Article

Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion

Kiran Jabeen [1], Muhammad Attique Khan [1], Majed Alhaisoni [2], Usman Tariq [3], Yu-Dong Zhang [4], Ameer Hamza [1], Artūras Mickus [5] and Robertas Damaševičius [5,*]

1   Department of Computer Science, HITEC University Taxila, Taxila 47080, Pakistan;
     kiran.jabeen@student.hitecuni.edu.pk (K.J.); attique.khan@hitecuni.edu.pk (M.A.K.);
     21-mscs-014@student.hitecuni.edu.pk (A.H.)
2   College of Computer Science and Engineering, University of Ha'il, Ha'il 55211, Saudi Arabia;
     m.alhaisoni@uoh.edu.sa
3   College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University,
     Al-Kharaj 11942, Saudi Arabia; u.tariq@psau.edu.sa
4   Department of Informatics, University of Leicester, Leicester LE1 7RH, UK; yudongzhang@ieee.org
5   Department of Applied Informatics, Vytautas Magnus University, LT-44404 Kaunas, Lithuania;
     arturas.mickus@vdu.lt
*   Correspondence: robertas.damasevicius@vdu.lt

Abstract: After lung cancer, breast cancer is the second leading cause of death in women. If breast cancer is detected early, mortality rates in women can be reduced. Because manual breast cancer diagnosis takes a long time, an automated system is required for early cancer detection. This paper proposes a new framework for breast cancer classification from ultrasound images that employs deep learning and the fusion of the best selected features. The proposed framework is divided into five major steps: (i) data augmentation is performed to increase the size of the original dataset for better learning of Convolutional Neural Network (CNN) models; (ii) a pre-trained DarkNet-53 model is considered and the output layer is modified based on the augmented dataset classes; (iii) the modified model is trained using transfer learning and features are extracted from the global average pooling layer; (iv) the best features are selected using two improved optimization algorithms known as reformed differential evaluation (RDE) and reformed gray wolf (RGW); and (v) the best selected features are fused using a new probability-based serial approach and classified using machine learning algorithms. The experiment was conducted on an augmented Breast Ultrasound Images (BUSI) dataset, and the best accuracy was 99.1%. When compared with recent techniques, the proposed framework outperforms them.

Keywords: breast cancer; data augmentation; deep learning; feature optimization; classification

1. Introduction

Breast cancer is one of the most common cancers in women; it starts in the breast and spreads to other parts of the body [1]. This cancer affects the breast glands [2] and is the second most common tumor in the world, next to lung tumors [3]. Breast cancer cells create a tumor that might be seen in X-ray images. In 2020, approximately 1.8 million cancer cases were diagnosed, with breast cancer accounting for 30% of those cases [4]. There are two types of breast cancer: malignant and benign. Cells are classified based on their various characteristics. It is critical to detect breast cancer at an early stage in order to reduce the mortality rate [5].

Many imaging tools are available for the prior recognition and early treatment of breast cancer. Breast ultrasound is one of the most commonly used modalities in clinical practice for the diagnosis process [6,7]. Epithelial cells that border the terminal duct lobular unit are the source of the breast cancer. In situ or noninvasive cancer cells are

those that remain inside the basement membrane of the draining duct and the basement membrane of the parts of the terminal duct lobular unit [8]. One of the most critical factors in predicting treatment decisions in breast cancer is the status of axillary lymph node metastases [9]. Ultrasound imaging is one of the most widely used test materials for detecting and categorizing breast disorders [10]. In addition to mammography, it is a common imaging modality used for performing radiological cancer diagnosis. The problems we may encounter in real life are not even reported. It is imperative to consider the presence of speckle, and to consider pre-processing such as wavelet-based denoising [11], in the first and second generations [12].

Ultrasound is non-invasive, well-tolerated by women, and radiation-free; therefore, it is a method that is frequently used in the diagnosis of breast tumors [9]. In dense breast tissue, ultrasound is a highly powerful diagnostic tool, often finding breast tumors that are missed by mammography [13]. Other types of medical imaging, such as magnetic resonance imaging (MRI) and mammography, are less portable and more costly than ultrasound imaging [14]. Computer-aided diagnosis (CAD) systems were developed to assist radiologists in the analysis of breast ultrasound tests [15,16]. Earlier CAD systems often relied on handmade visual information that is difficult to generalize across ultrasound images taken using different methods [17–22]. Recent developments have helped the construction of artificial intelligence (AI) technologies for the automated identification of breast tumors using ultrasound images [23–25]. A computerized method includes a few important steps such as the pre-processing of ultrasound images, tumor segmentation, extraction of features from the segmented tumor, and finally classification [26].

Recently, deep learning showed a huge improvement for cell segmentation [27], skin melanoma detection [28], hemorrhage detection [29], and a few more [30,31]. In medical imaging, deep learning was successful, especially for breast cancer [32], COVID-19 [33], Alzheimer's disease recognition [34], brain tumor [35] diagnostics, and more [36–38]. CNN is a type of deep learning that includes several hierarchies of layers. Through CNN, image pixels are transformed into features. The features are later utilized for infection detection and classification. In CNN, the features are extracted from the raw images. The features extracted from the raw images also produce some irrelevant features that later affect the classification performance. Therefore, it is essential to select only the most relevant features for a better classification precision rate [39].

The selection of the best features from the originally extracted features is an active research topic. Many selection algorithms are introduced in the literature and applied in medical imaging, such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and a few more. Using these methods, the best subset of the features instead of entire feature space. The main advantage of feature selection methods is that they improve system accuracy while decreasing computational time [40]. However, sometime during the best feature selection process, a few important features are also ignored, which impact on the system accuracy. Therefore, computer vision researchers introduced feature fusion techniques [41]. The fusion process increases the number of predictors and increases the accuracy of the system [42]. Some well-known feature fusion techniques are serial-based fusion and parallel fusion [43].

