IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 16, 2021, accepted May 4, 2021, date of publication May 11, 2021, date of current version May 19, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3079204*

# A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique

**ABEER SABER[1], MOHAMED SAKR[2], OSAMA M. ABO-SEIDA[1], ARABI KESHK[2], AND HUILING CHEN[3], (Associate Member, IEEE)**

[1]Department of Computer Science, Faculty of Computers and Information, Kafr El-Sheikh University, Kafr El-Sheikh 33511, Egypt
[2]Department of Computer Science, Faculty of Computers and Information, Menoufia University, Menoufia 32511, Egypt
[3]Department of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China

Corresponding author: Huiling Chen (chenhuiling.jlu@gmail.com)

**ABSTRACT** Breast cancer (BC) is one of the primary causes of cancer death among women. Early detection of BC allows patients to receive appropriate treatment, thus increasing the possibility of survival. In this work, a new deep-learning (DL) model based on the transfer-learning (TL) technique is developed to efficiently assist in the automatic detection and diagnosis of the BC suspected area based on two techniques namely 80-20 and cross-validation. DL architectures are modeled to be problem-specific. TL uses the knowledge gained during solving one problem in another relevant problem. In the proposed model, the features are extracted from the mammographic image analysis- society (MIAS) dataset using a pre-trained convolutional neural network (CNN) architecture such as Inception V3, ResNet50, Visual Geometry Group networks (VGG)-19, VGG-16, and Inception-V2 ResNet. Six evaluation metrics for evaluating the performance of the proposed model in terms of accuracy, sensitivity, specificity, precision, F-score, and area under the ROC curve (AUC) has been chosen. Experimental results show that the TL of the VGG16 model is powerful for BC diagnosis by classifying the mammogram breast images with overall accuracy, sensitivity, specificity, precision, F-score, and AUC of 98.96%, 97.83%, 99.13%, 97.35%, 97.66%, and 0.995, respectively for 80-20 method and 98.87%, 97.27%, 98.2%, 98.84%, 98.04%, and 0.993 for 10-fold cross-validation method.

**INDEX TERMS** Breast cancer, machine learning, deep-learning, transfer learning, image classification, convolutional neural networks.

## I. INTRODUCTION

Cancer tumor is related to abnormal cell growth, which invades the surrounding tissues in the human body. There are two types of tumor: benign and malignant. A benign tumor consists of non–cancerous cells that grow only locally and do not spread in the human body. in contrast, a malignant tumor consists of cancerous cells, which are capable of multiplying uncontrollably, spreading to various parts of the human body, and invading the tissues. In the USA, approximately 12% of women are expected to be diagnosed with BC over their lifetime. On average, one woman every two minutes is diagnosed with BC in the USA [1], [2]. This makes BC the most common type of cancer in women [3]. BC is a disease in which breast cells grow uncontrollably. The BC type depends on the cells that become cancerous. BC can start in various parts of the breast. Breast consists of three main parts: lobes, ducts, and connective tissue. Most BCs start in the ducts or lobules. Therefore, early BC detection is significant in increasing patient survival rates. The high morbidity and considerable cost of healthcare-associated with cancer have instigated researchers to implement more precise models for cancer detection. Mammography and biopsy are the two most common methodologies for BC detection. In mammography, radiologist uses a specific type of breast images to detect early symptoms of cancer in women. Studies have shown that mammography has led to a reduction in death rates caused by BC. A biopsy is another efficient diagnostic methodology for BC detection. Automatic identification and localization of cancer cells are the main challenges in BC images due to their variance in size, shape, and location. Other abnormalities, such as mastitis, adenopathy, and granuloma, may also be

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

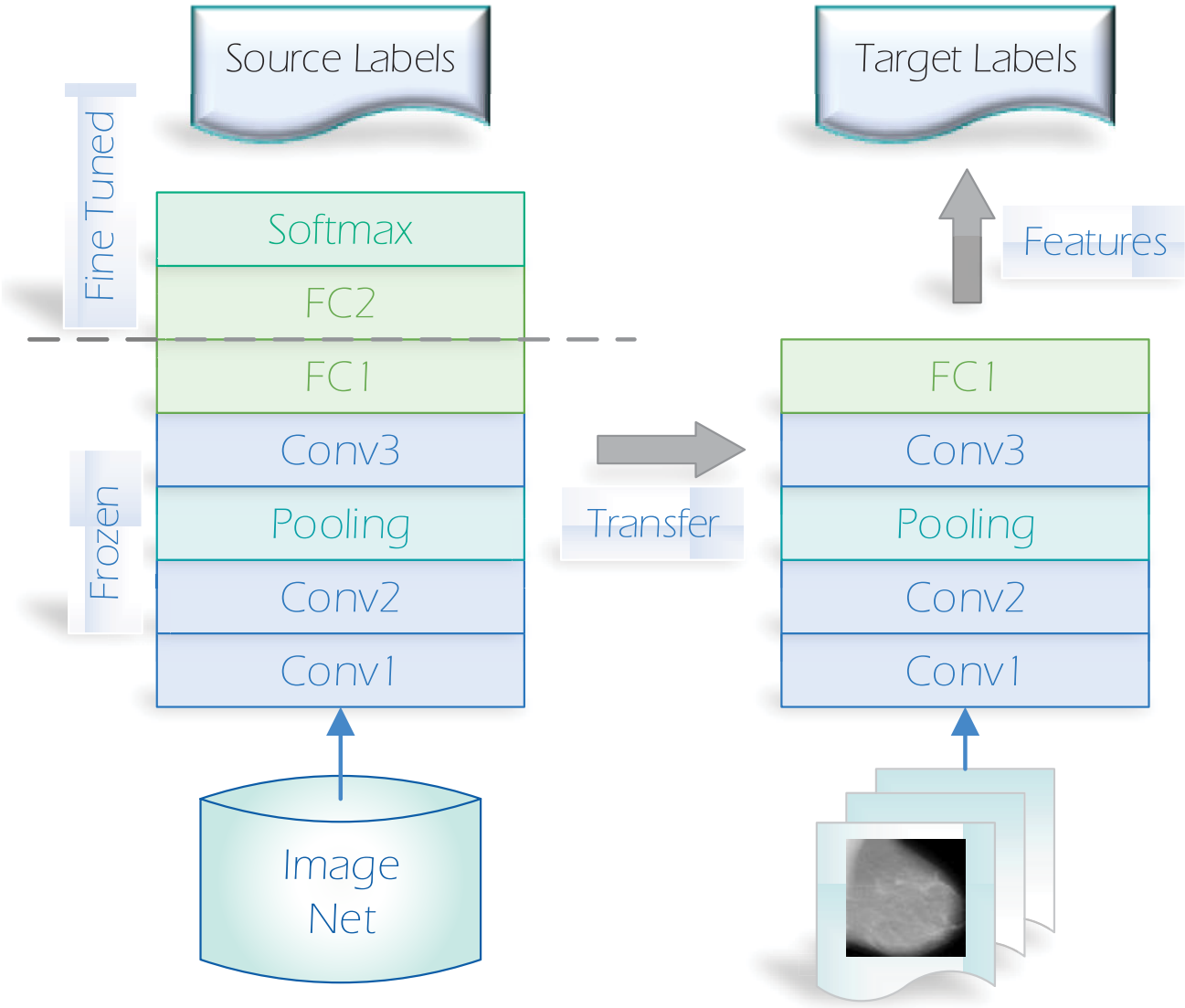


**FIGURE 1.** Transfer learning method.

found in breast images [4]. Machine learning (ML) techniques have found its wide applications in many fields such as prediction problems in educational field [5]–[9], bankruptcy prediction [10]–[16], pattern recognition [17]–[28], image editing [29]–[39], feature reduction [40]–[44], fault diagnosis [45]–[50], face recognition and micro-expression recognition [51]–[57], natural language processing [58], [59] and medical diagnosis [60]–[74]. Especially, it has found its great potential in BC diagnosis. In recent decades, various solutions for automatic cell classification in BC detection have been suggested by many researchers. In this context, some researchers have worked on nucleus analysis by extracting nucleus features that represent useful information in classifying cells into benign or malignant [75]. Similarly, grouping-based algorithms using the circular Hough-transform and various statistical features have also been exploited for nuclei segmentation and classification. However, due to the complex nature of classic ML techniques, such as preprocessing, segmentation, feature extraction, and other, the system's performance degrades in terms of efficiency and accuracy. Traditional ML challenges can be overcome by the DL method, which has emerged recently. This method is capable of achieving outstanding feature representation to solve image-classification and object-localization tasks. The most popular of the DL algorithms proposed in the literature are the CNNs. The CNN architecture is specially modified with the 2D input-image structure [76], [77]. A CNN-training task requires a large amount of data, which lack in the medical domain, especially in BC. A solution to this problem is to use the TL technique from a natural-images dataset, such as ImageNet, and implement a fine-tuning technique, as shown in Fig. 1. The TL concept can be exploited to enhance the performance of individual CNN architectures by combining their knowledge [78]. The major advantage of TL

is the enhancement of classification accuracy and the speed-up of the training process. An appropriate TL method is a model transfer; first, the network parameters are pre-trained using the source data, then these parameters are applied in the target domain, and finally the network parameters are adjusted for better performance [79]. In this context, a framework for multi-class BC detection and classification based on TL is proposed and implemented. The proposed model consists of two main components. The first component consists of six main phases (noise removal, histogram equalization, morphological analysis, segmentation, image resizing, data splitting, and data augmentation), which are applied to improve the breast images. Then, a pre-trained CNN such as, the Inception V3, VGG19, VGG16, ResNet50, and Inception-V2 ResNet, are used to transfer their learned parameters to the BC-classification task. The major objectives of this work are the automatic extraction of the affected patch using segmentation, reduction in training time, and improvement in classification performance.

This paper has the following contributions:

1) Reducing training time by extracting only the affected regions from breast images.

2) Using noise reduction, histogram equalization, and morphological analysis methods to improve the affected areas detection.

3) Improving the classification performance by changing the pre-trained networks classifier.

4) Solving the problem of overfitting.

Other contributions of this paper as follows:

- DL is introduced to help in BC automatic diagnosis.
- Compared between many pre-trained CNN such as Inception V3, ResNet50, VGG-16, VGG-19, and Inception-V2 ResNet results.
- Six different measures are used as accuracy, sensitivity, specificity, precision, AUC, and F-score.

This paper is organized as follows. In Section II, the related work is discussed, whereas a description of the proposed model for BC detection and classification using TL techniques is presented in Section III. The experimental results compared with real data are presented in Section IV. Finally, the paper is concluded in Section V.

