

ODET: Optimized Deep ELM-based Transfer Learning for Breast Cancer Explainable Detection

Ziquan Zhu¹, Shui-Hua Wang^{2,*}

¹1st Department of Civil Engineering, University of Florida, Gainesville, FL 32608, United States, zhu.ziquan@outlook.com

²2nd School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China, shuihuawang@ieee.org

Abstract

INTRODUCTION: Breast cancer is one of the most common malignant tumors in women, and the incidence rate is increasing year by year. Women in every country in the world may develop breast cancer at any age after puberty. The cause of breast cancer is not fully understood. At present, the main methods of breast cancer detection are inefficient. Researchers are trying to use computer technology to detect breast cancer. But there are some still limitations.

METHODS: We propose a network (ODET) to detect breast cancer based on ultrasound images. In this paper, we use ResNet50 as the backbone model. We make some modifications to the backbone model by deep ELM-based transfer learning. After these modifications, the network is named DET. However, DET still has some shortcomings because the parameters in DET are randomly assigned and will not change in the experiment. In this case, we select BA to optimize DET. The optimized DET is named ODET.

RESULTS: The proposed ODET gets the F1-score (F1), precision (PRE), specificity (SPE), sensitivity (SEN), and accuracy (ACC) are 93.16%±1.12%, 93.28%±1.36%, 98.63%±0.31%, 93.96%±1.85%, and 97.84%±0.37%, respectively.

CONCLUSION: It proves that the proposed ODET is an effective method for breast cancer detection.

Keywords: breast cancer; ResNet50; bat algorithm; extreme learning machine; ultrasound image;

Received on 30 June 2022, accepted on 05 September 2022, published on 29 September 2022

Copyright © 2022 Ziquan Zhu *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.v9i6.1747

*Corresponding author. Email: shuihuawang@ieee.org

1. Introduction

Breast cancer is one of the most common malignant tumors in women, and the incidence rate is increasing year by year. In Western Europe and North America, the incidence rate of breast cancer ranks first in female cancer. Women in every country in the world may develop breast cancer at any age after puberty. The cause of breast cancer is not fully understood. According to the recent research, the following factors may be related to the incidence of breast cancer: 1. breast cancer patients in the family; 2. early menarche or late cessation; 3. those who are old and unmarried or have no children after marriage; 4. no lactation after childbirth; 5. severe atypical hyperplasia of the breast; 6. abnormal

ovarian function; 7. excessive exposure to ionizing radiation.

The main methods of breast cancer detection are physical examination, blood tests, and imaging tests. These methods for breast cancer detection are inefficient. In the detection process, doctors are easily disturbed by other factors, such as doctors' lack of sleep, doctors' lack of professional knowledge, and so on.

With the rapid progress of artificial intelligence technology, researchers are trying to use computer technology to detect breast cancer. Gao, et al. [1] proposed a new model (SD-CNN) for the diagnosis of breast cancer. In this proposed model, features were extracted through the convolutional neural network (CNN). The proposed model obtained 90% accuracy and AUC=0.92. Wang, et al. [2] explored a method that was based on CNN. In their

experiment, CNN was the feature extraction, and ELM was used as the classifier. The CNN-GTD got 86.50% accuracy, 85.10% sensitivity, 88.02% specificity, 0.923 AUC. Rouhi, et al. [3] presented two models for the classification of breast tumors. The first model used a pre-trained ANN for the segmentation. In the second method, the segmentation was done by CNN. They got 96.87% sensitivity, 95.94% specificity, and 96.47% accuracy. Zuluaga-Gomez, et al. [4] explored a method for breast cancer diagnosis based on thermal images. The hyper-parameters of a CNN model were fine-tuned by a tree estimator. This method obtained 0.92 accuracy and 0.92 F1-score. Mahmood, et al. [5] presented a model to detect breast cancer. The presented model was based on the Faster R-CNN and deep CNNs and got 0.841 recall, 0.858 F1-measure, and 0.876 precision for the ICPR 2012 dataset, 0.583 recall, 0.691 F1-measure, and 0.848 precision for the ICPR 2014 dataset. Gupta and Chawla [6] combined CNN with two different machine learning methods to detect brain cancer. The two different machine learning methods were SVM and LR. In conclusion, CNN+LR got better performance than CNN+SVM.

Many researchers have used CNN to detect breast cancer from the above studies. Nevertheless, these methods have some disadvantages. In CNN, a small data set may lead to overfitting problems. However, it is difficult to have a large data set for medical image detection. This paper proposes the optimized deep ELM-based transfer learning network (ODET) to detect breast cancer based on ultrasound images. The main innovations are concluded as below:

- We propose an BA-guided optimized DET (ODET) model.
- We propose an BA-guided optimized DET (ODET) model.
- We validate ResNet50 is the best backbone model.
- This explainable ODET network is superior to six state-of-the-art models.

The rest sections of this paper are as follows. Section 2 is about the material. The methodology is presented in Section 3. Experiment results are shown in Section 4. The conclusion is given in Section 5.

2. Materials

We use the breast ultrasound images data set for this paper. This data set is public and available on the Kaggle website (Breast Ultrasound Images Dataset | Kaggle). This data set was completed in 2018 and mainly contained ultrasound images of breasts of women aged 25 to 75. This data set consisted of 780 images, and the pixel size was 500×500, which was saved in PNG. This data set was divided into normal, benign, and malignant categories. However, the data set is divided into two categories in this study. Benign and malignant images are unhealthy, and normal images are healthy. Some images of this public data set are given in Figure 1.