The following problems are considered in this article: (i) the available ultrasound images are not enough for the training of a good deep model as a model trained on a smaller number of images performs incorrect prediction; (ii) the similarity among benign and malignant breast cancer lesions is very high, which leads to misclassification; (iii) the features extracted from images contain irrelevant and redundant information that causes wrong predictions. To solve these problems, we propose a new fully automated deep learning-based method for breast cancer classification from ultrasound images.

The major contributions of this work are listed below.

- We modified a pre-trained deep model named DarkNet53 and trained it on augmented ultrasound images using transfer learning.

- The best features are selected using reformed deferential evolution (RDE) and reformed gray wolf (RGW) optimization algorithms.
- The best selected features are fused using a probability-based approach and classified using machine learning algorithms.

The rest of the manuscript is organized as follows. The related work of this manuscript is described in Section 2. Section 3 presents the proposed methodology, which includes deep learning, feature selection, and fusion. Results and analysis are discussed in Section 4. Finally, we conclude the proposed methodology in Section 5.

2. Related Work

Researchers present a number of computer vision-based automated methods for breast cancer classification using ultrasound images [44,45]. A few of them concentrated on the segmentation step, followed by feature extraction [46], and a few extracted features from raw images. Researchers used the preprocessing step in a few studies to improve the contrast of the input images and highlight the infected part for better feature extraction [47]. For example, Sadad et al. [48] presented a computer-aided diagnosis (CAD) method for the detection of breast cancer. They applied Hilbert Transform (HT) for reconstructing brightness-mode images from the rough data. After that, the tumor is segmented using a marker-controlled watershed transformation. In the subsequent step, shape, and textural features are extracted and classified using the K-Nearest Neighbor (KNN) classifier and the ensemble decision tree model. Badawy et al. [3] performed semantic segmentation, fuzzy logic, and deep learning for breast tumor segmentation and classification from ultrasound images. They used fuzzy logic in the preprocessing step and segmented the tumor using the semantic segmentation approach. Later, eight pre-trained models were applied for final tumor classification.

Mishra et al. [49] introduced a machine learning (ML) radiomics-based classification pipeline. The region of interest (ROI) was separated, and useful features were extracted. The extracted features were classified using machine learning classifiers for the final classification. The experimental process was conducted on the BUSI dataset and showed improved accuracy. Byra [14] introduced a deep learning-based framework for the classification of breast mass from ultrasound images. They used transfer learning (TL) and added deep representation scaling (DRS) layers between pre-trained CNN blocks to improve information flow. Only the parameters of the DRS layers were updated during network training to modify the pre-trained CNN to analyze breast mass classification from the input images. The results showed that the DRS method was significantly better compared with the recent techniques. Irfan et al. [5] introduced a Dilated Semantic Segmentation Network (Di-CNN) for the detection and classification of breast cancer. They considered a pre-trained DenseNet201 deep model and trained it using transfer learning that was later used for feature extraction. Additionally, they implemented a 24-layered CNN and parallel fused feature information with the pre-trained model and classified the nodules. The results showed that the fusion process improves the recognition accuracy.

Hussain et al. [50] presented a contextual level set method for segmentation of breast tumors. They designed a UNet-style encoder-decoder architecture network to learn high-level contextual aspects from semantic data. Xiangmin et al. [51] presented a deep doubly supervised transfer learning network for breast cancer classification. They introduced a Learning using Privileged Information (LUPI) paradigm, which was executed through the Maximum Mean Discrepancy (MMD) criterion. Later, they combined both techniques using a novel doubly supervised TL network (DDSTN) and achieved improved performance. Woo et al. [52] introduced a computerized diagnosis system for breast cancer classification using ultrasound images. They introduced an image fusion technique and combined it with image content representation and several CNN models. The experimental process was conducted on BUSI and private datasets and achieved notable performance. Byra et al. [53] presented a deep learning model for breast mass detection in ultrasound images. They considered the problem of variation in breast mass size, shape, and characteristics. To

solve these issues, they performed selective kernel U-Net CNN. Based on this approach, they fused the information and performed an experimental process on 882 breast images. Additionally, they considered three more datasets and achieved improved accuracy.

Kadry et al. [54] created a computerized technique for detecting breast tumor section (BTS) from breast MRI slices This study employs a combined thresholding and segmentation approach to improve and extract the BTS from 2D MRI slices. To improve the BTS, a tri-level thresholding based on the Slime Mould Algorithm and Shannon's Entropy is created, and Watershed Segmentation is implemented to mine the BTS. Following the extraction of the BTS, a comparison between the BTS and ground truth is carried out, and the required Image Performance Values are generated. Lahoura et al. [55] used an Extreme Learning Machine (ELM) to diagnose breast cancer. Second, the gain ratio feature selection approach is used to exclude unimportant features. Finally, a cloud computing-based method for remote breast cancer diagnostics is presented and validated on the Wisconsin Diagnostic Breast Cancer dataset.