## II. RELATED WORK

Ting *et al.* [80] implemented a deep CNN for BC-lesion classification. This network consisted of 1 input layer, 28 hidden layer, and 1 output layer. Overfitting was avoided using the feature-wise-data augmentation (FWDA) algorithm. Their proposed method sequentially achieved 89.47%, 90.50%, and 90.71% for sensitivity, accuracy, and specificity, respectively. Toğçar *et al.* [81] proposed the BreastNet, which consisted of convolutional, pooling, residual, and dense blocks, and it was capable of extracting the most effective features from breast images. BreastNet achieved better results than AlexNet, VGG-16, and VGG-19 models as its accuracy approached 98.80%. Abbas [82] presented a multi-layer DL architecture for classifying benign and malignant regions in breast images. This network consisted of four phases for extracting invariant features, which were transformed into deep-invariant features, and learning features for making the final decision. In [82], the MIAS dataset was used and achieved a 92%, 84.2%, 91.5%, and 0.91 for sensitivity, specificity, accuracy, and AUC, respectively. Using the same dataset, Sha *et al.* [83] presented a method for automatic detection and classification of the cancerous region in breast images. Their proposed method was based on CNNs and the grasshopper optimization algorithm. The results showed that this proposed method was capable of achieving 96%, 93%, and 92% for sensitivity, specificity, and accuracy, respectively. Charan *et al.* [84] trained a CNN for BC detection. Their proposed CNN consisted of six convolution layers, four average-pooling layers, and three fully-connected layers (FCLs). They used a size of 224 × 224 for the input image and the Softmax (SM) function to apply the classification results. The overall accuracy of this network was 65%, which was obtained using the MIAS database. In [85], Wahab *et al.* exploited a pre-trained CNN and transferred its learned parameters to another CNN for mitoses classification. Their proposed method achieved 0.50, 0.80, and 0.621 for precision, recall, and F-measure, respectively. In addition, for multi-class BC-classification purposes, Lotter *et al.* [75] proposed a model in which the features were extracted using a pre-trained ResNet50 network. Their model was capable of classifying lesions into five classes: mass, calcifications, focal asymmetry, architectural distortion, or no lesion. Their model achieved 96.2, 90.9, and 0.94 for sensitivity, specificity, and AUC, respectively. Jiang *et al.* [86] achieved better BC-classification accuracy in the case of TL from a pre-trained network in building networks from scratch. The accuracy approached 0.88 using GoogleNet and 0.83 using AlexNet on the film mammography number 3 (BCDR-F03) dataset. Khan *et al.* [87] implemented a model in which the breast-image features were extracted using pre-trained CNN architectures, namely, GoogleNet, VGGNet, and ResNet. The model's accuracy, which approached 97.525%, was evaluated using a standard benchmark dataset. Cao *et al.* [88] improved the performance of TL for BC-classification without any fine-tuning on the source network layers (ResNet-125). Instead, they used random forest dissimilarity for combining various feature groups. The ''ICIAR 2018'' dataset was used, and the classification accuracy was improved to 82.90%. Deniz *et al.* [89] fine-tuned the last three layers in the AlexNet and VGG16 models to classify breast tumors on the BreaKHis dataset. Their model achieved better accuracy than five other methods as it approached 91.37%. In the same dataset, Celik *et al.* [90] pre-trained the DenseNet-161 model and achieved 92.38% and 91.57% for the F-score and accuracy, respectively.

## III. MATERIALS AND METHODS

The proposed method for BC detection and classification consists of two main components. The first component is

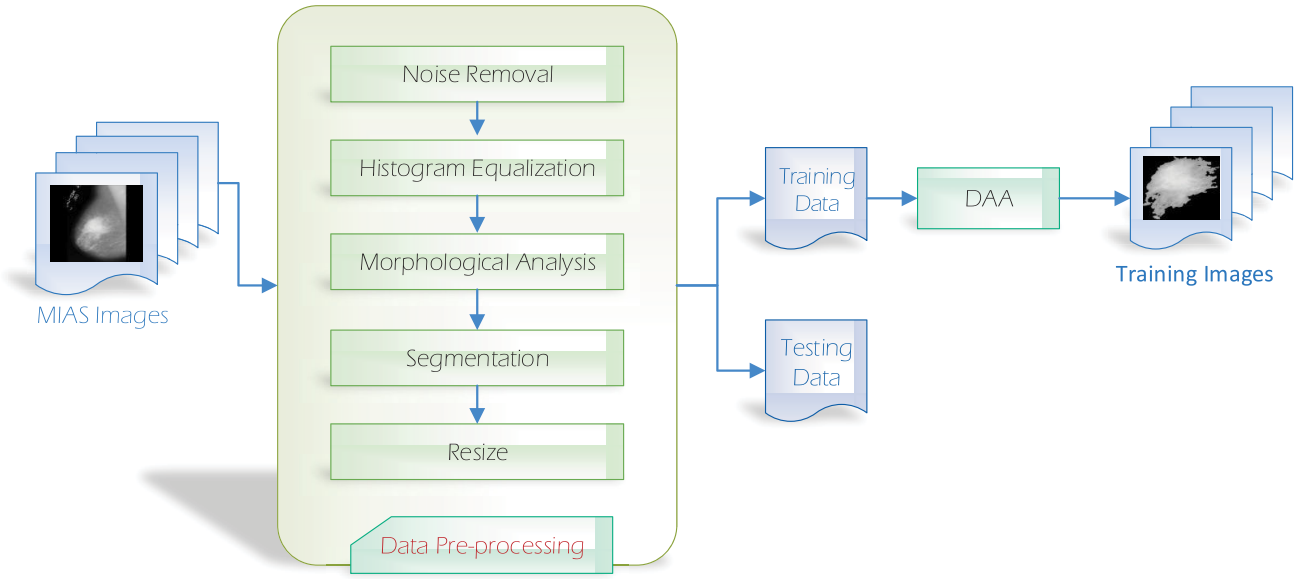


**FIGURE 2.** Data pre-processing.

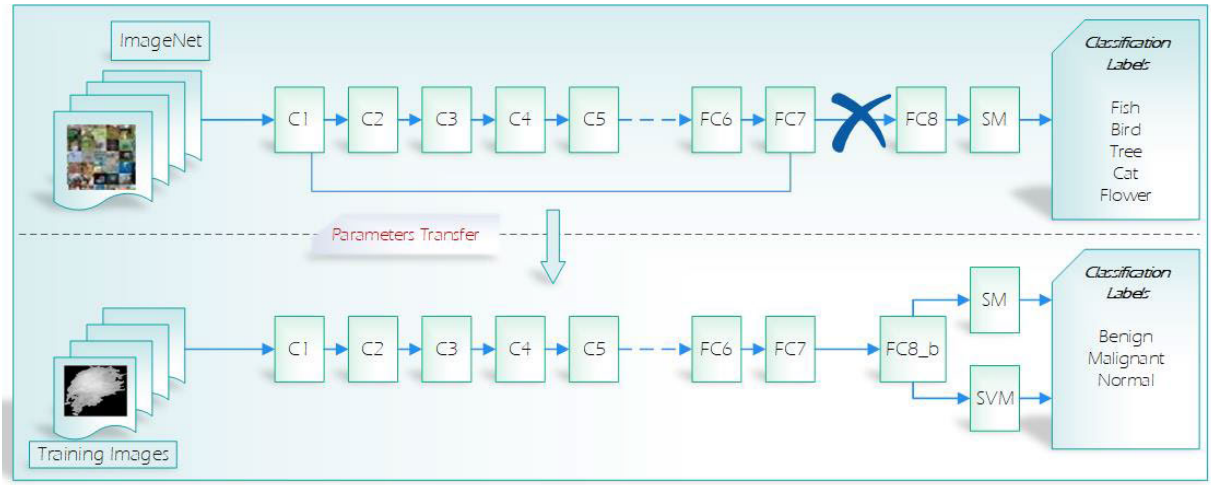


**FIGURE 3.** Transferring CNN parameters.

used for data preprocessing and the second for transferring the CNN parameters, as shown in Figs. 2 and 3.

### *A. DATA PREPROCESSING*

Image preprocessing is very important to remove the limits of observing abnormalities without undue influence from a mammogram. In this work, the tumor regions are automatically extracted using segmentation techniques before the learning process to reduce computation time. Image quality can be improved and the segmentation results can become more accurate using noise removal, histogram equalization, and morphological analysis before segmentation. As shown in Fig. 2, data preprocessing consists of seven phases.

#### 1) NOISE REMOVAL

A 2D median filter of a $3 \times 3$ size is applied to remove the digitization noise from the mammogram image.

#### 2) HISTOGRAM EQUALIZATION

Classical histogram equalization is applied to improve the contrast for all levels of the original image. This is accomplished by effectively distributing the most frequent gray level of the image that is, stretching the intensity range of the image. In mammogram images, histogram equalization is applied to make contrast adjustment so that image anomalies become more visible.

#### 3) MORPHOLOGICAL ANALYSIS

The morphological analysis is an important process for removing non-breast regions before segmentation so that the results are not affected. In morphological operations, the relevant structures are extracted from the input image after applying the structuring element (SE). The output image of this operation has the size of the input. The value of each pixel depends on the corresponding pixel in the input

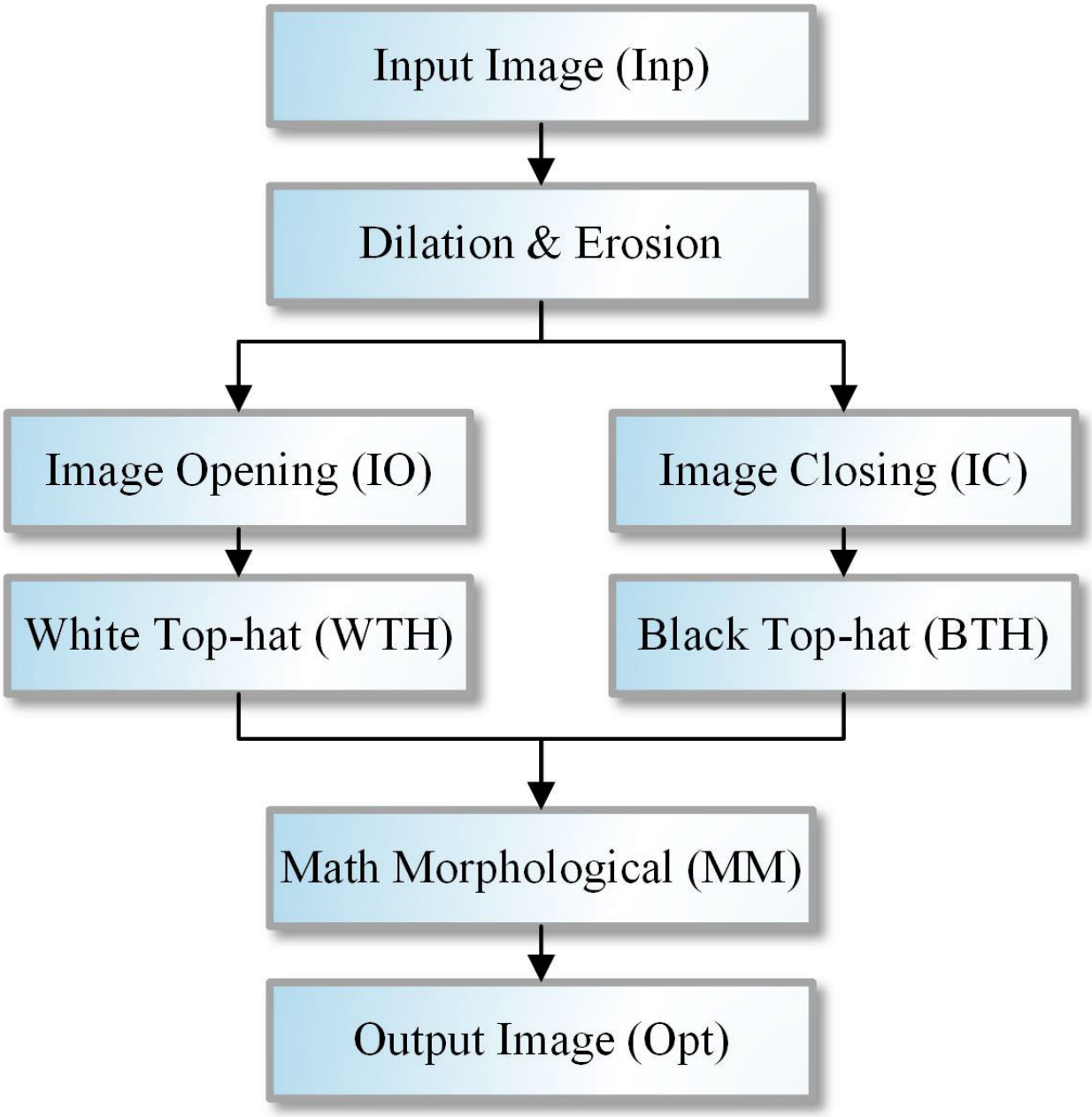


**FIGURE 4.** Mathematical morphological operation.

and its neighbors. The operations described in Fig. 4 can be estimated as follows [91]:

- Image Opening (IO)

$$IO = Inp \ominus SE \oplus SE \tag{1}$$

- Image Closing (IC)

$$IC = Inp \oplus SE \ominus SE \tag{2}$$

- White Top-hat (WTH)

$$WTH = Inp - IO \tag{3}$$

- Black Top-hat (BTH)

$$BTH = IC - Inp \tag{4}$$

- Mathematical Morphological (MM)

$$MM = Inp + WTH - BTH \tag{5}$$

where $\oplus$ and $\ominus$ refer to the dilation and erosion operations, respectively.

#### 4) SEGMENTATION

The computation time can be reduced, and the analysis can be focused on the region mostly affected by cancer using

a threshold-based segmentation method for automatic patch extraction [92].

#### 5) IMAGE RESIZING

The breast images are resized and converted into three channels: red green, and blue (RGB) to match the input size of the pretrained CNN architecture.

#### 6) DATA SPLITTING

The MIAS dataset is split into ''80%'' for the training set and ''20%'' for the testing set [93]–[95].

In addition, to overcome the problem of over-fitting, the experiments have been re-performed using a cross-validation technique with 10-folds. The cross-validation idea is the partitioning of the dataset to k folds with equal size. After that, k-1 folds will be used to train the classifier and the remaining fold will be used to test data to predict each sample label. The final result is the average of different data rounds [96].

#### 7) DATA AUGMENTATION ALGORITHM (DDA)

DL models work better when large datasets are used. Data augmentation is considered one of the most popular methods to increase the size of the dataset, which helps overcome overfitting when training a very small amount of data. In this work, the training data can be augmented using a set of transformations. DAA is implemented to increase the input data. First, the segmented images are rotated clockwise to 90°, 180°, 270°, and 360°. Then, every rotated image is flipped vertically. In this way, an input image will produce eight images. The detailed algorithm for data augmentation is shown in Alg. 1.

**Algorithm 1** Data Augmentation Algorithm (DDA)

**Input:**
Benign B, Malignant M, Normal N segmented mammogram image.

**Processing:**
Step1: ∀ B, rotate to 0°, 90°, 180°, 270°
Step2: Perform flip on all step1.
Step3: ∀ M, rotate to 0°, 90°, 180°, 270°
Step4: Perform flip on all step3
Step5: ∀ N, rotate to 0°, 90°, 180°, 270°
Step6: Perform flip on all step5
Repeat for all training data

**Output:**
Save steps1,2,3,4,5,6

### B. DEEP-CNN TRAINING BASED ON TL

In this work, the Inception V3, ResNet50, VGG19, VGG16, and Inception-V2 ResNet networks are used for feature extraction. These networks are trained using the ImageNet dataset. The filters in the network layers are used to recognize the input features such as colors, vertical, and horizontal lines. Subsequently, trivial shapes and small parts can be recognized. From the generated output, the class in which the input image belongs (i.e. cats, birds, and other) can be determined. Next, the pre-trained network for classifying different objects in a new dataset is applied (in this work for BC-classification to perform TL). The trained parameters from the source task, except for the last three (FCL, SM, and classification) layers are frozen and transferred to the target task, as shown in Fig. 3. Then, the extracted patches from the segmentation process during preprocessing are used to continue the network training. Hence, the newly-trained dense layers are few. Furthermore, the already-trained layers in the pre-trained network are combined with these layers for a new class classification. Thus, the training process can be created very quickly and very few training data are needed compared with the CNN training from scratch. The extracted features are then used to train support vector machine (SVM) and SM classifiers for applying classification task. Fine-tuning is conducted using the stochastic gradient-descent (SGD) method with momentum (SGDM), which is actually an improved version of SGD with the learning parameters shown in Table 1. SGDM' goal is to increase velocity in all dimensions, even in those with consistent gradient. Due to SGDM jittering, gradient high-velocity dimensions are reduced, whereas past gradients that have some momentum are reduced due to a saddle point when the current gradient is approximately zero [97], [98]. Here, the same hyperparameter setting is used in all experiments (before & after preprocessing). The ResNet50 network was proposed by the Microsoft research team [99], where 50 represented the number of deep layers. It contains 48-convolution, 1 average-pooling, and 1 max-pooling layer with a 224 × 224-input size. The residual block is a concatenation for three convolution layers. The overall architecture is shown in Fig. 5. The Inception-V2 ResNet network contains 148 deep layers, and it is capable of classifying 1000 classes. This network was developed by the Google research team. The network has an input-image size of 244 × 244, as shown in Fig. 5. A detailed description of the Inception-V2 ResNet, stem, and reduction blocks was discussed in [100]. The Inception V3 is a CNN developed by the Google research team. It contains 48 layers with an input-image size of 299 × 299. The Inception V3 network is trained using the ImageNet database, which contains one million training images in 1,000 categories. The Inception V3 has a decreased set of parameters due to factoring larger convolution layers into smaller ones and using different other

**TABLE 1. Parameter settings.**

| Sr. no | Parameter | Value |
|---|---|---|
| 1 | Minimum batch size | 10 |
| 2 | Maximum Epochs | 20 |
| 3 | Learn-rate drop factor | 0.5 |
| 4 | Initial-learn rate | 1e-4 |
| 5 | Learn-rate drop period | 5 |

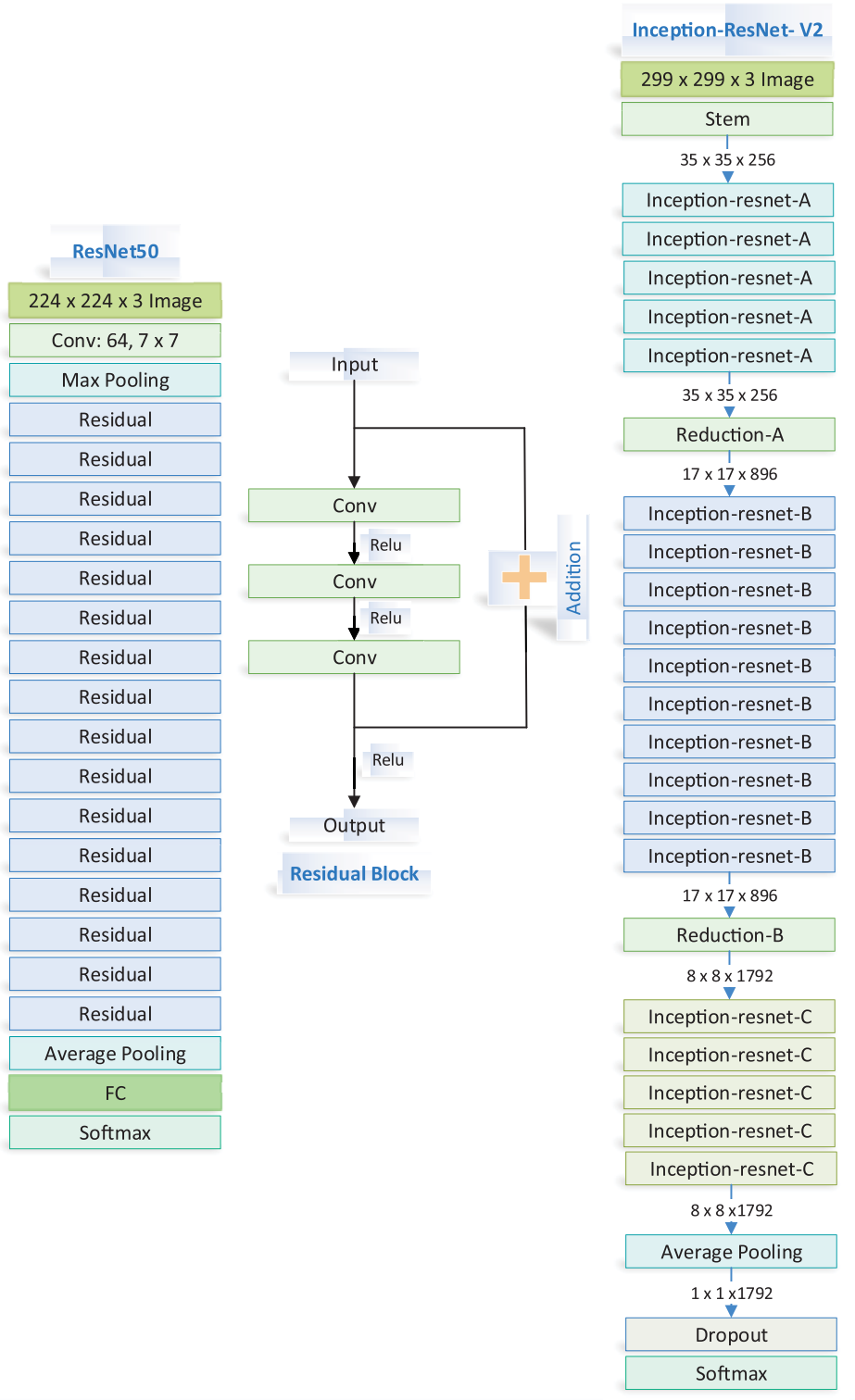


**FIGURE 5.** The ResNet50 and Inception V2-ResNet architectures.

means. A set of changes to the basic structure of the Inception V3 leads to a faster and more accurate architecture, which also works for smaller datasets as discussed in [101]. The RMSProp Optimizer is added to the Inception V3 network in addition to factorized 7 × 7 convolutions. The basic architecture of the Inception V3 network is presented in Fig. 6. The VGG19 is a CNN developed by the Visual Geometry Group at Oxford's and thus, the name VGG. The VGG19 is a variant of VGG models trained over the ImageNet database and contains 19 deep layers