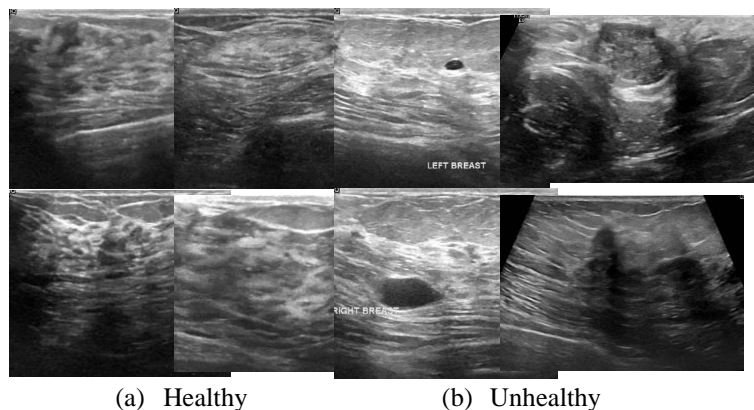


Figure 1. Some images of the public data set

3. Methodology

Table 1. Full Explanation and Acronym

Acronym	Full Explanation
Avr	Average
CNN	Convolutional neural network
ELM	Extreme learning machine
BA	Bat algorithm
FC	Fully connected
BN	Batch normalization
DCNN	Deep convolutional neural network
SLFN	Single hidden layer feedforward network
ACC	Accuracy
F1	F1-score
SPE	Specificity
SEN	Sensitivity
PRE	Precision
Std	Standard deviation
Grad-CAM	Gradient-weighted class activation mapping

Table 2. Full Explanation and Abbreviated Model

Abbreviated model	Full Explanation
ODET	The BA-guided optimized DET
DET	ResNet50 modified by deep ELM-based transfer learning
DETA	AlexNet modified by deep ELM-based transfer learning
DETD	DenseNet modified by deep ELM-based transfer learning
DETM	MobileNet modified by deep ELM-based transfer learning
DETR	ResNet18 modified by deep ELM-based transfer learning
DETV	VGG modified by deep ELM-based transfer learning
ODETA	The BA-guided optimized DETA
ODETD	The BA-guided optimized DETD
ODETM	The BA-guided optimized DETM
ODETR	The BA-guided optimized DETR
ODETV	The BA-guided optimized DETV

3.1. Proposed ODET

The acronym table and abbreviated model table are demonstrated in Table 1 and Table 2. Extracting effective features is the great significance for image classification [7]. Many researchers use computer technology to extract effective features from images [8]. With the continuous improvement of computer technology in image classification, many excellent computer models have been put forward [9], such as AlexNet [10], GoogleNet [11], VGG [12], DenseNet [13], MobileNet [14], and so on.

In this paper, the pre-trained model (ResNet50) [15] is selected as the backbone of our network. We make some modifications to the backbone model by deep ELM-based transfer learning. Firstly, we modify the pre-trained ResNet50 because there are only two output nodes in our experiment. Then, the last five layers of modified ResNet50 are replaced by ELM [16], which is the single hidden layer feedforward network (SLFN). After these modifications, the network is named DET. However, DET still has some shortcomings because the parameters in DET are randomly assigned and will not change in the experiment. In this case, we select the bat algorithm (BA) [17] to optimize DET. The optimized DET is named ODET. The pseudocode of ODET is demonstrated in Table 3. The architecture of our proposed ODET is presented in Figure 2.

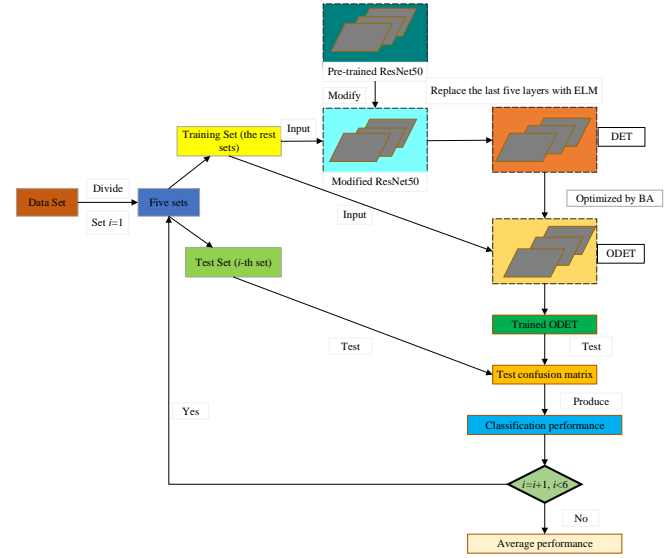


Figure 2. The architecture of our proposed ODET

Table 3. The pseudocode of ODET

- Step 1: The pre-trained ResNet50 is selected as the backbone of our model.
- Step 2: The data set is divided into five sets and set $i=1$.
 - Step 2.1: The i -th set is the test set and the rest sets as the training set.
- Step 3: Some modifications are made to the backbone model.
 - Step 3.1: The FC1000, softmax, and classification layer are deleted.
 - Step 3.2: Six layers are added, which are FC128, ReLU, BN, FC2, softmax, and classification layer.
 - Step 3.3: The last five layers of the modified ResNet50 are substituted by ELM.
 - Step 3.4 The network is named DET after these modifications:
- Step 4: The BA-guided Optimization process in the ODET
 - Step 4.1: The BA is selected to optimize the DET.
 - Step 4.2: The optimized DET is named ODET.
 - Step 4.3: The ODET is trained on the training set.
 - Step 4.3.1: The training set is input.
 - Step 4.3.2: The target is the labels of the training set.
 - Step 4.4: The trained ODET is tested on the test set.
 - Step 4.5: The test classification performance of the ODET is reported.
 - Step 4.6: Set $i=i+1$, if $i<6$, go to Step 2.
- Step 6: Average the test classification performance.

3.2. Comparison of Backbone Models

The more layers in the CNN model, the more features should be extracted [18]. However, if the depth of the CNN model is continuously increased, gradient dispersion or gradient explosion could occur [19]. Batch normalization (BN) is a solution to deal with these problems [20]. However, another problem is degradation, which is that with the increasing number of layers in the CNN model, the accuracy is declining. He, et al. [15] proposed residual learning to deal with this problem. The framework of residual learning is given in Figure 3.