Maqsood et al. [56] offered a brain tumor diagnosis technique based on edge detection and the U-NET model. The suggested tumor segmentation system is based on image enhancement, edge detection, and classification using fuzzy logic. The contrast enhancement approach is used to pre-process the input pictures, and a fuzzy logic-based edge detection method is utilized to identify the edge in the source images, and dual tree-complex wavelet transform is employed at different scale levels. The decaying sub-band pictures are used to calculate the features, which are then classified using the U-NET CNN classification, which detects meningioma in brain images. Rajinikanth et al. [57] created an automated breast cancer diagnosis system utilizing breast thermal images. First, they captured images of various breast orientations. They then extracted healthy/DCIS image patches, processed the patches with image processing, used the Marine Predators Algorithm for feature extraction and feature optimization, and performed classification using the Decision Tree (DT) classifier, which achieved higher accuracy (>92%) when compared with other methods. In [58], the authors presented a novel layer connectivity based architecture for the low contrast nodules segmentation from ultrasound images. They employed dense connectivity and combined it with high-level coarse segmentation. Later, the dilated filter was applied to refine the nodule. Moreover, a class imbalance loss function is also proposed to improve the accuracy of the proposed architecture.

Based on the techniques mentioned above, we discovered that most researchers do not pay attention to the preprocessing step. Typically, researchers performed the segmentation step first, followed by the extraction of features. A few of them used feature fusion to improve their classification results. They did not, however, concentrate on the selection of optimal features. They also ignored computational time, which is now an important factor. In this paper, we proposed an optimal deep learning feature fusion framework for breast mass classification. A summary of a few of the latest techniques is given below Table 1.

Table 1. Summary of existing techniques for breast cancer classification.

Reference	Methods	Features	Dataset
[47], 2021	Shape Adaptive CNN	Deep learning	Breast Ultrasound Images (BUSI)
[48], 2020	Hilbert transform and Watershed	Textural features	BUSI
[3], 2021	Fuzzy Logic and Semantic Segmentation	Deep features	BUSI
[49], 2021	Machine learning and radiomics	Textural and geometric features	BUSI
[14], 2021	CNN and deep representation scaling	Deep features through scaling layers	BUSI
[50], 2020	U-Net Encoder-Decoder CNN architecture	High level contextual features	BUSI
[56], 2021	U-Net and Fuzzy logic	CNN features	BUSI

3. Proposed Methodology

The proposed framework for breast cancer classification using ultrasound images is presented in this section. Figure 1 illustrates the architecture of the proposed framework. Initial data augmentation is performed on the original ultrasound images and then passed to the fine-tuned deep network DarkNet53 for training purposes. Training is performed using TL and extract features from the global average pool layer. Extracted features are refined using the reformed feature optimization techniques, such as reformed differential evolution (RDE) and reformed gray wolf (RGW) algorithms. The best selected features are fused using a probability-based approach. Finally, the fused features are classified using machine learning classifiers. A detailed description of each step is given below.
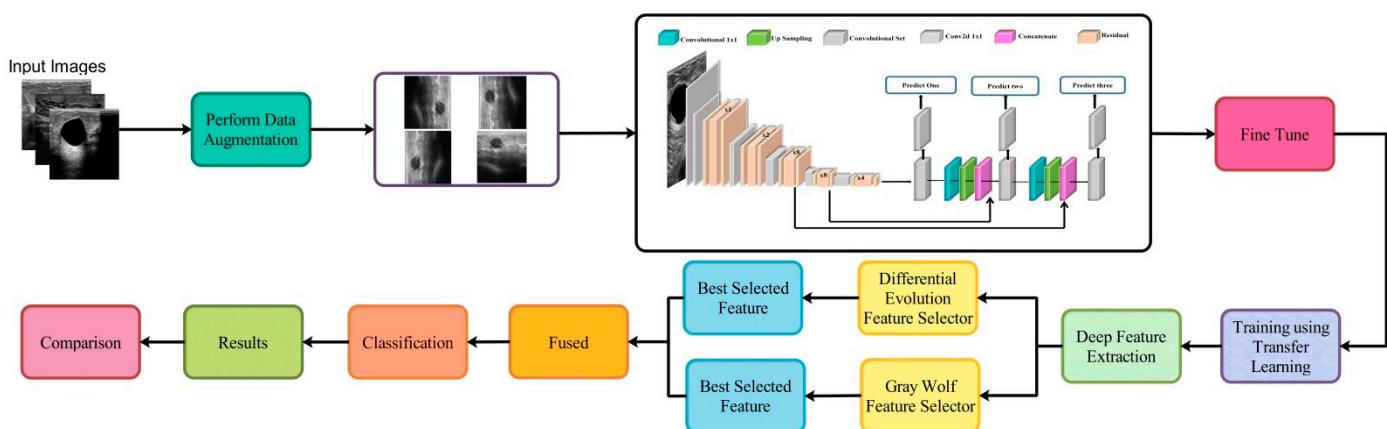

Figure 1. The proposed framework for breast cancer classification using Ultrasound Images.

3.1. Dataset Augmentation

Data augmentation has been an important research area in recent years in the domain of deep learning. In deep learning, neural networks required many training samples; however, existing data sets in the medical domain belong to the low resource domain. Therefore, a data augmentation step is necessary to increase the diversity of the original dataset.

In this work, the BUSI dataset is used for the validation process. There are 780 images in the collection with an average image size of 500×500 pixels. This dataset consists of three total categories: normal (133 images), malignant (210 images), and benign (487 images) [59], as illustrated in Figure 2. We divided this entire dataset into the training and testing of ratio 50:50. After this, the training images of each class were normal (56 images), malignant (105 images), and benign (243 images). This dataset is not enough to train the deep learning model; therefore, a data augmentation step is employed. Three operations such as horizontal flip, vertical flip, and rotate 90 are implemented and performed on original ultrasound images to increase the diversity of the original dataset. These implemented operations are performed multiple times until the number of images in each class has reached 4000. After the augmentation process, the number of images in the dataset is 12,000.

