93%**	**98.14%**	**96.50%**	**94.10%**

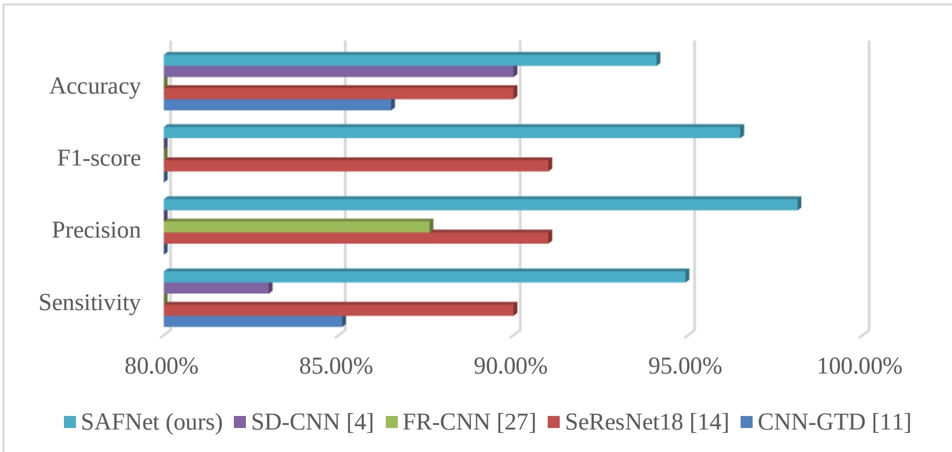


Figure 11: Comparison with other state-of-the-art breast cancer detection models

5. Discussion

The SAFNet achieved an average sensitivity of 94.93%, an average precision of 98.14%, an average F1 score of 96.50%, and an average accuracy of 94.10% for the 5-fold cross-validation. All four metrics fluctuated around 95%, which revealed the good generalization performance of the proposed SAFNet for breast cancer detection in USIs. Meanwhile, the running time for the 5-fold cross-validation of the SAFNet was merely 274.83 seconds, which was affordable for clinical applications.

We discovered that the spatial attention mechanism could boost the generalization ability of the SAFNet for breast cancer detection in terms of all four metrics, although the improvement was not considerable. It can be inferred from the results that the spatial attention module is an effective and simple method to boost the generalization ability of the proposed model for breast cancer diagnosis.

The SAFNet with the late fusion of the three RNNs achieved better sensitivity, F1-score, and accuracy than the SAFNets with the three individual RNNs. The precision of the SAFNet with the late fusion was 98.14%, which was also close to the best precision of 98.30% in the list. Therefore, we held the view that the late fusion mechanism can be beneficial for the generalization ability of the SAFNet for breast cancer detection in USIs.

From the Grad-CAMs, we can claim that our SAFNet can locate the potential lesion areas in the breast USIs, which contributes to the promising results.

The potential reasons for the outstanding generalization ability of the SAFNet include that the spatial attention module improves the representation learning capability of the pre-trained ResNet-18 backbone, the RNN classifiers can avoid the overfitting problem, and the late fusion of the classifiers can eliminate the bad effects of the random parameters in the RNNs.

6. Conclusion

This study put forward a novel breast cancer detection approach called SAFNet based on the ultrasound images. A pre-trained ResNet-18 was embedded with the spatial attention module to serve as the backbone model in the SAFNet to extract features from the breast USIs. We trained three RNNs as the classifiers and fused their predictions as the final output of the SAFNet. A public breast USI dataset was utilized to evaluate the generalization ability of the SAFNet based on 5-fold cross-validation. Extensive experiments were conducted, and our SAFNet outperformed four existing approaches in terms of sensitivity, precision, F1 score, and accuracy. The Grad-CAMs also revealed the high performance of the SAFNet to locate lesion areas. In all, our SAFNet is accurate in detecting breast cancer from USIs.

For future research, we shall collect more breast USIs as the number of normal USIs in the current dataset is small. In addition, we shall attempt to apply vision transformers for breast cancer detection. We will also try to harness optimization algorithms to train the RNNs [54-57]. Moreover, image segmentation shall be studied in the future [58-61].

Statements of ethical approval

This work does not contain any studies with human participants or animals performed by any authors.

Declaration of competing interests

The authors declare that they have no competing interests.

Funding

The paper is partially supported by Hope Foundation for Cancer Research, UK (RM60G0680), Royal Society International Exchanges Cost Share Award, UK (RP202G0230), Medical Research Council Confidence in Concept Award, UK (MC_PC_17171), British Heart Foundation Accelerator Award, UK (AA/18/3/34220), Sino-UK Industrial Fund, UK (RP202G0289), Global Challenges Research Fund (GCRF), UK (P202PF11), LIAS Pioneering Partnerships award, UK (P202ED10), Data Science Enhancement Fund, UK (P202RE237). Siyuan Lu holds the CSC scholarship with University of Leicester.