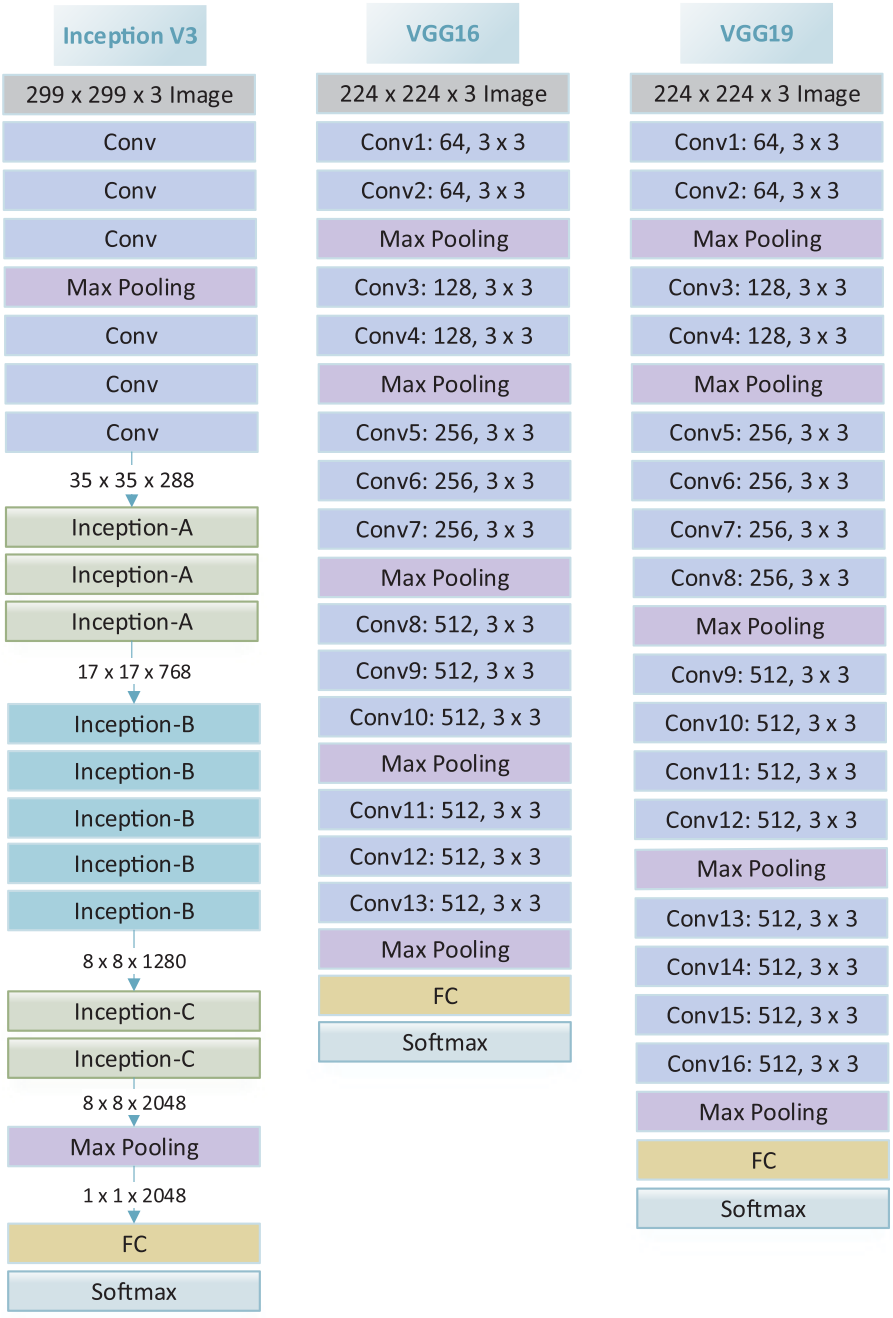


**FIGURE 6.** The Inception V3, VGG16, and VGG19 architectures.

(16 convolution and 3 max-pooling layers) with an input-image size of 244 × 244. The kernel size used in the VGG19 is 3 × 3 with 1 stride size, whereas max-pooling is performed in a 2 × 2-pixel window with a stride size equal to 2. There are different variants of the VGG such as VGG16 and others. The major disadvantage of this CNN is its large size in terms of the number of parameters to be trained. The VGG19 CNN is bigger than the VGG16. However, since the VGG19 performs almost as well as the VGG16, many people use the VGG16 [102]. The basic VGG19 architecture is presented in Fig. 6. The VGG16 is trained over the ImageNet database. Its architecture is deep and very simple. As shown in Fig. 6, it consists of 13 convolution layers and 5 max-pooling layers, followed by three FCLs and an SM classifier. The input is a 224 × 244-RGB image. The applied filters are 3 × 3 with a stride equal 1, whereas max-pooling is a 2 × 2-pixel window with a stride equal to 2 [102].

## IV. RESULTS

### A. DATASET DESCRIPTION

As shown in Fig. 7, the digital database for screening mammography (DDSM), MIAS, and private datasets are the most

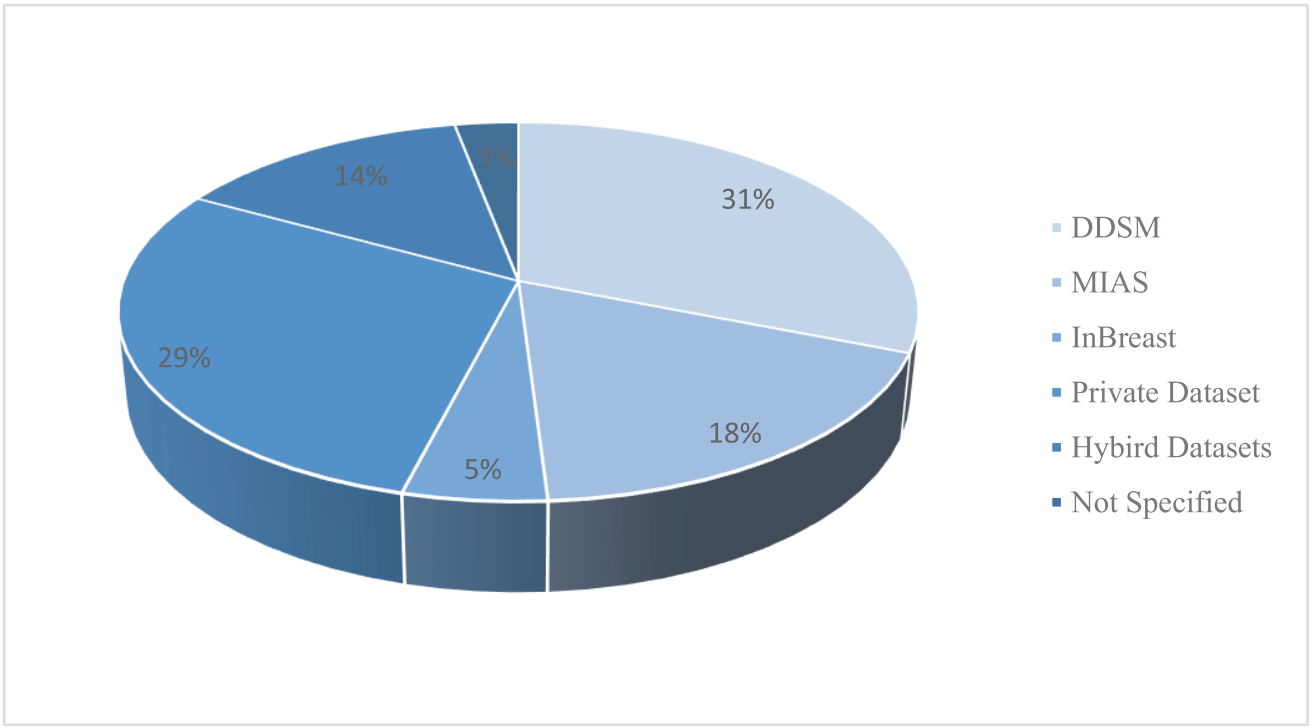


**FIGURE 7.** Dataset frequency usage for breast-tumor classification.

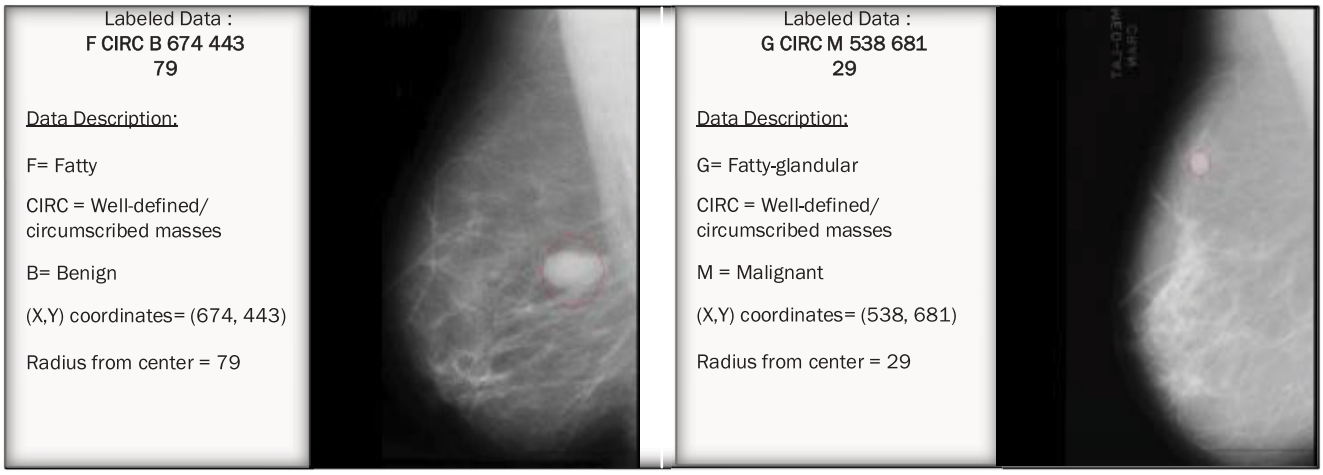


**FIGURE 8.** Tumor description in the mammogram images.

popular databases used for BC-classification models based on the statistics discussed on [103]. In this work, the applied mammogram database was provided by MIAS. Every image has a 1024 × 1024 size in portable gray map (PGM) format. The MIAS includes 322 images in three classes, 61 images for the benign case, 52 images for the malignant case, and 209 for the normal case. Data details are shown in Table 2. It provides details for ground-truth information on the mammogram images such as background tissue, abnormality present class, tumor type, abnormality center coordinates, and approximate radius for enclosing the abnormality circle. The abnormality class is presented by six forms; calcification (CALC), well-defined circumscribed masses (CIRC), spiculated masses (SPIC), other ill-defined masses (MISC), architectural distortion (ARCH), and asymmetry. A tumor region in the mammogram images is presented in Fig. 8.