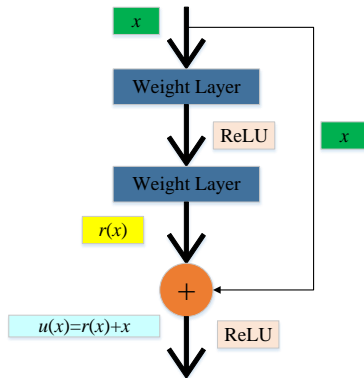


Figure 3. The framework of residual learning

The input is represented as x , the original target is denoted as $u(x)$, and the $r(x)$ is obtained through the residual function:

$$r(x) = u(x) - x. \quad (1)$$

The original target:

$$u(x) = r(x) + x. \quad (1)$$

There are some shortcut connections in the residual learning network [21]. These connections skip some layers and pass the original data directly to subsequent layers [22]. When the extreme case is encountered that the new layer does not learn any data, the new layer in the residual learning network can directly copy the original data. Therefore, the residual learning network can cope with the degradation problem [23]. In the experiment, we will compare ResNet50 with AlexNet, DenseNet, MobileNet, ResNet18, and VGG.

3.3. Proposed Deep ELM-based Transfer Learning

We make some modifications to the backbone model by deep ELM-based transfer learning [24], as shown on the left

side of Figure 4. Firstly, we modify the pre-trained ResNet50 because there are only two output nodes in our experiment. There are only two categories in this paper. Therefore, the FC1000, softmax, and classification layer of the pre-trained backbone model are substituted by six layers, which are FC128, ReLU, BN, FC2, softmax, and classification layer. We replace the last five layers of the modified ResNet50 with ELM. After these modifications, the network is called DET. The structure of ELM is presented on the right side of Figure 4. There are only three layers in the ELM: input, hidden, and output layers. The orange box is the input, the blue circle represents the hidden nodes in the hidden layer, and the yellow box presents the output.

There are only three training steps in ELM. Given a data set with N samples:

$$\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T \in R^n, i = 1, \dots, N. \quad (2)$$

$$\mathbf{k}_i = (k_{i1}, \dots, k_{im})^T \in R^m, i = 1, \dots, N. \quad (3)$$

Where the input dimension and the output dimension are represented by n and m , respectively.

The output matrix of the hidden layer is computed as:

$$\mathbf{M}_{\text{ELM}} = \sum_{j=1}^s b(\mathbf{o}_j \mathbf{x}_i + l_j), i = 1, \dots, N. \quad (4)$$

Where S is the number of hidden nodes in the hidden layer, b represents the sigmoid function, \mathbf{o}_j is the weight vector, which connects the input nodes with the j -th hidden node, and the bias of the j -th hidden node is presented by l_j .

Another expression of this formula is:

$$\mathbf{M}_{\text{ELM}} = \begin{bmatrix} b(\mathbf{o}_1 \mathbf{x}_1 + l_1) & \dots & b(\mathbf{o}_s \mathbf{x}_1 + l_s) \\ \vdots & \ddots & \vdots \\ b(\mathbf{o}_1 \mathbf{x}_N + l_1) & \dots & b(\mathbf{o}_s \mathbf{x}_N + l_s) \end{bmatrix}_{N \times S}. \quad (6)$$

Then the final output weight is calculated as:

$$\mathbf{h} = \mathbf{M}_{\text{ELM}}^+ \mathbf{k}. \quad (5)$$

Where the final output weight is given as \mathbf{h} , $\mathbf{M}_{\text{ELM}}^+$ is the pseudo-inverse matrix of the output matrix of the hidden layer (\mathbf{M}_{ELM}), and the label of the data set is represented as $\mathbf{k} = (\mathbf{k}_1, \dots, \mathbf{k}_N)^T$. The training of ELM is completed through three calculation steps. In recent years, the ELM has attracted more and more researchers' attention and has been widely used in various machine learning tasks, such as chemistry [25], geography [26], and so on.

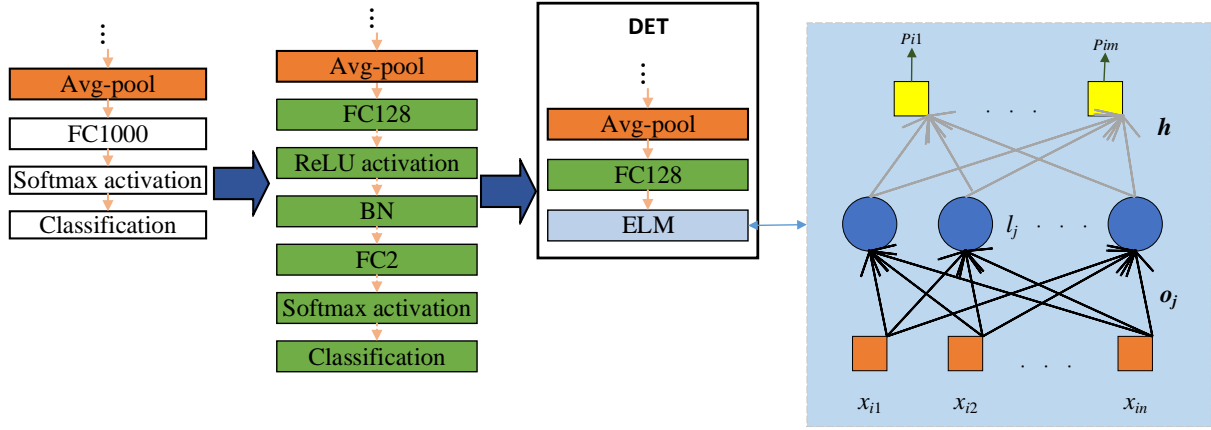


Figure 4. Illustration of DET

3.4. Proposed BA-guided Optimization

Although the proposed DET can reduce the computation and avoid overfitting problems, some parameters in the DET are set randomly and will not change, which also leads to some problems [27]. Poor weight and bias will lead to the decline of model performance [28]. For better classification performance, BA is selected for the optimization of DET. The optimized DET is named ODET. The BA was proposed based on the behaviour of bats. Bats sense distance through echolocation and distinguish the difference between food/prey and background barriers in some magical way. BA finds the best location through specific strategies based on bat behaviour. The parameters are continuously updated based on the optimal solution in each iteration. The equations of BA are as follows:

$$F = \sum_{i=1}^N (P_i - k_i)^2. \quad (6)$$

Where P_i is the output of ELM.