3.2. Modified DarkNet-53 Model

DarkNet-53 is a 53-layer deep convolutional neural network. It serves as the basis for the YOLOv3 object detection method. It can ensure super expression of features while avoiding the gradient problem produced by a too-deep network by combining Resnet's qualities. The structure of the DarkNet-53 model is shown in Figure 3. It combines the residual network with the deep residual network. It contains successive 1×1 and 3×3 convolution layers and residual blocks. The convolutional layer is defined as follows:

$$a_m^n = \sum_{j \in X_i} a_j^{n-1} * y_{jm}^n + z_m^n \tag{1}$$
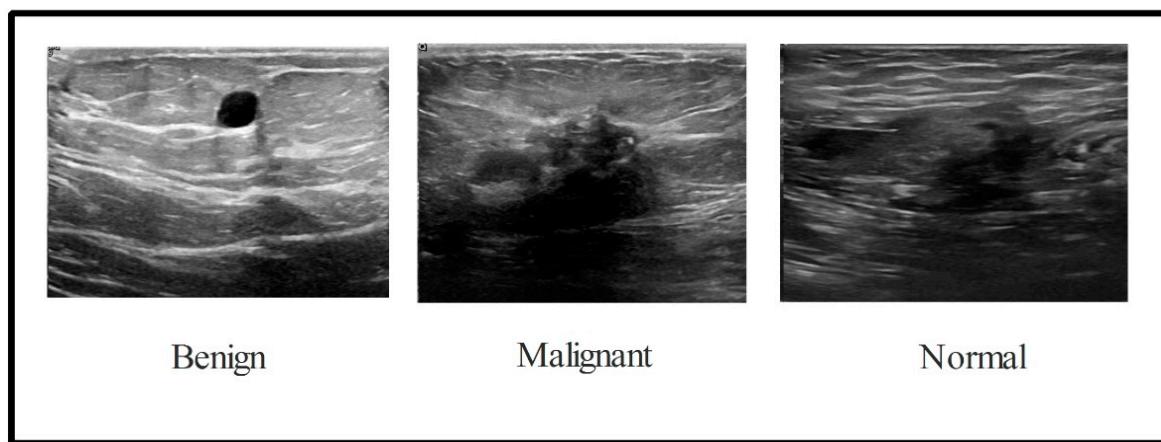
**Figure 2.** Sample ultrasound images of the BUSI dataset [59].
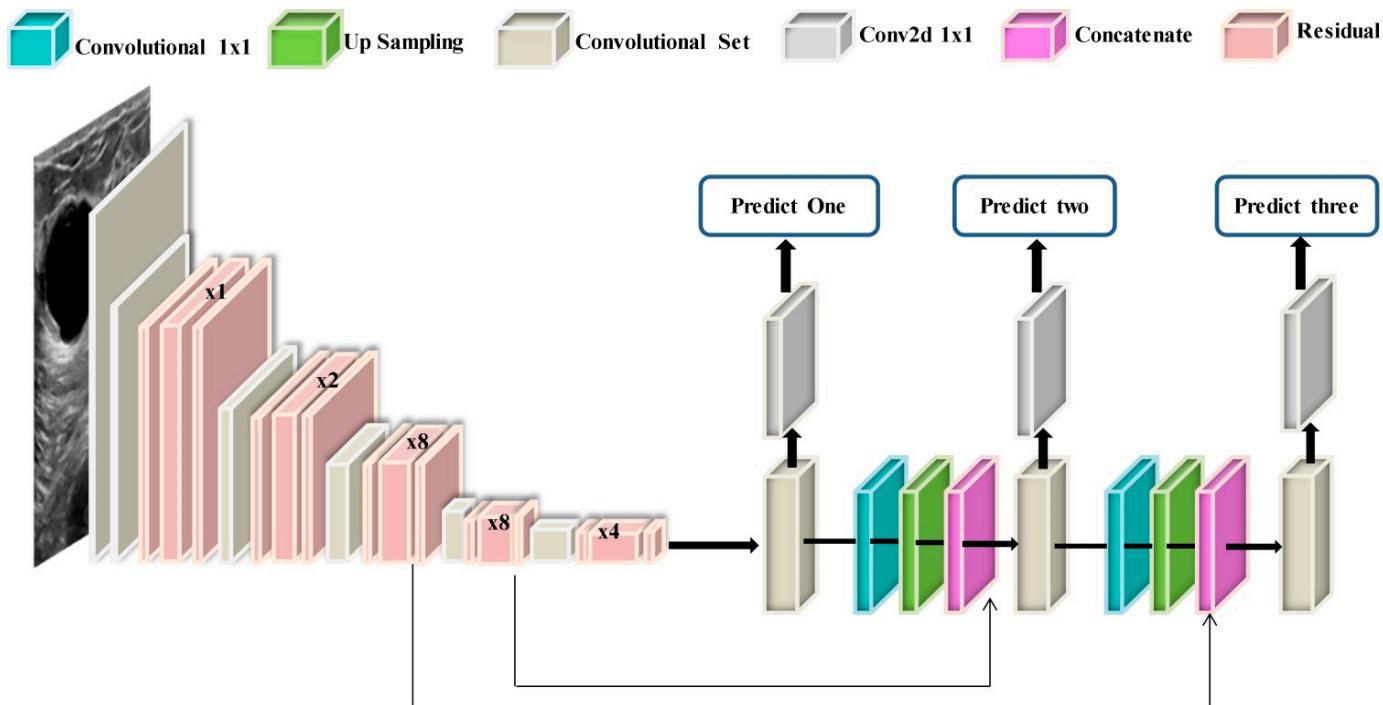
Figure 3. Structure of Modified DarkNet-53 deep model.

In Equation (1), the input image is twisted by several convolution kernels to produce m separate feature maps a_m^n, which is represented in layer n by the m feature map. The symbol * represents the convolution operation. The feature vector of the image is represented by X_i and the j element of the m convolution kernel in the layer n is represented by y_j^n.