### B. EXPERIMENTAL ANALYSIS

In this section, several experiments conducted for investigating the performance of the proposed model on the MIAS dataset are presented. Here, TL is applied to five DL models (Inception V3, Inception-V2 ResNet, VGG16, VGG19, and ResNet50) and compared in terms of accuracy, precision, sensitivity, specificity, and AUC. The dataset was divided into three classes "Benign, Malignant, and Normal." Then, it was split to 80% and 20% for the training and testing tasks,

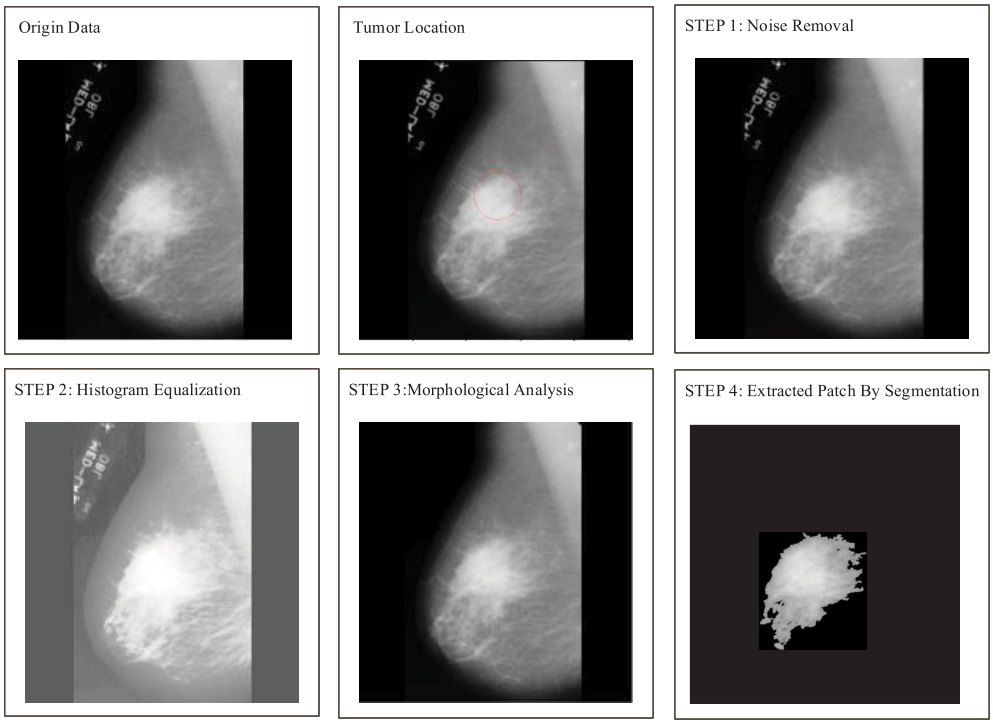


**FIGURE 9.** MIAS data preprocessing results.

**TABLE 2.** MIAS data description.

| Class | Sub-class of abnormality present | Number |
|---|---|---|
| Benign | CIRC | 19 |
| | CALC | 10 |
| | SPIC | 11 |
| | MISC | 7 |
| | ARCH | 9 |
| | ASYM | 6 |
| | | Total = 62 |
| Malignant | CIRC | 4 |
| | CALC | 13 |
| | SPIC | 8 |
| | MISC | 7 |
| | ARCH | 10 |
| | ASYM | 9 |
| | | Total = 51 |
| Normal | — | 209 |

respectively. The efficiency of the proposed models was measured using the evaluation metrics for three classes, as shown in Table 3 and Eqs. 6–10. The benefits of preprocessing were investigated by conducting experiments twice, before and after preprocessing. The classifier performance results without preprocessing are presented in Table 4. It can be observed that the Inception-V2 ResNet achieves the best performance results in terms of accuracy, whereas the Inception V3 was ranked the second-best in terms of accuracy. On the other hand, the VGG16 achieves the best results in terms of sensitivity and specificity with 55.76% and 69.68%, respectively. Also, the ResNet50 achieves better results in terms of precision, AUC, and the F-score with values of 34.84%, 0.55, and 34.00%. The first component of the proposed model for preprocessing phase results is described in Fig. 9. The training data to be used as a training input for the proposed CNN are then augmented using DAA, as shown in Fig. 10. The results presented in Table 5 confirm that the VGG16 achieves the best results in the case of TL in the BC detection mechanism with SM classifier.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{9}$$

$$\text{F - score} = \frac{2 * \text{Percision} * \text{Sensitivity}}{Percision + Sensitivity} \tag{10}$$

The detailed results per class are presented in Table 6. From this table, it can be observed that: 1) In the benign case, the VGG16 was ranked first in terms of accuracy, sensitivity, AUC, and the F-score, whereas the VGG19 was ranked first in terms of specificity and precision. 2) In the malignant case, the ResNet50 achieved the best accuracy, specificity, precision, and the F-score. In the last case (Normal),

$$\text{Accuracy} = \frac{P_{BB} + P_{MM} + P_{NN}}{P_{BB} + P_{MB} + P_{NB} + P_{BM} + P_{MM} + P_{NM} + P_{BN} + P_{MN} + P_{NN}} \tag{6}$$

**TABLE 3.** Evaluation metrics for BC-classification.

| | | Predict | | |
|---|---|---|---|---|
| | | Benign | Malignant | Normal |
| Actual | Benign | $P_{BB}$ (TP) | $P_{MB}$ | $P_{NB}$ |
| | Malignant | $P_{BM}$ | $P_{MM}$ (TP) | $P_{NM}$ |
| | Normal | $P_{BN}$ | $P_{MN}$ | $P_{NN}$ (TP) |

**TABLE 4.** BC-classification performance of various CNNs before preprocessing.

| CNN | Classifier Performance | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
| Inception V3 | 62.50 | 33 | 63.7 | 33.8 | 0.43 | 29.70 |
| VGG19 | 54.69 | 21.21 | 63.9 | 27.67 | 0.43 | 24.00 |
| VGG16 | 59.38 | 55.76 | 69.68 | 32.1 | 0.49 | 30.30 |
| ResNet50 | 54.69 | 34.77 | 66.62 | 34.84 | 0.55 | 34.00 |
| Inception-V2 ResNet | 64.06 | 0.22 | 55.12 | 32.5 | 0.51 | 26.00 |

**TABLE 5.** BC-classification performance of various CNNs after preprocessing using 80:20 and SM classifier.

| CNN | Classifier Performance | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
| Inception V3 | 96.19 | 92.6 | 96.7 | 91.3 | 0.99 | 91.8 |
| VGG19 | 94.35 | 89.86 | 94.8 | 88 | 0.97 | 88.3 |
| VGG16 | 96.77 | 96 | 98 | 91 | 0.99 | 93 |
| ResNet50 | 95.27 | 92 | 95.6 | 90 | 0.97 | 91 |
| Inception-V2 ResNet | 93.42 | 90.66 | 96 | 82.7 | 0.978 | 86 |

**TABLE 6.** BC-classification performance of various CNNs per class using 80:20 and SM classifier.

| CNN | Class | Classifier performance per class | | | | | |
|---|---|---|---|---|---|---|---|
| | | Accuracy (%) | Sensitivity | Specificity | Precision | AUC | F-score |
| Inception V3 | Benign | 96.89 | 0.96 | 0.97 | 0.87 | 0.99 | 0.91 |
| | Malignant | 96.02 | 0.86 | 0.98 | 0.89 | 0.98 | 0.874 |
| | Normal | 95.67 | 0.96 | 0.95 | 0.98 | 0.99 | 0.97 |
| VGG19 | Benign | 94.12 | 0.80 | 0.982 | 0.93 | 0.96 | 0.86 |
| | Malignant | 95.5 | 0.946 | 0.956 | 0.76 | 0.986 | 0.84 |
| | Normal | 93.43 | 0.95 | 0.905 | 0.95 | 0.973 | 0.95 |
| VGG16 | Benign | 97.06 | 0.99 | 0.97 | 0.85 | 0.992 | 0.914 |
| | Malignant | 97.4 | 0.95 | 0.98 | 0.88 | 0.99 | 0.91 |
| | Normal | 95.85 | 0.94 | 0.99 | 1.0 | 0.992 | 0.97 |
| ResNet50 | Benign | 94.81 | 0.89 | 0.96 | 0.83 | 0.95 | 0.86 |
| | Malignant | 97.58 | 0.92 | 0.99 | 0.92 | 0.99 | 0.92 |
| | Normal | 93.43 | 0.94 | 0.92 | 0.96 | 0.97 | 0.95 |
| Inception-V2 ResNet | Benign | 92.21 | 0.89 | 0.93 | 0.67 | 0.965 | 0.764 |
| | Malignant | 96.02 | 0.93 | 0.97 | 0.82 | 0.988 | 0.87 |
| | Normal | 92.04 | 0.90 | 0.98 | 0.99 | 0.98 | 0.94 |

**TABLE 7.** BC-classification performance of various CNNs after preprocessing using 10-fold cross-validation and SM classifier.

| CNN | Classifier Performance | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
| Inception V3 | 96.41 | 93.2 | 97.6 | 92.1 | 0.99 | 92.6 |
| VGG19 | 94.44 | 90.62 | 93.22 | 89.02 | 0.98 | 89.81 |
| VGG16 | 96.65 | 95.44 | 96.92 | 91.5 | 0.98 | 93.42 |
| ResNet50 | 96.01 | 91.12 | 96.5 | 90.2 | 0.98 | 90.65 |
| Inception-V2 ResNet | 93.83 | 91.44 | 93.2 | 83.1 | 0.98 | 87.07 |

**TABLE 8.** BC-classification performance of various CNNs after preprocessing using 80:20 and SVM classifier.

| CNN | Classifier Performance | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
| Inception V3 | 98.15 | 96.93 | 98.57 | 95.4 | 0.993 | 96 |
| VGG19 | 95.84 | 91.33 | 96.37 | 90.66 | 0.982 | 91.33 |
| VGG16 | 98.96 | 97.83 | 99.13 | 97.35 | 0.995 | 97.66 |
| ResNet50 | 97.11 | 94.73 | 97.6 | 93.03 | 0.978 | 93.66 |
| Inception-V2 ResNet | 94.23 | 88 | 94.96 | 87.2 | 0.984 | 87.33 |