References

1. Amrane, M., S. Oukid, I. Gagaoua, and T. Ensari, *Breast Cancer Classification Using Machine Learning.* 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018

2. Roy, S., W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli, E. Peschiera, R. Trevisan, G. Maschietto, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R.J.G. van Sloun, E. Ricci, and L. Demi, *Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound.* IEEE Transactions on Medical Imaging, 2020. **39**(8): p. 2676-2687
3. Li, C., L. Hou, B.Y. Sharma, H. Li, C. Chen, Y. Li, X. Zhao, H. Huang, Z. Cai, and H. Chen, *Developing a new intelligent system for the diagnosis of tuberculous pleural effusion.* Computer methods and programs in biomedicine, 2018. **153**: p. 211-225
4. Su, H., D. Zhao, F. Yu, A.A. Heidari, Y. Zhang, H. Chen, C. Li, J. Pan, and S. Quan, *Horizontal and vertical search artificial bee colony for image segmentation of COVID-19 X-ray images.* Computers in Biology and Medicine, 2022. **142**: p. 105181
5. Chen, X., H. Huang, A.A. Heidari, C. Sun, Y. Lv, W. Gui, G. Liang, Z. Gu, H. Chen, C. Li, and P. Chen, *An efficient multilevel thresholding image segmentation method based on the slime mould algorithm with bee foraging mechanism: A real case with lupus nephritis images.* Computers in Biology and Medicine, 2022. **142**: p. 105179
6. Zhao, X., X. Zhang, Z. Cai, X. Tian, X. Wang, Y. Huang, H. Chen, and L. Hu, *Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients.* Computational biology and chemistry, 2019. **78**: p. 481-490
7. Rouhi, R., M. Jafari, S. Kasaei, and P. Keshavarzian, *Benign and malignant breast tumors classification based on region growing and CNN segmentation.* Expert Systems with Applications, 2015. **42**(3): p. 990-1002
8. Gao, F., T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, and B. Patel, *SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis.* Computerized Medical Imaging and Graphics, 2018. **70**: p. 53-62
9. Aslan, M.F., Y. Celik, K. Sabanci, and A. Durdu, *Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data.* International Journal of Intelligent Systems and Applications in Engineering, 2018. **6**(4): p. 289–293
10. Dai, B., R.-C. Chen, S.-Z. Zhu, and W.-W. Zhang, *Using Random Forest Algorithm for Breast Cancer Diagnosis*, in *2018 International Symposium on Computer, Consumer and Control (IS3C)*. 2018. p. 449-452.
11. Ghasemzadeh, A., S. Sarbazi Azad, and E. Esmaeili, *Breast cancer detection based on Gabor-wavelet transform and machine learning methods.* International Journal of Machine Learning and Cybernetics, 2018. **10**(7): p. 1603-1612
12. Gupta, M. and B. Gupta, *A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques.* Proceedings of the Second International Conference on Computing Methodologies and Communication (ICCMC 2018), 2018
13. Heidari, M., A.Z. Khuzani, A.B. Hollingsworth, G. Danala, S. Mirniaharikandehei, Y. Qiu, H. Liu, and B. Zheng, *Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm.* Physics in Medicine & Biology, 2018. **63**(3): p. 035020
14. Hussain, L., W. Aziz, S. Saeed, S. Rathore, and M. Rafique, *Automated Breast Cancer Detection Using Machine Learning Techniques by Extracting Different Feature Extracting Strategies*, in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. 2018. p. 327-331.
15. Wang, Z., M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, *Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features.* IEEE Access, 2019. **7**: p. 105146-105158
16. Islam, M.M., M.R. Haque, H. Iqbal, M.M. Hasan, M. Hasan, and M.N. Kabir, *Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques.* SN Computer Science, 2020. **1**(5)