The next important layer is the batch normalization (BN) layer.

$$a_{out} = \frac{\propto (a_m^n - \partial)}{\sqrt{\omega^2 + \varphi}} + \gamma \qquad (2)$$

In Equation (2), the scaling factor is represented by $\propto$, the mean of all outputs is represented by ∂, the input variance is represented by ω, φ is a constant offset represented by γ, and the convolution calculation result is denoted by a_{out}. The result of BN denoted by a_{out}. The output is normalized using Batch Normalization corresponding to the same distribution of the coefficients of the same batch of eigenvalues. Following that, it has a convolutional layer that can accelerate network convergence, as well as avoiding over-fitting. The next layer is also known as an activation layer. In DarNet53, a leaky ReLu

layer is included as an activation function. This function increases the nonlinearity of the network:

$$x_j = \begin{cases} y_j, & if\ a_{out} \geq 0 \\ \frac{y_j}{b_j}, & if\ a_{out} < 0 \end{cases} \tag{3}$$

In Equation (3), the input value is denoted by y_j, the activation value is represented by x_j, and the fixed parameter in the interval (1, +∞) is denoted by b_j. Another important layer in this network is pooling layer. This layer is employed for the downsampling of weights in the network. The max-pooling layer is used in this network. In the last example, all weights are combined in one layer in the form of a 1D array, also called features. These extracted features are finally classified in the output layer. The depth of this model is 53, the size is 155 MB, the number of parameters is 41.6 million, and the image input size is 256-by-256. The detailed layer-wise architecture is given in Figure 4.

	Types	Filters	Size	Output
	Convolutional	32	3 x 3	256 x 256
	Convolutional	64	3 x 3 / 2	128 x 128
	Convolutional	32	1 x 1	
1 x	Convolutional	64	3 x 3	
	Residual			128 x 128
	Convolutional	128	3 x 3 / 2	64 x 64
	Convolutional	64	1 x 1	
2 x	Convolutional	128	3 x 3	
	Residual			64 x 64
	Convolutional	256	3 x 3 / 2	32 x 32
	Convolutional	128	1 x 1	
8 x	Convolutional	256	3 x 3	
	Residual			32 x 32
	Convolutional	512	3 x 3 / 2	16 x 16
	Convolutional	256	1 x 1	
8 x	Convolutional	512	3 x 3	
	Residual			16 x 16
	Convolutional	1024	3 x 3 / 2	8 x 8
	Convolutional	512	1 x 1	
4 x	Convolutional	1024	3 x 3	
	Residual			8 x 8
	Avgpool		Global	
	Connected		3	
	Softmax			

Figure 4. The layer wise architecture of Modified DarkNet-53 deep model.

3.3. Transfer Learning

Transfer learning (TL) is a machine learning approach in which a pre-trained model is reused for another task [60]. Reusing or transferring data from previous learned tasks for the newly learned tasks has the potential to dramatically improve the sampling efficiency of a supervised learning agent from a practical standpoint [61]. Here, TL is employed for the deep feature extraction. For this purpose, initially pre-trained model is fine-tuned and then trained using TL. Mathematically, TL is defined as follows:

A domain $d = \{Y, \ p(y)\}$ is described by two parameters: a feature space Y, and a distribution of marginal probabilities $f(y)$, where $y = \{y1, \ y2, \ y3, \ldots yn\} \in Y$. If there are two different domains, then they either have dissimilar marginal probabilities $(p(Y_p) \neq p(Y_q))$ or feature space $(Y_p \neq Y_q)$.

Task: Given a particular domain d, there are two components of task $t \ \{X, g(.)\}$: a label space X, and a predictive function $g \ (.)$; this is not visible, but can be derived from training data $\{(m_j, n_j j\{1, 2, 3, \ldots N\}$, where $m_j Y$ and $n_j \ X$. From a probabilistic point, $f(m_j)$ may also be written as $p(n_j | m_j)$, thus we can rewrite the task t as $t = \{X, \ P(x|Y)\}$. If two tasks are dissimilar, their label spaces may differ $(X_p \neq X_q)$ or result in dissimilar distributions with conditional probabilities $(p(X_p | Y_p) \neq p(X_q | Y_q))$.