The ν -th bat gets the velocity e_ν^t and position d_ν^t when the time is t . The updated velocity and position are calculated as:

$$f_\nu = f_{\min} + (f_{\max} - f_{\min})\vartheta. \quad (7)$$

$$e_\nu^{t+1} = e_\nu^t + (d_\nu^t - d_*)f_\nu. \quad (8)$$

$$d_\nu^{t+1} = d_\nu^t + e_\nu^t. \quad (9)$$

Where the frequency of the ν -th bat is f_ν , f_{\max} and f_{\min} are the maximum frequency and minimum frequency, respectively, $\vartheta \in [0, 1]$ is a random variable, d_* is the current best location.

When a position is selected from the current optimal position, a new alternative position will be generated by the random walk:

$$d_{\text{new}} = d_{\text{old}} + \alpha w^t. \quad (10)$$

Where $\alpha \in [0, 1]$ is a random variable, w^t is the average loudness of the bat population when time is t .

When the ν -th bat locks the prey, the loudness w_ν^t and emission rate z_ν^t will be updated:

$$w_\nu^{t+1} = 0.9w_\nu^t. \quad (11)$$

$$z_\nu^t = z_\nu^0 (1 - e^{-0.99t}). \quad (12)$$

To better understand the optimization process, the pseudocode of BA-guided Optimization is given in Table 4.

Table 4. The pseudocode of the BA-guided Optimization

Step 1: Initialize the position d_ν , the velocity e_ν , the loudness w_ν , and the emission rate z_ν .
Step 2: While ($t <$ the maximum number of iterations).
Step 2.1: Generate new positions.
Step 2.2: Update the position and velocity.
Step 2.3: If ($\text{rand} > z_\nu$).
Step 2.3.1: Choose a position from the current optimal positions.
Step 2.3.2: Generate a new alternative position by the random walk.
Step 2.4: Generate a new position by flying randomly.
Step 2.5: If ($\text{rand} < w_\nu$ & $F(d_\nu) < F(d_*)$).
Step 2.5.1: Store the new position.
Step 2.5.2: Update the loudness and emission rate.
Step 2.6: Find the best d_* .
Step 3: Output the best d_* .

3.5. Evaluation

We define unhealthy breasts as the positive class and healthy breasts as the negative class. Five widely used measurements are calculated to evaluate the proposed ODET, which are F1-score (F1), precision (PRE), specificity (SPE), sensitivity (SEN), and accuracy (ACC). Their equations are given below:

$$\begin{cases} F1 = \frac{2 \times TP}{2TP + FP + FN} \\ PRE = \frac{TP}{TP + FP} \\ SPE = \frac{TN}{TN + FP} \\ SEN = \frac{TP}{TP + FN} \\ ACC = \frac{TP + TN}{TP + TN + FP + FN} \end{cases} \quad (13)$$

where FP, FN, TP, and TN are false positive, false negative, true positive, and true negative, respectively.

4. Experimental Results and Discussion

4.1. The classification performance of ODET

We use five-fold cross-validation to evaluate the proposed ODET. The classification performance is given in Table 5. The comparison figure of DET and ODET is given in Figure 5. From the figure, we can see that the classification performance is significantly improved after the BA-guided Optimization. The accuracy of ODET is higher than 97%. Meanwhile, other measurements are higher than 93%. Therefore, these results prove that the proposed ODET is an effective model to detect breast cancer.

Table 5. Classification performance of DET and ODET

Method	Fold	ACC	SPE	PRE	SEN	F1
ODET	F 1	98.10	98.86	94.34	94.34	94.34
	F 2	97.78	98.47	92.59	94.34	93.46
	F 3	97.46	98.86	94.12	90.57	92.31
	F 4	98.41	98.86	94.44	96.23	95.33
	F 5	97.46	98.09	90.91	94.34	92.59
	Avr	97.84	98.63	93.28	93.96	93.16
	Std	±0.37	±0.31	±1.36	±1.85	±1.12
DET	F 1	93.02	95.42	78.18	81.13	79.63
	F 2	93.97	98.09	88.64	73.58	80.41
	F 3	93.97	95.80	80.36	84.91	82.57
	F 4	95.56	97.33	86.79	86.79	86.79

F 5	93.65	96.18	81.13	81.13	81.13
Avr	94.03	96.56	83.02	81.51	82.11
Std	±0.84	±1.00	±4.00	±4.53	±2.53

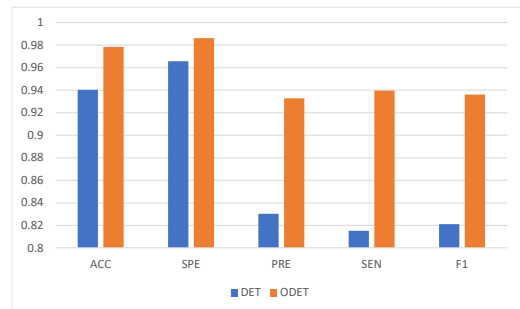


Figure 5. The comparison figure of DET and ODET

4.2. Comparison of different backbone models

Six CNN models are tested in this paper, which are ResNet50, AlexNet, DenseNet, MobileNet, ResNet18, and VGG. The classification performances of these backbone models are presented in Table 6. The comparison figure of different backbone models is shown in Figure 6. ResNet50 is the backbone model that achieves the best results than other backbones, except SPE and PRE. Nevertheless, these two measurements are only 1% less than the highest.

ResNet50 is a backbone model that can achieve such good results because residual learning can solve the problem of degradation. Compared with ResNet18, ResNet50 has more layers, which may extract more features. Therefore, it will achieve better results. AlexNet and VGG have too many parameters, which requires more epochs for aggregation. However, in this paper, the max-epoch is set to 1 based on the dataset size. MobileNet has fewer layers and cannot extract features well compared with Resnet50. DenseNet has no residual learning so it may encounter degradation problems.