17. Lahoura, V., H. Singh, A. Aggarwal, B. Sharma, M.A. Mohammed, R. Damasevicius, S. Kadry, and K. Cengiz, *Cloud Computing-Based Framework for Breast Cancer Diagnosis Using Extreme Learning Machine.* Diagnostics (Basel), 2021. **11**(2)
18. Zuluaga-Gomez, J., Z. Al Masry, K. Benaggoune, S. Meraghni, and N. Zerhouni, *A CNN-based methodology for breast cancer diagnosis using thermal images.* Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2021. **9**(2): p. 131-145
19. Rehman, O., H. Zhuang, A. Muhamed Ali, A. Ibrahim, and Z. Li, *Validation of miRNAs as Breast Cancer Biomarkers with a Machine Learning Approach.* Cancers (Basel), 2019. **11**(3)
20. Singh, D. and A.K. Singh, *Role of image thermography in early breast cancer detection- Past, present and future.* Comput Methods Programs Biomed, 2020. **183**: p. 105074
21. Stark, G.F., G.R. Hart, B.J. Nartowt, and J. Deng, *Predicting breast cancer risk using personal health data and machine learning models.* PLoS One, 2019. **14**(12): p. e0226765
22. Tapak, L., N. Shirmohammadi-Khorram, P. Amini, B. Alafchi, O. Hamidi, and J. Poorolajal, *Prediction of survival and metastasis in breast cancer patients using machine learning classifiers.* Clinical Epidemiology and Global Health, 2019. **7**(3): p. 293-299
23. Zheng, J., D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, *Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis.* IEEE Access, 2020. **8**: p. 96946-96954
24. Khuriwal, N. and N. Mishra, *Breast Cancer Detection From Histopathological Images Using Deep Learning.* 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 2018: p. 22-25
25. Kadam, V.J., S.M. Jadhav, and K. Vijayakumar, *Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax Regression.* Journal of Medical Systems, 2019. **43**(8): p. 263
26. Mercan, E., S. Mehta, J. Bartlett, L.G. Shapiro, D.L. Weaver, and J.G. Elmore, *Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions.* JAMA Network Open, 2019. **2**(8): p. e198777
27. Turkki, R., D. Byckhov, M. Lundin, J. Isola, S. Nordling, P.E. Kovanen, C. Verrill, K. von Smitten, H. Joensuu, J. Lundin, and N. Linder, *Breast cancer outcome prediction with tumour tissue images and machine learning.* Breast Cancer Research and Treatment, 2019. **177**(1): p. 41-52
28. Zeebaree, D.Q., H. Haron, A.M. Abdulazeez, and D.A. Zebari, *Machine learning and Region Growing for Breast Cancer Segmentation.* 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019: p. 88-93
29. Sharma, S. and R. Mehra, *Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images-a Comparative Insight.* Journal of Digital Imaging, 2020. **33**(3): p. 632-654
30. Zuluaga-Gomez, J., Z. Al Masry, K. Benaggoune, S. Meraghni, and N. Zerhouni, *A CNN-based methodology for breast cancer diagnosis using thermal images.* Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2020. **9**(2): p. 131-145
31. Mahmood, T., M. Arsalan, M. Owais, M.B. Lee, and K.R. Park, *Artificial Intelligence-Based Mitosis Detection in Breast Cancer Histopathology Images Using Faster R-CNN and Deep CNNs.* Journal of Clinical Medicine, 2020. **9**(3)
32. Al-Dhabyani, W., M. Gomaa, H. Khaled, and A. Fahmy, *Dataset of breast ultrasound images.* Data Brief, 2020. **28**: p. 104863
33. He, K., X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, IEEE: Las Vegas, NV, USA. p. 770-778.
34. Woo, S., J. Park, J.-Y. Lee, and I.S. Kweon, *CBAM: Convolutional Block Attention Module.* ECCV, 2018
35. Xiao, Y., H. Yin, S.H. Wang, and Y.D. Zhang, *TReC: Transferred ResNet and CBAM for Detecting Brain Diseases.* Frontiers in Neuroinformatics, 2021. **15**: p. 781551

36. Guang-Bin, H., Z. Qin-Yu, and S. Chee-Kheong, *Extreme learning machine: Theory and applications.* Neurocomputing, 2006. **70**(1-3): p. 489-501

37. Pao, Y.H., G.H. Park, and D.J. Sobajic, *Learning and generalization characteristics of random vector functional-link net.* Neurocomputing, 1994. **6**: p. 163-180

38. Schmidt, W.F., M.A. Kraaijveld, and R.P.W. Duin, *Feedforward neural networks with random weights*, in *Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems*. 1992. p. 1-4.

39. Vuković, N., M. Petrović, and Z. Miljković, *A comprehensive experimental evaluation of orthogonal polynomial expanded random vector functional link neural networks for regression.* Applied Soft Computing, 2018. **70**: p. 1083-1096

40. Zhang, Y., J. Wu, Z. Cai, B. Du, and P.S. Yu, *An unsupervised parameter learning model for RVFL neural network.* Neural Networks, 2019. **112**: p. 85-97

41. Wong, P.K., W. Huang, C.M. Vong, and Z. Yang, *Adaptive neural tracking control for automotive engine idle speed regulation using extreme learning machine.* Neural Computing and Applications, 2019

42. Bartlett, P.L., *The Sample Complexity of Pattern Classification with Neural Networks: The Size of the Weights is More Important than the Size of the Network.* IEEE TRANSACTIONS ON INFORMATION THEORY, 1998. **44**(2): p. 525-536

43. Baek, S., M. Song, J. Jang, G. Kim, and S.B. Paik, *Face detection in untrained deep neural networks.* Nature Communications, 2021. **12**(1): p. 7328