The visual process of transfer learning is illustrated in Figure 5. The knowledge of the original model (source domain) is transferred to the modified deep model (target domain). After that, this modified model is trained, and the following hyper parameters are utilized: learning rate is 0.001, mini batch size is 16, epochs are 200, and the learning method is the stochastic gradient descent. The features are extracted from the Global Average Pooling (GAP) layer of the modified deep model. The extracted features are later optimized using two reformed optimization algorithms.
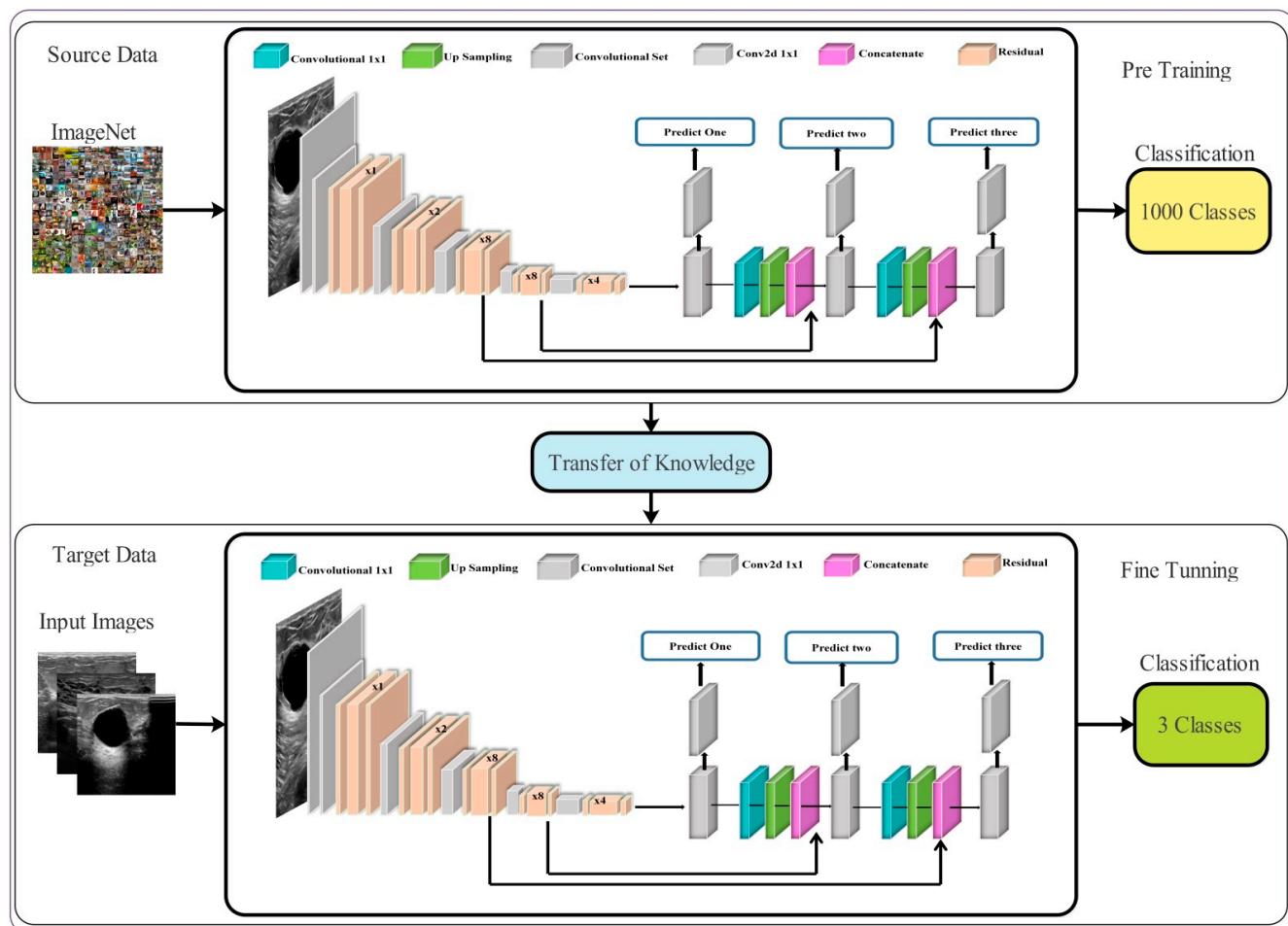

Figure 5. Transfer learning-based training of modified model and extract features.

3.4. Best Features Selection

In this work, two optimization algorithms are reformed for the selection of best features such as differential evolution and gray wolf and fused their information for the final classification. The vector size after performing a differential evolution algorithm is 4788×818. Here, 818 is the number of features and 4788 is the number of images. The vector size after performing the gray wolf optimization algorithm is 4788×734.

3.4.1. Reformed Differential Evolution (RDE) Algorithm

The DE algorithm searches the solution space using the differences between individuals as a guide. The DE's main idea is to scale and differentiate two different specific vectors in the same population, then add a third individual vector to this population to generate a mutation independent vector, which is crossed with the parent independent vector with a certain possibility to produce an intended individual vector. Finally, greedy selection is applied to the generated individual vector and the parent independent vector, and the consistently better vector is preserved for the future generation. The DE's fundamental evolution processes are as follows:

Initialization: D-dimensional vectors (D) are used as the starting solution in the DE algorithm. The population number can be represented by P, each independent factor can be denoted by $z_j\,(Y) = (z_{j1}(Y),\, z_{j2}(Y), z_{j3}(Y), \ldots, z_{jn}(Y))$, and $z_j\,(Y) \in$ denotes the deep extracted features. The starting population is produced in $[z_{min}, z_{max}]$. Here, the number of D-dimensional vectors is denoted by D, population numbers are represented by P, and $z_j(Y)$ represents the *jth* individual.

$$z_{jw} = z_{min} + rand(0,1) \times (z_{max} + z_{min}) \tag{4}$$

where Y denotes the *Yth* generation, the maximum and minimum values of the search space are representing by z_{max} and z_{min}, respectively, and $rand(0,1)$ indicates a random number that falls inside $(0,1)$ the normal distribution.

Mutation Operation: The DE method generates a mutation vector $M_{j,Y}$ for each individual $z_{j,Y}$ in the existing population (target vector) using the mutation operation. A specific mutation technique can generate a relevant mutation vector for each derived target vector. Several DE mutation strategies are established based on the varied generating ways of mutation people. The five most widely utilized mutation techniques are:

DE/rand/1:
$$M_{j,Y} = z_{r1,\,Y} + L \cdot (z_{r2,\,Y} - z_{r3,\,Y}) \tag{5}$$

DE/best/1:
$$M_{j,Y} = z_{best,\,Y} + L \cdot (z_{r1,\,Y} - z_{r2,\,Y}) \tag{6}$$

DE/rand-to-best/1:
$$M_{j,Y} = z_{j,\,Y} + L \cdot (z_{best,\,Y} - z_{j,\,Y}) + L \cdot (z_{r1,\,Y} - z_{r2,\,Y}) \tag{7}$$

DE/rand/2:
$$M_{j,Y} = z_{r1,\,Y} + L \cdot (z_{r2,\,Y} - z_{r3,\,Y}) + L \cdot (z_{r4,\,Y} - z_{r5,\,Y}) \tag{8}$$

Random exclusive integers are created and denoted by $r_1,\, r_2, r_3, r_4$ and r_5 within $[1,\,D]$. To scale a divergence vector, the scaling factor E is a positive constant value. In the *Yth* generation, $z_{best,\,Y}$ is an independent vector with the best global value.