Table 6. Classification performance of different backbone models

Backbone model	Fold	ACC	SPE	PRE	SEN	F1
ResNet50	F 1	98.10	98.86	94.34	94.34	94.34
	F 2	97.78	98.47	92.59	94.34	93.46
	F 3	97.46	98.86	94.12	90.57	92.31
	F 4	98.41	98.86	94.44	96.23	95.33
	F 5	97.46	98.09	90.91	94.34	92.59
	Avr	97.84	98.63	93.28	93.96	93.16
	Std	±0.37	±0.31	±1.36	±1.85	±1.12
DenseNet	F 1	96.52	98.86	93.75	84.91	89.11

	F 2	95.57	98.48	91.49	81.13	86.00
	F 3	95.57	98.86	93.33	79.25	85.71
	F 4	95.89	97.72	88.46	86.79	87.62
	F 5	95.57	99.24	95.35	77.36	85.42
	Avr	95.82	98.63	92.48	81.89	86.77
	Std	±0.37	±0.52	±2.35	±3.50	±1.40
MobileNet	F 1	96.20	98.10	90.20	86.79	88.46
	F 2	96.52	98.10	90.38	88.68	89.52
	F 3	96.84	97.72	89.10	92.45	90.74
	F 4	96.52	97.33	87.50	92.45	89.91
	F 5	95.25	96.20	82.76	90.57	86.49
	Avr	96.27	97.49	87.99	90.19	89.02
	Std	±0.54	±0.71	±2.81	±2.20	±1.47
ResNet18	F 1	95.25	96.58	83.93	88.68	86.24
	F 2	95.89	98.48	91.67	83.02	87.13
	F 3	95.89	96.96	85.71	90.57	88.07
	F 4	93.35	95.82	79.63	81.13	80.37
	F 5	93.99	96.58	82.69	81.13	81.90
	Avr	94.87	96.88	84.73	84.91	84.74
	Std	±1.03	±0.88	±4.00	±3.96	±3.03
VGG	F 1	96.19	98.09	90.20	86.79	88.46
	F 2	93.33	100.00	100.00	60.38	75.29
	F 3	88.25	93.89	66.67	60.38	63.37
	F 4	96.51	98.47	92.00	86.79	89.32
	F 5	83.49	99.62	66.67	3.77	7.14
	Avr	91.56	98.02	83.11	59.62	64.72
	Std	±5.00	±2.18	±1.38	±3.03	±3.03
AlexNet	F 1	97.78	98.86	94.23	92.45	93.33
	F 2	95.87	98.09	90.00	84.91	87.38
	F 3	97.46	99.62	97.87	86.79	92.00
	F 4	94.92	97.33	86.27	83.02	84.62
	F 5	92.06	100.00	100.00	52.83	69.14
	Avr	95.52	98.78	93.68	80.00	85.29
	Std	±2.06	±0.98	±5.02	±13.95	±8.67

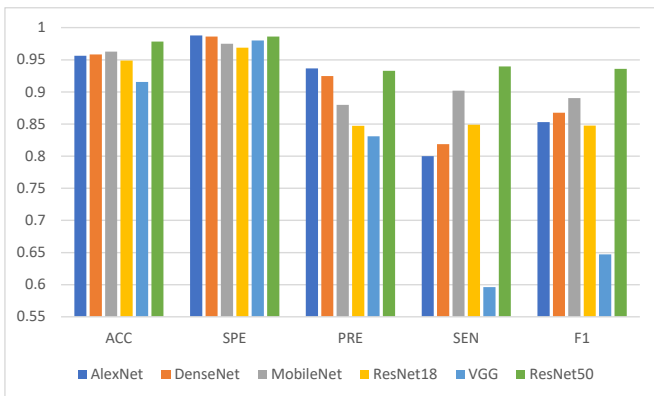


Figure 6. The comparison figure of different backbones

4.3. Effects of deep ELM-based transfer learning

We compare the classification performance of CNN models with CNN models modified by deep ELM-based transfer learning. To avoid contingency, we test six CNN models, which are ResNet50, AlexNet, DenseNet,

MobileNet, ResNet18, and VGG. These CNN models modified by deep ELM-based transfer learning are abbreviated as DET (ResNet50), DETA (AlexNet), DETD (DenseNet), DETM (MobileNet), DETR (ResNet18), and DETV (VGG). The experimental results of CNN models and CNN models modified by deep ELM-based transfer learning are presented in Table 7 and Table 8, respectively. The comparison figure is given in Figure 7. Based on the comparison results, we can see that the CNN models modified by deep ELM-based transfer learning can generally obtain better classification performance than the CNN models.

Table 7. Classification performance of CNN models

Model	Fold	ACC	SPE	PRE	SEN	F1
ResNet50	F 1	91.75	91.22	68.49	94.34	79.37
	F 2	93.33	93.89	75.00	90.57	82.05
	F 3	91.74	92.37	70.15	88.68	78.33
	F 4	94.92	96.18	82.46	88.68	85.45
	F 5	93.65	94.66	77.05	88.68	82.46
	Avr	93.08	93.66	74.63	90.19	81.53
	Std	±1.21	±1.73	±5.00	±2.20	±2.51
AlexNet	F 1	87.62	97.71	76.92	37.74	50.63
	F 2	83.17	100.00	-	0.00	-
	F 3	85.71	98.47	75.00	22.64	34.78
	F 4	83.81	100.00	100.00	3.77	7.27
	F 5	83.17	100.00	-	0.00	-
	Avr	84.70	99.24	83.97	12.83	30.90
	Std	±1.73	±0.97	±11.36	±15.02	±17.92
DenseNet	F 1	94.30	97.72	87.23	77.36	82.00
	F 2	93.35	96.20	80.77	79.25	80.00
	F 3	92.41	95.82	78.43	75.47	76.92
	F 4	93.67	96.20	81.13	81.13	81.13
	F 5	91.77	97.34	82.93	64.15	72.34
	Avr	93.10	96.66	82.10	75.47	78.48
	Std	±0.90	±0.74	±2.94	±5.97	±3.52
MobileNet	F 1	92.41	91.64	69.86	96.23	80.95
	F 2	87.98	86.69	58.82	94.34	72.46
	F 3	90.19	89.35	64.10	94.34	76.34
	F 4	91.14	92.02	68.66	86.79	76.67
	F 5	88.29	87.45	59.76	92.45	72.59
	Avr	90.00	89.43	64.24	92.83	75.80
	Std	±1.68	±2.14	±4.49	±3.25	±3.13
ResNet18	F 1	89.56	90.88	64.71	83.02	72.73
	F 2	92.72	95.06	76.79	81.13	78.90
	F 3	90.82	92.78	69.36	81.13	74.78
	F 4	89.87	92.02	66.67	79.25	72.41
	F 5	90.19	93.16	69.00	75.47	72.07
	Avr	90.63	92.78	69.30	80.00	74.18
	Std	±1.13	±1.38	±4.10	±2.56	±2.54
VGG	F 1	83.18	100.00	-	0.00	-
	F 2	83.18	100.00	-	0.00	-
	F 3	83.18	100.00	-	0.00	-
	F 4	92.70	93.13	72.73	90.57	80.67
	F 5	83.18	100.00	-	0.00	-