44. Huang, F., K. Yin, J. Huang, L. Gui, and P. Wang, *Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine.* Engineering Geology, 2017. **223**: p. 11-22

45. Deng, W.Y., Z. Bai, G.B. Huang, and Q.H. Zheng, *A Fast SVD-Hidden-nodes based Extreme Learning Machine for Large-Scale Data Analytics.* Neural Networks, 2016. **77**: p. 14-28

46. He, Q., X. Jin, C. Du, F. Zhuang, and Z. Shi, *Clustering in extreme learning machine feature space.* Neurocomputing, 2014. **128**: p. 88-95

47. Bian, X., C. Zhang, X. Tan, M. Dymek, Y. Guo, L. Lin, B. Cheng, and X. Hu, *A boosting extreme learning machine for near-infrared spectral quantitative analysis of diesel fuel and edible blend oil samples.* Analytical Methods, 2017. **9**(20): p. 2983-2989

48. Zhu, Z., X. Zhu, F. Kong, and W. Guo, *A rapid method on identifying disqualified raw goat's milk based on total bacterial count by using dielectric spectra.* Journal of Food Engineering, 2018. **239**: p. 40-51

49. Suganthan, P.N., *Letter: On non-iterative learning algorithms with closed-form solution.* Applied Soft Computing, 2018. **70**: p. 1078-1082

50. Bisoi, R., P.K. Dash, and S.P. Mishra, *Modes decomposition method in fusion with robust random vector functional link network for crude oil price forecasting.* Applied Soft Computing, 2019. **80**: p. 475-493

51. Zhang, L. and P.N. Suganthan, *Visual Tracking With Convolutional Random Vector Functional Link Network.* IEEE Transactions on Cybernetics, 2017. **47**(10): p. 3243-3253

52. Ren, Y., P.N. Suganthan, N. Srikanth, and G. Amaratunga, *Random vector functional link network for short-term electricity load demand forecasting.* Information Sciences, 2016. **367-368**: p. 1078-1093

53. Selvaraju, R.R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.* International Journal of Computer Vision, 2017. **128**: p. 336-359

54. Xia, J., Z. Wang, D. Yang, R. Li, G. Liang, H. Chen, A.A. Heidari, H. Turabieh, M. Mafarja, and Z. Pan, *Performance optimization of support vector machine with oppositional grasshopper optimization for acute appendicitis diagnosis.* Computers in Biology and Medicine, 2022: p. 105206

55. Xia, J., D. Yang, H. Zhou, Y. Chen, H. Zhang, T. Liu, A.A. Heidari, H. Chen, and Z. Pan, *Evolving kernel extreme learning machine for medical diagnosis via a disperse foraging sine cosine algorithm.* Computers in Biology and Medicine, 2022. **141**: p. 105137

56. Hu, J., z. Han, A.A. Heidari, Y. Shou, H. Ye, L. Wang, X. Huang, H. Chen, Y. Chen, and P. Wu, *Detection of COVID-19 severity using blood gas analysis parameters and Harris hawks optimized extreme learning machine.* Computers in Biology and Medicine, 2022. **142**: p. 105166

57. Hu, J., Y. Liu, A.A. Heidari, Y. Bano, A. Ibrohimov, G. Liang, H. Chen, X. Chen, A. Zaguia, and H. Turabieh, *An effective model for predicting serum albumin level in hemodialysis patients.* Computers in Biology and Medicine, 2022. **140**: p. 105054

58. Zhang, Q., Z. Wang, A.A. Heidari, W. Gui, Q. Shao, H. Chen, A. Zaguia, H. Turabieh, and M. Chen, *Gaussian Barebone Salp Swarm Algorithm with Stochastic Fractal Search for medical image segmentation: A COVID-19 case study.* Computers in Biology and Medicine, 2021. **139**: p. 104941

59. Liu, L., D. Zhao, F. Yu, A.A. Heidari, J. Ru, H. Chen, M. Mafarja, H. Turabieh, and Z. Pan, *Performance optimization of differential evolution with slime mould algorithm for multilevel breast cancer image segmentation.* Computers in Biology and Medicine, 2021. **138**: p. 104910

60. Liu, L., D. Zhao, F. Yu, A.A. Heidari, C. Li, J. Ouyang, H. Chen, M. Mafarja, H. Turabieh, and J. Pan, *Ant colony optimization with Cauchy and greedy Levy mutations for multilevel COVID 19 X-ray image segmentation.* Computers in Biology and Medicine, 2021. **136**: p. 104609

61. Yu, M., M. Han, X. Li, X. Wei, H. Jiang, H. Chen, and R. Yu, *Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: Thyroid ultrasound image case study.* Computers in Biology and Medicine, 2022. **144**: p. 105347