Crossover Operation: To construct a test vector $v_{j,Y} = (v_{1,Y}, v_{2,Y}, v_{3,Y}, \ldots, v_{j,y})$, each pair of target vectors $z_{j,Y}$ and their matching mutation vectors $M_{j,Y}$ are crossed.

A binomial crossover is defined as follows in the DE algorithm:

$$v_{j,Y} = \begin{cases} M_{j,Y} & if\ (rand_i(0,1) \leq C)\ or\ (i = i_{rand},\ i = 1,2,3, \ldots, K) \\ z_{j,y} & Otherwise \end{cases} \tag{9}$$

where C denotes the crossover frequency and is a constant on $[0, 1]$. This is used to limit the quantity of the duplicated mutation vector. The selected integer on $[1, K]$, which is random, is denoted by i_{rand}.

Selection operation: If the parameter values reach the upper or lower bounds, they can be regenerated in a random and uniform manner within the specified range. The values of all the objective functions of the test vectors are then evaluated, and the selection operation is carried out. Each test vector's objective function $f(v_{j,Y})$ is matched to the associated target vector's optimal solution value of the associated target vector in the current sample. If the test vector's objective function is much less than or similar to the target vector's, the target vector is replaced by the test vector for the upcoming generation. The target vector is kept for the following generation if this is not the case.

$$z_{j,Y+1} = \begin{cases} v_{j,Y} & if \ (f(v_{j,Y}) \le f(z_{j,Y})) \\ z_{j,Y} & Otherwise \end{cases} \tag{10}$$

After obtaining the selected features $v_{j,Y}$, features are further refined using another threshold function called the selected standard error of mean (SSEoM). Using this new threshold function, the $Sl(k)$ features are selected as a final phase.

$$Tr = \begin{cases} Sl(k) & for \ v_{j,Y} \ge SM \\ Ignore, & Elsewhere \end{cases} \tag{11}$$

where Tr is a threshold function and SM is the standard error mean.

3.4.2. Reformed Binary Gray Wolf (RBGW) Optimization

The key update Equation for bGWO1 in this approach is provided as follows:

$$l_j^{h+1} = crossover \ (l_1, l_2, l_3) \tag{12}$$

$crossover(l, m, n)$ is an appropriate crossover between solutions l, m, n and l_1, l_2, l_3, which are binary vectors showing the effect of wolves moving towards alpha, beta, and delta grey wolves, in that order. l_1, l_2, l_3 can be computed by using the following Equation (13):

$$l_1^t = \begin{cases} 1 & if \ (l_a^t + bistep_a^t) \ge 1 \\ 0 & Otherwise \end{cases} \tag{13}$$

where position vector in dimension t is denoted by l_1^t and binary step is represented by $bistep_a^t$ in dimension t. It can be computed by using Equation (15):

$$bistep_a^t = \begin{cases} 1 & if \ costep_a^t \ge rand \\ 0 & Otherwise \end{cases} \tag{14}$$

where $rand$ is an integer picked at random from a uniformly distributed $\in [0, 1]$, and the continuous value of the size step is denoted by $costep_a^t$; this can be computed by the following Equation (15):

$$costep_a^t = \frac{1}{1 + e^{-10(X_1^t Y_a^t - 0.5)}} \tag{15}$$

where X_1^t and D_a^t are computed through Equations (16) and (17) that were later employed for the threshold selection as follows:

$$\vec{X} = 2c \cdot \vec{r_1} - c \tag{16}$$

$$\vec{D}_\alpha = \left| \vec{C_1} \cdot \vec{X}_a - \vec{X} \right| \tag{17}$$

$$l_2^t = \begin{cases} 1 & if\ \left(l_b^t + bistep_b^t\right) \geq 1 \\ 0 & Otherwise \end{cases} \tag{18}$$

where in Equation (16), $\overrightarrow{X}$ is the updated position of prey, $\overrightarrow{r_1}$ denotes the random distribution, and c is constantly reduced in the scope of (2,0). In Equation (18), $\overrightarrow{D_\alpha}$ represent the distances of prey from each gray wolf and $\overrightarrow{C_1}$ represent the coefficient variable. In Equation (18), the position vector in dimension t is denoted by l_2^t and the binary step is represented by $bistep_b^t$ in dimension t. It can be computed by using the following Equation (19):

$$bistep_b^t = \begin{cases} 1 & if\ costep_b^t \geq rand \\ 0 & Otherwise \end{cases} \tag{19}$$

where *rand* is an integer picked at random from an uniformly distributed $\in [0,1]$ and the continuous valued of size step is denoted by $costep_b^t$; this can be computed by the following Equation (20):

$$costep_b^t = \frac{1}{1 + e^{-10(X_1^t D_b^t - 0.5)}} \tag{20}$$

where D_b^t in dimension t can be computed by Equation (21).