**TABLE 9.** BC-classification performance of various CNNs per class using 80:20 and SVM classifier.

| CNN | Class | Classifier performance per class | | | | | |
|---|---|---|---|---|---|---|---|
| | | Accuracy (%) | Sensitivity | Specificity | Precision | AUC | F-score |
| Inception V3 | Benign | 98.62 | 0.99 | 0.985 | 0.936 | 0.997 | 0.96 |
| | Malignant | 98.1 | 0.945 | 0.9876 | 0.934 | 0.99 | 0.94 |
| | Normal | 97.75 | 0.973 | 0.984 | 0.992 | 0.992 | 0.98 |
| VGG19 | Benign | 95.67 | 0.889 | 0.972 | 0.881 | 0.971 | 0.89 |
| | Malignant | 96.19 | 0.888 | 0.975 | 0.869 | 0.99 | 0.88 |
| | Normal | 95.67 | 0.963 | 0.944 | 0.97 | 0.985 | 0.97 |
| VGG16 | Benign | 99.31 | 0.99 | 0.993 | 0.97 | 0.997 | 0.98 |
| | Malignant | 98.62 | 0.956 | 0.991 | 0.956 | 0.992 | 0.96 |
| | Normal | 98.96 | 0.989 | 0.99 | 0.99 | 0.996 | 0.99 |
| ResNet50 | Benign | 97.4 | 0.935 | 0.982 | 0.927 | 0.96 | 0.93 |
| | Malignant | 97.23 | 0.941 | 0.977 | 0.88 | 0.992 | 0.91 |
| | Normal | 96.71 | 0.966 | 0.969 | 0.984 | 0.982 | 0.97 |
| Inception-V2 ResNet | Benign | 94.12 | 0.872 | 0.955 | 0.809 | 0.975 | 0.84 |
| | Malignant | 94.64 | 0.821 | 0.971 | 0.847 | 0.99 | 0.83 |
| | Normal | 93.94 | 0.947 | 0.923 | 0.96 | 0.989 | 0.95 |

**TABLE 10.** BC-classification performance of various CNNs after preprocessing using 10-fold cross-validation and SVM classifier.

| CNN | Classifier Performance | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
| Inception V3 | 98.45 | 97.2 | 98.9 | 93.5 | 0.994 | 95.31 |
| VGG19 | 95.92 | 92.41 | 95.21 | 91.96 | 0.99 | 92.18 |
| VGG16 | 98.87 | 97.27 | 98.2 | 98.84 | 0.993 | 98.04 |
| ResNet50 | 96.87 | 94.24 | 96.99 | 95.45 | 0.97 | 94.84 |
| Inception-V2 ResNet | 94.76 | 88.86 | 94.72 | 88.14 | 0.987 | 88.49 |

**TABLE 11.** Comparison between the proposed model and existing models.

| Method | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | AUC | F-score (%) |
|---|---|---|---|---|---|---|
| Abbas (2016) | 91.5 | 92 | 84.2 | - | 0.91 | - |
| Charan (2018) | 65 | - | - | - | - | - |
| Ting et al. (2019) | 90.50 | 89.47 | 90.71 | - | 0.90 | - |
| Sha et al. (2020) | 92 | 96 | 93 | - | - | - |
| Proposed | 98.96 | 97.83 | 99.13 | 97.35 | 0.995 | 97.66 |

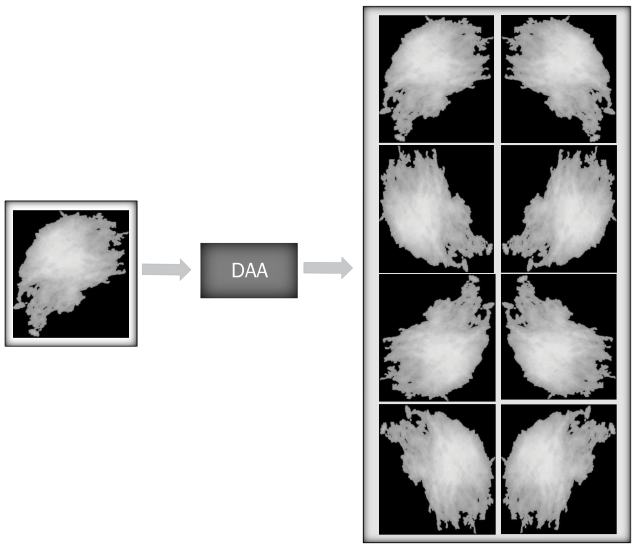


**FIGURE 10.** Results obtained using the data augmentation algorithm.

the VGG16 was ranked first in terms of accuracy, specificity, precision, and AUC, whereas the Inception V3 was ranked first in sensitivity.

The results of 10-fold cross-validation are shown in Table 7. It can be noted that the cross-validation method achieved better results than the 80-20 technique in all CNNs except VGG16. The results obtained from the SVM classifier achieved better than the results obtained from the SM classifier as presented in details in Tables 8 - 10. The experiments performed are presented in Table 11, where the performance is compared with four other existing models. The analysis results confirm that the proposed model performs better than other existing models in terms of accuracy, sensitivity, specificity, and AUC.

## V. CONCLUSION

In this paper, a novel deep learning model for improving the classification results on the MIAS dataset was proposed. The purpose of this model is to help medical doctors in BC detection and diagnosis. The MIAS images were divided into three different classes, benign, malignant, and normal. The original MIAS dataset was pre-processed for noise removal, improving contrast in breast images, non-breast region removal, and determining the cancerous area. The data augmentation concept was also proposed for increasing the size of a

dataset to enhance the performance of the CNN structure. Then, the freezing and fine-tuning strategies were used to improve the mass-lesion classification accuracy of the mentioned dataset. The VGG16 model achieved the best accuracy, sensitivity, specificity, AUC, and the F-score compared with four other models. Finally, it can be concluded that integrating the CNN using learning transfer in the screening mechanism, a clear improvement can be achieved compared with other existing approaches. The results showed 98.96% accuracy, 97.83% sensitivity, 99.13% specificity, 97.35% precision, 97.66% F-score, and 0.995 AUC. These results are better than the other mentioned methods.

In future work, the proposed method can be further used to diagnosis or prognosis of paraquat-poisoned patients [104]–[109], identification of poisoning status [110]–[112], diagnosis of tuberculous pleural effusion [113], differentiation of malignant and benign thyroid nodules [114], early diagnosis of Parkinson's disease [115]–[119], RNA secondary structure prediction [120], detection of erythemato-squamous diseases [121], online recognition of foreign fibers in cotton [122].

## ACKNOWLEDGMENT

*(The authors have equally contributed to this work.)*

## REFERENCES

[1] K. P. Lowry, "Long-term outcomes and cost-effectiveness of breast cancer screening with digital breast tomosynthesis in the United States," *J. Nat. Cancer Inst.*, vol. 112, no. 6, pp. 582–589, 2020.

[2] S. A. George, "Barriers to breast cancer screening: An integrative review," *Health Care Women Int.*, vol. 21, no. 1, pp. 53–65, Jan. 2000, doi: 10.1080/073993300245401.

[3] P. H. Viale, "The American cancer society's facts & figures: 2020 edition," *J. Adv. Practitioner Oncol.*, vol. 11, no. 2, p. 135, Mar. 2020, doi: 10.6004/jadpro.2020.11.2.1.

[4] J. Zhou, L. Luo, Q. Dou, H. Chen, C. Chen, G. Li, Z. Jiang, and P. Heng, "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images," *J. Magn. Reson. Imag.*, vol. 50, no. 4, pp. 1144–1151, Oct. 2019, doi: 10.1002/jmri.26721.

[5] B. A. Soomro, G. R. Lakhan, S. Mangi, and N. Shah, "Predicting entrepreneurial intention among business students of public sector universities of Pakistan: An application of the entrepreneurial event model," *World J. Entrepreneurship, Manage. Sustain. Develop.*, vol. 16, no. 3, pp. 219–2301, 2020.

[6] W. Zhu, C. Ma, X. Zhao, M. Wang, A. A. Heidari, H. Chen, and C. Li, "Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine," *IEEE Access*, vol. 8, pp. 61107–61123, 2020.

[7] A. Lin, Q. Wu, A. A. Heidari, Y. Xu, H. Chen, W. Geng, and C. Li, "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.

[8] J. Tu, A. Lin, H. Chen, Y. Li, and C. Li, "Predict the entrepreneurial intention of fresh graduate students based on an adaptive support vector machine framework," *Math. Problems Eng.*, vol. 2019, pp. 1–16, Jan. 2019.

[9] Y. Wei, N. Ni, D. Liu, H. Chen, M. Wang, Q. Li, X. Cui, and H. Ye, "An improved grey wolf optimization strategy enhanced SVM and its application in predicting the second major," *Math. Problems Eng.*, vol. 2017, pp. 1–12, Feb. 2017.

[10] Y. Zhang, R. Liu, A. A. Heidari, X. Wang, Y. Chen, M. Wang, and H. Chen, "Towards augmented kernel extreme learning models for bankruptcy prediction: Algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, pp. 185–212, Mar. 2021.

[11] C. Yu, M. Chen, K. Cheng, X. Zhao, C. Ma, F. Kuang, and H. Chen, "SGOA: Annealing-behaved grasshopper optimizer for global tasks," *Eng. Comput.*, vol. 2, pp. 1–28, Jan. 2021.

[12] Z. Cai, J. Gu, J. Luo, Q. Zhang, H. Chen, Z. Pan, Y. Li, and C. Li, "Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112814.

[13] Y. Xu, H. Chen, A. A. Heidari, J. Luo, Q. Zhang, X. Zhao, and C. Li, "An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks," *Expert Syst. Appl.*, vol. 129, pp. 135–155, Sep. 2019.

[14] J. Luo, H. Chen, Q. Zhang, Y. Xu, H. Huang, and X. Zhao, "An improved grasshopper optimization algorithm with application to financial stress prediction," *Appl. Math. Model.*, vol. 64, pp. 654–668, Dec. 2018.

[15] D. Zhao, C. Huang, Y. Wei, F. Yu, M. Wang, and H. Chen, "An effective computational model for bankruptcy prediction using kernel extreme learning machine approach," *Comput. Econ.*, vol. 49, no. 2, pp. 325–341, Feb. 2017.

[16] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong, J. Li, and X. Xu, "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Eng. Appl. Artif. Intell.*, vol. 63, pp. 54–68, Aug. 2017.

[17] T. Wang, L. Zhao, P. Huang, X. Zhang, and J. Xu, "Haze concentration adaptive network for image dehazing," *Neurocomputing*, vol. 439, pp. 75–85, Jun. 2021.

[18] X. Zhang, J. Wang, T. Wang, R. Jiang, J. Xu, and L. Zhao, "Robust feature learning for adversarial defense via hierarchical feature alignment," *Inf. Sci.*, vol. 560, pp. 256–270, Jun. 2021.

[19] P. Huang, L. Zhao, R. Jiang, T. Wang, and X. Zhang, "Self-filtering image dehazing with self-supporting module," *Neurocomputing*, vol. 432, pp. 57–69, Apr. 2021.

[20] T. Wang, X. Zhang, R. Jiang, L. Zhao, H. Chen, and W. Luo, "Video deblurring via spatiotemporal pyramid network and adversarial gradient prior," *Comput. Vis. Image Understand.*, vol. 203, Feb. 2021, Art. no. 103135.

[21] Y. Caixia, X. Chang, L. Minnan, Z. Qinghua, Z. Xiaoqin, L. Zhihui, and N. Feiping, "Self-weighted robust LDA for multiclass classification with edge classes," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 1, pp. 1–19, 2020.

[22] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 31, 2020, doi: 10.1109/TNNLS.2020.3009209.

[23] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Comput. Vis. Image Understand.*, vols. 197–198, Aug. 2020, Art. no. 103003.

[24] X. Zhang, R. Jiang, T. Wang, P. Huang, and L. Zhao, "Attention-based interpolation network for video deblurring," *Neurocomputing*, Sep. 2020, doi: 10.1016/j.neucom.2020.04.147.