Avr	85.08	98.63	72.73	18.11	80.67
Std	±3.81	±2.75	±0.00	±36.23	±0.00

Table 8. Classification performance of CNN models modified by deep ELM-based transfer learning

Model	Fold	ACC	SPE	PRE	SEN	F1
DET	F 1	93.02	95.42	78.18	81.13	79.63
	F 2	93.97	98.09	88.64	73.58	80.41
	F 3	93.97	95.80	80.36	84.91	82.57
	F 4	95.56	97.33	86.79	86.79	86.79
	F 5	93.65	96.18	81.13	81.13	81.13
	Avr	94.03	96.56	83.02	81.51	82.11
	Std	±0.84	±1.00	±4.00	±4.53	±2.53
DETA	F 1	95.56	98.47	91.49	81.13	86.00
	F 2	93.65	98.86	92.31	67.92	78.26
	F 3	93.33	97.71	86.36	71.70	78.35
	F 4	92.06	96.18	79.17	71.70	75.25
	F 5	82.54	99.24	0.00	0.00	-
	Avr	91.43	98.09	69.87	58.49	79.46
	Std	±4.58	±1.08	±35.25	±29.57	±3.97
DETD	F 1	93.99	97.34	85.42	77.36	81.19
	F 2	93.35	98.48	90.00	67.92	77.42
	F 3	93.67	98.86	92.31	67.92	78.26
	F 4	94.94	99.24	95.12	73.58	82.98
	F 5	91.14	98.10	85.71	56.60	68.18
	Avr	93.42	98.40	89.71	68.68	77.61
	Std	±0.59	±0.71	±3.55	±4.00	±2.23
DETM	F 1	94.30	97.72	87.23	77.36	82.00
	F 2	93.67	96.20	81.13	81.13	81.13
	F 3	92.09	95.06	75.93	77.36	76.64
	F 4	92.06	96.18	76.36	79.25	77.78
	F 5	93.99	96.96	84.00	79.25	81.55
	Avr	93.29	96.20	80.93	78.87	79.82
	Std	±0.91	±1.09	±4.56	±1.56	±2.24
DETR	F 1	92.09	96.58	80.43	69.81	74.75
	F 2	93.67	98.48	90.24	69.81	78.72
	F 3	93.04	98.10	87.80	67.92	76.60
	F 4	92.41	96.96	82.22	69.81	75.51
	F 5	91.77	96.96	80.00	67.92	73.47
	Avr	92.59	97.34	84.14	69.06	75.81
	Std	±0.61	±0.78	±3.99	±0.82	±1.50
DETV	F 1	93.02	99.62	96.97	60.38	74.42
	F 2	83.17	100.00	-	0.00	-
	F 3	83.17	100.00	-	0.00	-
	F 4	96.19	98.47	91.84	84.91	88.24
	F 5	83.17	100.00	-	0.00	-
	Avr	87.75	99.62	94.40	29.06	81.33
	Std	±5.82	±0.63	±2.57	±37.34	±6.91

(- means no value)

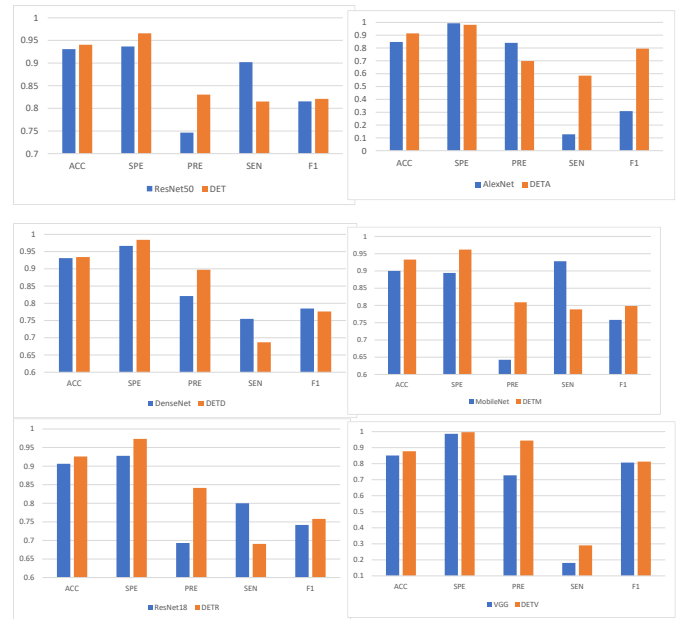
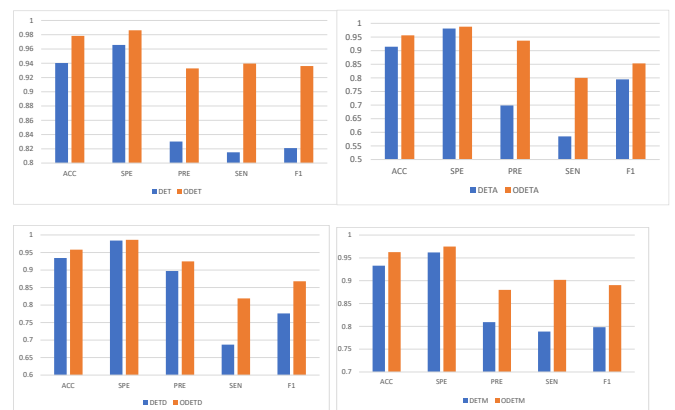


Figure 7. Effects of deep ELM-based transfer learning

4.4. Effects of the BA-guided Optimization

Six CNN models are selected to test the importance of BA-guided Optimization, which are ResNet50, AlexNet, DenseNet, MobileNet, ResNet18, and VGG. These CNN models modified by BA-guided Optimization and deep ELM-based transfer learning are abbreviated as ODET (ResNet50), ODETA (AlexNet), ODETD (DenseNet), ODETM (MobileNet), ODETR (ResNet18), and ODETV (VGG). The classification performance of CNN models modified by BA-guided Optimization and deep ELM-based transfer learning and CNN models modified by deep ELM-based transfer learning is shown in Table 6 and Table 8, respectively. For a more intuitive view, the comparison figure is presented in Figure 8. After the BA-guided Optimization, the classification performance of six different models is greatly improved. We can conclude that BA-guided Optimization can have a good influence on breast cancer detection.