[25] W. Liu, X. Chang, L. Chen, D. Phung, X. Zhang, Y. Yang, and A. G. Hauptmann, "Pair-based uncertainty and diversity promoting early active learning for person re-identification," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 2, pp. 1–15, Mar. 2020.

[26] X. Zhang, T. Wang, W. Luo, and P. Huang, "Multi-level fusion and attention-guided CNN for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Dec. 22, 2020, doi: 10.1109/TCSVT.2020.3046625.

[27] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Nov. 3, 2020, doi: 10.1109/TCSVT.2020.3035722.

[28] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 238–255, Jan. 2021.

[29] H. Zhao, H. Guo, X. Jin, J. Shen, X. Mao, and J. Liu, "Parallel and efficient approximate nearest patch matching for image editing applications," *Neurocomputing*, vol. 305, pp. 39–50, Aug. 2018.

[30] Y. Zhao, X. Jin, Y. Xu, H. Zhao, M. Ai, and K. Zhou, "Parallel style-aware image cloning for artworks," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 2, pp. 229–240, Feb. 2015.

[31] Y. Yang, H. Zhao, L. You, R. Tu, X. Wu, and X. Jin, "Semantic portrait color transfer with Internet images," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 523–541, Jan. 2017.

[32] H.-L. Zhao, G.-Z. Nie, X.-J. Li, X.-G. Jin, and Z.-G. Pan, "Structure-aware nonlocal optimization framework for image colorization," *J. Comput. Sci. Technol.*, vol. 30, no. 3, pp. 478–488, May 2015.

[33] H. Zhao, X. Jin, and X. Mao, "Real-time directional stylization of images and videos," *Multimedia Tools Appl.*, vol. 63, no. 3, pp. 647–661, Apr. 2013.
[34] S. Lu, X. Jin, H. Zhao, and Y. Zhao, "Real-time image marbleization," *Multimedia Tools Appl.*, vol. 64, no. 3, pp. 795–808, Jun. 2013.
[35] S. Lu, A. Jaffer, X. Jin, H. Zhao, and X. Mao, "Mathematical marbling," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 26–35, Nov. 2012.
[36] H. Zhao, H. Zhang, and X. Jin, "Efficient image decolorization with a multimodal contrast-preserving measure," *Comput. Graph.*, vol. 70, pp. 251–260, Feb. 2018.
[37] X. Li, H. Huang, H. Zhao, Y. Wang, and M. Hu, "Learning a convolutional neural network for propagation-based stereo image segmentation," *Vis. Comput.*, vol. 36, no. 1, pp. 39–52, Jan. 2020.
[38] X. Li, H. Zhao, H. Huang, Z. Hu, and L. Xiao, "Interactive image recoloring by combining global and local optimization," *Multimedia Tools Appl.*, vol. 75, no. 11, pp. 6431–6443, Jun. 2016.
[39] H. Huang, X. Li, H. Zhao, G. Nie, Z. Hu, and L. Xiao, "Manifold-preserving image colorization with nonlocal estimation," *Multimedia Tools Appl.*, vol. 74, no. 18, pp. 7555–7568, Sep. 2015.
[40] N. Gu, M. Fan, L. Du, and D. Ren, "Efficient sequential feature selection based on adaptive eigenspace model," *Neurocomputing*, vol. 161, pp. 199–209, Aug. 2015.
[41] P. Zhou, J. Chen, M. Fan, L. Du, Y.-D. Shen, and X. Li, "Unsupervised feature selection for balanced clustering," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105417.
[42] M. Fan, N. Gu, H. Qiao, and B. Zhang, "Dimensionality reduction: An interpretation from manifold regularization perspective," *Inf. Sci.*, vol. 277, pp. 694–714, Sep. 2014.
[43] M. Fan, X. Zhang, H. Qiao, and B. Zhang, "Efficient isometric multi-manifold learning based on the self-organizing method," *Inf. Sci.*, vol. 345, pp. 325–339, Jun. 2016.
[44] C. Yang, L. Bruzzone, H. Zhao, Y. Tan, and R. Guan, "Superpixel-based unsupervised band selection for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7230–7245, Dec. 2018.
[45] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowl.-Based Syst.*, vol. 144, pp. 65–76, Mar. 2018.
[46] S. Wang and J. Xiang, "A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps," *Soft Comput.*, vol. 24, no. 4, pp. 2983–2997, Feb. 2020.
[47] Q. Gao, "A multi-sensor fault detection strategy for axial piston pump using the Walsh transform method," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 4, 2018, Art. no. 1550147718772531.
[48] W. Song, J. Xiang, and Y. Zhong, "A simulation model based fault diagnosis method for bearings," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3857–3867, Jun. 2018.
[49] X. Liu, H. Huang, and J. Xiang, "A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105653.
[50] H. Yu, K. Yuan, W. Li, N. Zhao, W. Chen, C. Huang, H. Chen, and M. Wang, "Improved butterfly optimizer-configured extreme learning machine for fault diagnosis," *Complexity*, vol. 2021, Feb. 2021, Art. no. 6315010.
[51] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," *Neural Process. Lett.*, vol. 39, no. 1, pp. 25–43, Feb. 2014.
[52] S.-J. Wang, W.-J. Yan, T. Sun, G. Zhao, and X. Fu, "Sparse tensor canonical correlation analysis for micro-expression recognition," *Neurocomputing*, vol. 214, pp. 218–232, Nov. 2016.
[53] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, Oct. 2016.
[54] W.-J. Yan and Y.-H. Chen, "Measuring dynamic micro-expressions via feature extraction methods," *J. Comput. Sci.*, vol. 25, pp. 318–326, Mar. 2018.
[55] F. Qu, S.-J. Wang, W.-J. Yan, H. Li, S. Wu, and X. Fu, "CAS(ME): A database for spontaneous macro-expression and micro-expression spotting and recognition," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 424–436, Oct. 2018.
[56] Y. Wang, J. See, Y.-H. Oh, R. C.-W. Phan, Y. Rahulamathavan, H.-C. Ling, S.-W. Tan, and X. Li, "Effective recognition of facial micro-expressions with video motion magnification," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21665–21690, Oct. 2017.
[57] S.-J. Wang, Y. He, J. Li, and X. Fu, "MESNet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos," *IEEE Trans. Image Process.*, vol. 30, pp. 3956–3969, 2021.
[58] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, "A content-based recommender system for computer science publications," *Knowl.-Based Syst.*, vol. 157, pp. 1–9, Oct. 2018.
[59] R. Guan, H. Zhang, Y. Liang, F. Giunchiglia, L. Huang, and X. Feng, "Deep feature-based text clustering and its explanation," *IEEE Trans. Knowl. Data Eng.*, early access, Oct. 6, 2020, doi: 10.1109/TKDE.2020.3028943.
[60] C. Feng, Z. Zhu, Z. Cui, V. Ushakov, J. Dreher, W. Luo, R. Gu, X. Wu, and F. Krueger, "Prediction of trust propensity from intrinsic brain morphology and functional connectome," *Hum. Brain Mapping*, vol. 42, no. 1, pp. 175–191, Jan. 2021.
[61] H. Wang, X. Wu, and L. Yao, "Identifying cortical brain directed connectivity networks from high-density EEG for emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Jul. 30, 2020, doi: 10.1109/TAFFC.2020.3006847.
[62] H. Zhang, R. Li, X. Wen, Q. Li, and X. Wu, "Altered time-frequency feature in default mode network of autism based on improved Hilbert-huang transform," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 485–492, Feb. 2021, doi: 10.1109/JBHI.2020.2993109.
[63] Z. Li, X. Wu, X. Xu, H. Wang, Z. Guo, Z. Zhan, and L. Yao, "The recognition of multiple anxiety levels based on electroencephalograph," *IEEE Trans. Affect. Comput.*, early access, Aug. 20, 2019, doi: 10.1109/TAFFC.2019.2936198.
[64] Z. Yao, Y. Fu, J. Wu, W. Zhang, Y. Yu, Z. Zhang, X. Wu, Y. Wang, and B. Hu, "Morphological changes in subregions of hippocampus and amygdala in major depressive disorder patients," *Brain Imag. Behav.*, vol. 14, no. 3, pp. 653–667, Jun. 2020.
[65] X. Fei, J. Wang, S. Ying, Z. Hu, and J. Shi, "Projective parameter transfer based sparse multiple empirical kernel learning machine for diagnosis of brain disease," *Neurocomputing*, vol. 413, pp. 271–283, Nov. 2020.
[66] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowl.-Based Syst.*, vol. 164, pp. 96–106, Jan. 2019.
[67] H. Chen, B. Yang, D. Liu, W. Liu, Y. Liu, X. Zhang, and L. Hu, "Using blood indexes to predict overweight statuses: An extreme learning machine-based approach," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0143003.
[68] X. Sun, Y. Liu, D. Wei, M. Xu, H. Chen, and J. Han, "Selection of interdependent genes via dynamic relevance analysis for cancer diagnosis," *J. Biomed. Informat.*, vol. 46, no. 2, pp. 252–258, Apr. 2013.
[69] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Appl. Soft Comput.*, vol. 88, Mar. 2020, Art. no. 105946.
[70] M. Wang, H. Chen, B. Yang, X. Zhao, L. Hu, Z. Cai, H. Huang, and C. Tong, "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, Dec. 2017.
[71] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, B. Yang, and D. Liu, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowl.-Based Syst.*, vol. 96, pp. 61–75, Mar. 2016.
[72] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, and H. Chen, "Medical image classification using spatial adjacent histogram based on adaptive local binary patterns," *Comput. Biol. Med.*, vol. 72, pp. 185–200, May 2016.
[73] W. Shan, Z. Qiao, A. A. Heidari, H. Chen, H. Turabieh, and Y. Teng, "Double adaptive weights for stabilization of moth flame optimizer: Balance analysis, engineering cases, and medical diagnosis," *Knowl.-Based Syst.*, vol. 214, Feb. 2021, Art. no. 106728.
[74] H. Zhao, X. Qiu, W. Lu, H. Huang, and X. Jin, "High-quality retinal vessel segmentation using generative adversarial network with a large receptive field," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 828–842, Sep. 2020.

[75] W. Lotter, A. Rahman Diab, B. Haslam, J. G. Kim, G. Grisot, E. Wu, K. Wu, J. O. Onieva, J. L. Boxerman, M. Wang, M. Bandler, G. Vijayaraghavan, and A. G. Sorensen, "Robust breast cancer detection in mammography and digital breast tomosynthesis using annotation-efficient deep learning approach," 2019, *arXiv:1912.11027*. [Online]. Available: http://arxiv.org/abs/1912.11027

[76] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, Jul. 2018.

[77] A. Saber, A. M. Al-Zoghby, and S. Elmougy, "Big-data aggregating, linking, integrating and representing using semantic Web technologies," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl. (AMLTA)*, in Advances in Intelligent Systems and Computing, 2018, pp. 331–342, doi: 10.1007/978-3-319-74690-6_33.

[78] K. Weiss, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016, doi: 10.1186/s40537-016-0043-6.

[79] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, K. H. Cha, and C. D. Richter, "Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms," *Phys. Med. Biol.*, vol. 62, no. 23, pp. 8894–8908, Nov. 2017, doi: 10.1088/1361-6560/aa93d4.

[80] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Syst. Appl.*, vol. 120, pp. 103–115, Apr. 2019.

[81] M. Toğaçar, K. B. Özkurt, B. Ergen, and Z. Cömert, "BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer," *Phys. A, Stat. Mech. Appl.*, vol. 545, May 2020, Art. no. 123592, doi: 10.1016/j.physa.2019.123592.

[82] Q. Abbas, "DeepCAD: A computer-aided diagnosis system for mammographic masses using deep invariant features," *Computers*, vol. 5, no. 4, p. 28, Oct. 2016, doi: 10.3390/computers5040028.

[83] Z. Sha, L. Hu, and B. D. Rouyendegh, "Deep learning and optimization algorithms for automatic breast cancer detection," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 2, pp. 495–506, Jun. 2020, doi: 10.1002/ima.22400.

[84] S. Charan, M. J. Khan, and K. Khurshid, "Breast cancer detection in mammograms using convolutional neural network," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, Mar. 2018, pp. 1–5, doi: 10.1109/icomet.2018.8346384.

[85] N. Wahab, A. Khan, and Y. S. Lee, "Transfer learning based deep CNN for segmentation and detection of mitoses in breast cancer histopathological images," *Microscopy*, vol. 68, no. 3, pp. 216–233, Jun. 2019, doi: 10.1093/jmicro/dfz002.

[86] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proc. 5th Int. Conf. Bioinf. Comput. Biol.*, Jan. 2017, pp. 59–62, doi: 10.1145/3035012.3035022.

[87] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, Jul. 2019, doi: 10.1016/j.patrec.2019.03.022.

[88] H. Cao, *Improve the Performance of Transfer Learning Without Fine-Tuning Using Dissimilarity-Based Multi-View Learning for Breast Cancer Histology Images* (Lecture Notes in Computer Science Image Analysis and Recognition). 2018, pp. 779–787, doi: 10.1007/978-3-319-93000-8_88.

[89] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health Inf. Sci. Syst.*, vol. 6, no. 1, pp. 1–7, Dec. 2018, doi: 10.1007/s13755-018-0057-x.

[90] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images," *Pattern Recognit. Lett.*, vol. 133, pp. 232–239, May 2020, doi: 10.1016/j.patrec.2020.03.011.

[91] K. A. M. Said and A. B. Jambek, "A study on image processing using mathematical morphological," in *Proc. 3rd Int. Conf. Electron. Design (ICED)*, Aug. 2016, pp. 507–512, doi: 10.1109/iced.2016.7804697.

[92] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*. [Online]. Available: http://arxiv.org/abs/1606.05718

[93] M. Sakr, A. Saber, O. M. Abo-Seida, and A. Keshk, "Machine learning for breast cancer classification using k-star algorithm," *Appl. Math. Inf. Sci. J.*, vol. 14, no. 5, pp. 855–863, 2020.

[94] A. G. Hussien, M. Amin, M. Wang, G. Liang, A. Alsanad, A. Gumaei, and H. Chen, "Crow search algorithm: Theory, recent advances, and applications," *IEEE Access*, vol. 8, pp. 173548–173565, 2020.

[95] Q. H. Nguyen, H.-B. Ly, L. S. Ho, N. Al-Ansari, H. V. Le, V. Q. Tran, I. Prakash, and B. T. Pham, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Problems Eng.*, vol. 2021, pp. 1–15, Feb. 2021.

[96] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. AI*, Aug. 1995, vol. 14. no. 2, pp. 1137–1145.

[97] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, Jan. 1999.

[98] I. M. J. Sutskever, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. Int. Conf. Mach. Learn.*, Feb. 2013, pp. 1139–1147.

[99] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/cvpr.2016.90.

[100] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017, pp. 4278–4284.

[101] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818-2826, doi: 10.1109/cvpr.2016.308.

[102] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[103] S. B. Y. Tasdemir, K. Tasdemir, and Z. Aydin, "A review of mammographic region of interest classification," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 5, p. e1357, Sep. 2020, doi: 10.1002/widm.1357.

[104] X. Zhao, 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," *Comput. Biol. Chem.*, vol. 78, pp. 481–490, Feb. 2019.

[105] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Comput. Biol. Med.*, vol. 59, pp. 116–124, Apr. 2015.

[106] L. Hu, H. Li, Z. Cai, F. Lin, G. Hong, H. Chen, and Z. Lu, "A new machine-learning method to prognosticate paraquat poisoned patients by combining coagulation, liver, and kidney indices," *PLoS ONE*, vol. 12, no. 10, Oct. 2017, Art. no. e0186427.

[107] L. Hu, F. Lin, H. Li, C. Tong, Z. Pan, J. Li, and H. Chen, "An intelligent prognostic system for analyzing patients with paraquat poisoning using arterial blood gas indexes," *J. Pharmacol. Toxicol. Methods*, vol. 84, pp. 78–85, Mar. 2017.

[108] H. Chen, L. Hu, H. Li, G. Hong, T. Zhang, J. Ma, and Z. Lu, "An effective machine learning approach for prognosis of paraquat poisoning patients using blood routine indexes," *Basic Clin. Pharmacol. Toxicol.*, vol. 120, no. 1, pp. 86–96, Jan. 2017.

[109] C. Wen, F. Lin, B. Huang, Z. Zhang, X. Wang, J. Ma, G. Lin, H. Chen, and L. Hu, "Metabolomics analysis in acute paraquat poisoning patients based on UPLC-Q-TOF-MS and machine learning approach," *Chem. Res. Toxicol.*, vol. 32, no. 4, pp. 629–637, Apr. 2019.

[110] J. Zhu, X. Zhao, H. Li, H. Chen, and G. Wu, "An effective machine learning approach for identifying the glyphosate poisoning status in rats using blood routine test," *IEEE Access*, vol. 6, pp. 15653–15662, 2018.

[111] J. Zhu, F. Zhu, S. Huang, G. Wang, H. Chen, X. Zhao, and S.-Y. Zhang, "A new evolutionary machine learning approach for identifying pyrene induced hepatotoxicity and renal dysfunction in rats," *IEEE Access*, vol. 7, pp. 15320–15329, 2019.

[112] Y. Xu, K. Yu, P. Wang, H. Chen, X. Zhao, and J. Zhu, "A new hybrid machine learning approach for prediction of phenanthrene toxicity on mice," *IEEE Access*, vol. 7, pp. 138461–138472, 2019.

[113] C. Li, 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," *Comput. Methods Programs Biomed.*, vol. 153, pp. 211–225, Jan. 2018.

[114] J. Xia, H. Chen, Q. Li, M. Zhou, L. Chen, Z. Cai, Y. Fang, and H. Zhou, "Ultrasound-based differentiation of malignant and benign thyroid nodules: An extreme learning machine approach," *Comput. Methods Programs Biomed.*, vol. 147, pp. 37–49, Aug. 2017.

[115] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, Apr. 2016.

[116] W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen, "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomed. Signal Process. Control*, vol. 8, no. 4, pp. 364–373, Jul. 2013.
[117] H.-L. Chen, C.-C. Huang, X.-G. Yu, X. Xu, X. Sun, G. Wang, and S.-J. Wang, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 263–271, Jan. 2013.
[118] Z. Cai, J. Gu, and H.-L. Chen, "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, 2017.
[119] Z. Cai, J. Gu, C. Wen, D. Zhao, C. Huang, H. Huang, C. Tong, J. Li, and H. Chen, "An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach," *Comput. Math. Methods Med.*, vol. 2018, Jun. 2018, Art. no. 2396952.
[120] G. Wang, W. Zhang, Q. Ning, and H. Chen, "A novel framework based on ACO and PSO for RNA secondary structure prediction," *Math. Problems Eng.*, vol. 2013, Jan. 2013, Art. no. 796304.
[121] T. Liu, L. Hu, C. Ma, Z.-Y. Wang, and H.-L. Chen, "A fast approach for detection of erythemato-squamous diseases based on extreme learning machine with maximum relevance minimum redundancy feature selection," *Int. J. Syst. Sci.*, vol. 46, no. 5, pp. 919–931, Apr. 2015.
[122] X. Zhao, D. Li, B. Yang, S. Liu, Z. Pan, and H. Chen, "An efficient and effective automatic recognition system for online recognition of foreign fibers in cotton," *IEEE Access*, vol. 4, pp. 8465–8475, 2016.

**ABEER SABER** was born in Damietta, Domyat, Egypt, in 1992. She received the B.Sc. and M.Sc. degrees in computer science from Mansoura University, Egypt, in 2013 and 2018, respectively. She is currently an Assistant Lecturer in computer science with the Faculty of Computers and Information, Kafr El-Sheikh University, Egypt. She has published many research papers in prestigious international conferences and articles in reputable journals. Her current research interests include big data analysis, semantic Web, linked open data, data mining, and machine learning.

**MOHAMED SAKR** was born in Shebin El-Kom, Menoufia, Egypt, in 1990. He received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Menoufia University, Egypt, in 2011, 2014, and 2019, respectively. He is currently a Lecturer of computer science with the Faculty of Computers and Information, Menoufia University. He has published many research papers published in prestigious international conferences and articles in reputable journals. His current research interests include big data analysis, anomaly detection, data mining, and machine learning.

**OSAMA M. ABO-SEIDA** was born in Kotor, Tanta, Egypt, in 1968. He received the B.Sc. and M.Sc. degrees in mathematics and the Ph.D. degree from Tanta University, Egypt, in 1990, 1994, and 1997, respectively. He is currently a Professor of mathematics and the Dean of the Faculty of Computers and Information, Kafr El-Sheikh University. He has authored or coauthored many articles in international repute journals. His research interests include wave propagation, applied mathematics, computional electromagnetics, and magnetical networks.

**ARABI KESHK** received the B.Sc. degree in electronic engineering and the M.Sc. degree in computer science and engineering from the Faculty of Electronic Engineering, Menoufia University, in 1987 and 1995, respectively, and the Ph.D. degree in electronic engineering from Osaka University, Japan, in 2001. His research interests include software testing, software engineering, distributed systems, cloud computing, machine learning, the IoT, big data analytics, and bioinformatics.

**HUILING CHEN** (Associate Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Jilin University, China. He is currently an Associate Professor with the College of Computer Science and Artificial Intelligence, Wenzhou University, China. He has published more than 100 articles in international journals and conference proceedings, including *Information Sciences*, *Pattern Recognition*, *Future Generation Computer System*, *Expert Systems with Applications*, *Knowledge-based Systems, Neurocomputing*, and PAKDD. His research interests include center on machine learning and data mining and their applications to medical diagnosis and bankruptcy prediction. He is currently serving as an Associate Editor for IEEE Access. He is also a reviewer for many journals, such as *Applied Soft Computing*, *Artificial Intelligence in Medicine*, *Knowledge-based Systems*, and *Future Generation Computer System*.