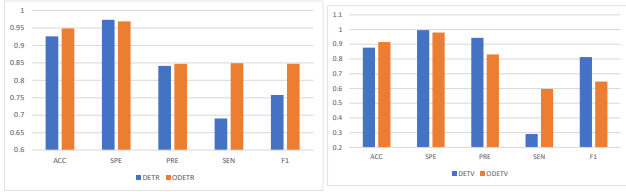


Figure 8. Effects of the BA-guided Optimization

4.5. Explainability of ODET

A deep convolutional neural network (DCNN) is like a black box. Researchers cannot figure out how DCNN makes classification and detection. In this paper, we use Gradient-weighted class activation mapping (Grad-CAM) [29] to explain the proposed ODET. The Grad-CAM has been applied for many fields, such as [30-32]. The raw images and Grad-CAM images are shown in Figure 9. In the Grad-CAM image, the greatest attention is represented by the red region. On the contrary, the blue region is the lowest attention. It can be discovered that breast cancers are within the red regions in the ODET. In conclusion, the ODET can capture breast cancers based on ultrasound images.

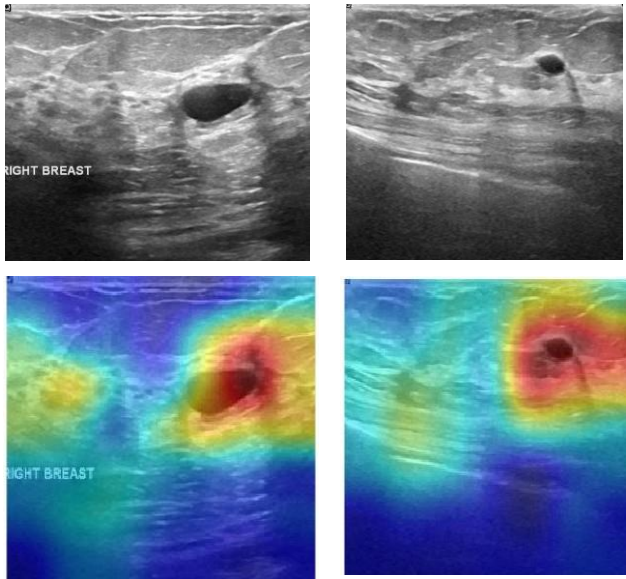


Figure 9. Explainability of the proposed ODET

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

We compare our proposed ODET with six other different methods, which are SD-CNN [1], CNN-GTD [2], GA-ANNs [3], SeResNet18 [4], Faster R-CNN+CNNs [5], and CNN+LR [6], respectively. Their results are listed in Table 9.

The comparison figure is given in Figure 10. In breast cancer detection, the method proposed in this paper is better than other state-of-the-art methods, which proves that our method is an effective method for breast cancer detection. The CNN model can capture high-level features. At the same time, we select the ELM as the classifier to prevent overfitting problems on small datasets.

Table 9. Comparison with state-of-the-art methods

Methods	ACC	SPE	PRE	SEN	F1
SD-CNN	90.00	94.00	-	83.00	-
CNN-GTD	86.50	88.02	-	85.10	-
GA-ANNs	96.47	95.94	-	96.87	-
SeResNet18	90.00	-	91.00	90.00	91.00
Faster R-CNN+CNNs	-	-	87.60	-	-
CNN+LR	96.87	-	86.93	-	82.64
ODET (Ours)	97.84	98.63	93.28	93.96	93.16

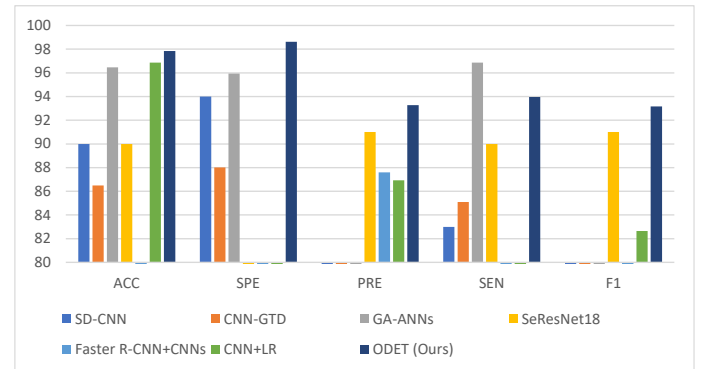


Figure 10. The comparison with different state-of-the-art methods

5. Conclusion

We propose a network (ODET) to detect breast cancer based on ultrasound images. In this paper, we use ResNet50 as the backbone model. We make some modifications to the backbone model by deep ELM-based transfer learning. After these modifications, the network is named DET. However, DET still has some shortcomings

because the parameters in DET are randomly assigned and will not change in the experiment. In this case, we select BA to optimize DET. The optimized DET is named ODET. The proposed ODET gets the F1-score (F1), precision (PRE), specificity (SPE), sensitivity (SEN), and accuracy (ACC) are $93.16\% \pm 1.12\%$, $93.28\% \pm 1.36\%$, $98.63\% \pm 0.31\%$, $93.96\% \pm 1.85\%$, and $97.84\% \pm 0.37\%$, respectively. The proposed ODET overperforms other state-of-the-art methods. It proves that our method is an effective method for breast cancer detection.

The proposed model yields great classification performance. But there are still some limitations to this paper. This paper has only the training and test sets and no validation set. In this paper, the public data set is classified into two categories, but this data set has three categories: normal, benign, and malignant. We should test different types of SLFN to compare the classification performance.

In the future, we will conduct experiments in more public breast cancer data sets to test the effectiveness of the proposed method. We will try to apply other methods [33, 34] to segment and classify breast cancer images, such as transformer, VIT, etc. Recently, many methods have proven their effectiveness, such as UNet, transformer, and so on. In the future, we will try these technologies in our study.

Author Contributions.

Ziquan Zhu: Conceptualization, Software, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Shui-Hua Wang: Methodology, Software, Validation, Investigation, Resources, Writing - Review & Editing, Supervision, Funding acquisition.

Funding.

The paper is partially supported by 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).

Data Availability Statement.

Publicly available datasets were analysed in this study. This data can be found here: Breast Ultrasound Images Dataset | Kaggle.

Conflicts of Interest.

The authors declare no conflict of interest.

References

- [1] **Journal article:** F. Gao *et al.*, "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 53-62, 2018.
- [2] **Journal article:** Z. Wang *et al.*, "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features," *IEEE Access*, vol. 7, pp. 105146-105158, 2019.
- [3] **Journal article:** R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990-1002, 2015.
- [4] **Journal article:** J. Zuluaga-Gomez, Z. Al Masry, K. Benagoune, 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*, vol. 9, no. 2, pp. 131-145, 2021.
- [5] **Journal article:** T. Mahmood, 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*, vol. 9, no. 3, p. 749, 2020.
- [6] **Journal article:** K. Gupta and N. Chawla, "Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN," *Procedia Computer Science*, vol. 167, pp. 878-889, 2020.
- [7] **Journal article:** B. Ay, C. Turker, E. Emre, K. Ay, and G. Aydin, "Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features," *Computers in Biology and Medicine*, vol. 147, p. 105725, 2022.
- [8] **Journal article:** Y. Zhang, W. Liu, X. Wang, and H. Gu, "A novel wind turbine fault diagnosis method based on compressed sensing and DTL-CNN," *Renewable Energy*, 2022.
- [9] **Conference:** J. Fang, H. Lin, X. Chen, and K. Zeng, "A Hybrid Network of CNN and Transformer for Lightweight Image Super-Resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1103-1112.
- [10] **Journal article:** A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [11] **Conference:** C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [12] **Journal article:** K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [13] **Conference:** G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [14] **Conference:** Z. Qin, Z. Zhang, X. Chen, C. Wang, and Y. Peng, "Fd-mobilenet: Improved mobilenet with a fast downsampling strategy," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018: IEEE, pp. 1363-1367.
- [15] **Conference:** K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [16] **Journal article:** G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.
- [17] **Journal article:** X.-S. Yang, "Bat algorithm for multi-objective optimisation," *arXiv preprint arXiv:1203.6571*, 2012.
- [18] **Conference:** P. Gavrikov and J. Keuper, "CNN Filter DB: An Empirical Investigation of Trained Convolutional

- Filters," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19066-19076.
- [19] **Journal article:** Y. Shi, Z. Wang, X. Du, G. Ling, W. Jia, and Y. Lu, "Research on the membrane fouling diagnosis of MBR membrane module based on ECA-CNN," *Journal of Environmental Chemical Engineering*, vol. 10, no. 3, p. 107649, 2022.
- [20] **Journal article:** J. Miettinen *et al.*, "Whitening CNN-Based Rotor System Fault Diagnosis Model Features," *Applied Sciences*, vol. 12, no. 9, p. 4411, 2022.
- [21] **Conference:** L. Shen, M. Ziaeefard, B. Meyer, W. Gross, and J. J. Clark, "Conjugate Adder Net (CAddNet)-A Space-Efficient Approximate CNN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2793-2797.
- [22] **Journal article:** X. Miao, Y. Zhang, J. Zhang, and X. Liang, "Hierarchical CNN Classification of Hyperspectral Images Based on 3D Attention Soft Augmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022.
- [23] **Journal article:** M. M. Althobaiti *et al.*, "Deep Transfer Learning-Based Breast Cancer Detection and Classification Model Using Photoacoustic Multimodal Images," *BioMed Research International*, vol. 2022, 2022.
- [24] **Journal article:** M. A. Ozdemir, D. H. Kisa, O. Guren, and A. Akan, "Hand gesture classification using time-frequency images and transfer learning based on CNN," *Biomedical Signal Processing and Control*, vol. 77, p. 103787, 2022.
- [25] **Journal article:** X. Hu, M. Hu, and X. Yang, "A Novel Fault Diagnosis Method for TE Process Based on Optimal Extreme Learning Machine," *Applied Sciences*, vol. 12, no. 7, p. 3388, 2022.
- [26] **Journal article:** A. Mohaghegh, S. Farzin, and M. V. Anaraki, "A new framework for missing data estimation and reconstruction based on the geographical input information, data mining, and multi-criteria decision-making; theory and application in missing groundwater data of Damghan Plain, Iran," *Groundwater for Sustainable Development*, vol. 17, p. 100767, 2022.
- [27] **Journal article:** G. Ahmed, T. Chu, and K. Loo, "Novel Multicolumn Kernel Extreme Learning Machine for Food Detection via Optimal Features from CNN," *arXiv preprint arXiv:2205.07348*, 2022.
- [28] **Journal article:** L. Boussaad and A. Boucetta, "Extreme Learning Machine-Based Age-Invariant Face Recognition With Deep Convolutional Descriptors," *International Journal of Applied Metaheuristic Computing (IJAMC)*, vol. 13, no. 1, pp. 1-18, 2022.
- [29] **Conference:** R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.
- [30] **Conference:** S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.
- [31] **Conference:** L. Chen, J. Chen, H. Hajimirsadeghi, and G. Mori, "Adapting grad-cam for embedding networks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2794-2803.
- [32] **Conference:** A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018: IEEE, pp. 839-847.
- [33] **Journal article:** Y. Zhang and L. Wu, "Improved image filter based on SPCNN," *Science in China Series F: Information Sciences*, vol. 51, no. 12, pp. 2115-2125, 2008.
- [34] **Journal article:** Y. Zhang and L. Wu, "Segment-based coding of color images," *Science in China Series F: Information Sciences*, vol. 52, no. 6, pp. 914-925, 2009.