$$D_b^t = \left| \overrightarrow{C_2} \cdot \overrightarrow{X_b} - \overrightarrow{X} \right| \tag{21}$$

$$l_3^t = \begin{cases} 1 & if\ \left(l_c^t + bistep_c^t\right) \geq 1 \\ 0 & Otherwise \end{cases} \tag{22}$$

where the position vector in dimension t is denoted by l_3^t and the binary step is represented by $bistep_c^t$ in dimension t. It can be computed by using the following Equation (23):

$$bistep_c^t = \begin{cases} 1 & if\ costep_c^t \geq rand \\ 0 & Otherwise \end{cases} \tag{23}$$

where *rand* is an integer picked at random and uniformly distributed $\in [0,1]$, and the continuous value of the size step is denoted by $costep_c^t$; this can be computed by Equation (24).

$$costep_c^t = \frac{1}{1 + e^{-10(X_1^t Y_c^t - 0.5)}} \tag{24}$$

where Y_c^t in dimension t can be computed by Equation (25).

$$\overrightarrow{Y_c} = \left| \overrightarrow{S_3} \cdot \overrightarrow{R_c} - \overrightarrow{R} \right| a = 1, \tag{25}$$

A stochastic crossover process is used per dimension to crossover u, v, w solutions.

$$l_t = \begin{cases} u_t & if\ rand < \frac{1}{2} \\ v_t & \frac{1}{2} \leq rand < \frac{2}{5} \\ w_t & Otherwise \end{cases} \tag{26}$$

Binary values are u_t, v_t and w_t. These are three parameters in dimension t. The output of the crossover is denoted by l_t in dimension t.

The algorithm is summarized in Algorithm 1.

Algorithm 1. Reformed Features Optimization Algorithm

Input: g the pack's total number of grey wolves,
Gth the number of optimization iterations.
Output: l_α Binary position of the grey wolf that is optimal,
$f(l_\alpha)$ Best fitness value
Begin
1.   Create a population of g wolves with random positions $\in [0, 1]$
2.   Find a, b, c solutions that are based on fitness.
3.   **While** Criteria for stopping not met **do**
        For each $wolf_j \in pack$ **do**
        Calculate l_1, l_2, l_3 using Equations (13), (18) and (22).
        $l_j^{h+1} \leftarrow$ crossover among l_1, l_2, l_3 using Equation (26). **end**

        I   Update c, X, S.
        II   Examine the individual wolf positions.
        III   Update a, b, c.
   End

3.5. Feature Fusion and Classification

The best selected features from the RDE and RGW algorithms are finally fused in one feature vector for the final classification. For the fusion of selected deep features, a probability-based serial approach is adopted. In this approach, initially probability is computed for both selected vectors and only one feature is employed based on the high probability value. Based on the high probability value feature, a comparison is conducted and features are fused in one matrix. The main purpose of this comparison is to tackle the problem of redundant features of both vectors. The final fused features are next classified using machine learning algorithms for the final classification. The size of the vector is 4788×704 after fusion.

4. Experimental Results and Analysis

Experimental Setup: During the training of fine-tuned deep learning model, the following hyper parameters are employed, such as a learning rate of 0.001, mini batch size of 16, epochs at 200, the optimization method is Adam, and the feature activation function is sigmoid. Moreover, the multiclass cross entropy loss function is employed for the calculation of loss.

All experiments are performed on MATLAB2020b using a desktop computer Core i7 with 8GB of graphics card and 16GB RAM.

The following experiments have been performed to validate the proposed method:

(i)   Classification using modified DarkNet53 features in training/testing ratio of 50:50;
(ii)   Classification using modified DarkNet53 features in training/testing ratio 70:30;
(iii)   Classification using modified DarkNet53 features in training/testing ratio 60:40;
(iv)   Classification using DE based best feature selection on training/testing ratio 50:50;
(v)   Classification using the Gray Wolf algorithm based best feature selection in training/testing ratio 50:50, and
(vi)   Fusion of best selected features and classification using several classifiers, including the support vector machine (SVM), KNN, decision trees (DT), etc.

The results of the proposed method are discussed in this section in terms of tables and visual plots. Different training and testing ratios are considered for analysis, such as 70:30, 60:40, and 50:50. The cross-validation value is selected at 10 for all experiments.

4.1. Results

The results of the first experiment are given in Table 2. This table presented the best accuracy obtained of 99.3% for Cubic SVM. A few other parameters are also computed for this classifier, such as sensitivity rate, precision rate, F1 score, FNR, and time complexity, and their values are 99.2, 99.2, 99.2, 0.8%, and 20.69 (s), respectively. The Q-SVM and

MGSVM obtained the second-best accuracy of 99.2%. The rest of the classifiers such as ESD, LSVM, ESKNN, FKNN, LD, CGSVM, and WKNN and their accuracy values are 98.9%, 98.9%, 98.7%, 98.7%, 98.6%, 98.2%, and 97.9%, respectively.

Table 2. Classification results of DarknNet53 using ultrasound images, where the training/testing ratio is 50:50.

Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR	Classification Time (s)
CSVM	99.2	99.2	99.2	99.3	0.8	200.697
MGSVM	99.2	99.2	99.2	99.2	0.8	207.879
QSVM	99.16	99.16	99.16	99.2	0.84	159.21
ESD	98.8	98.8	98.8	98.9	1.2	198.053
LSVM	98.93	98.93	98.93	98.9	1.07	122.98
ESKNN	98.6	98.6	98.6	98.7	1.4	189.79
FKNN	98.7	98.7	98.7	98.7	1.3	130.664
LD	98.6	98.6	98.6	98.6	1.4	120.909
CGSVM	98.16	98.2	98.17	98.2	1.84	133.085
WKNN	97.9	97.93	97.91	97.9	2.1	129.357

The sensitivity rate of Cubic SVM is validated through the confusion matrix illustrated in Figure 6. In addition, the computational time of each classifier is noted, and the best time is 120.909 (s) for LDA, and the worst time is 207.879 (s) for MGSVM.
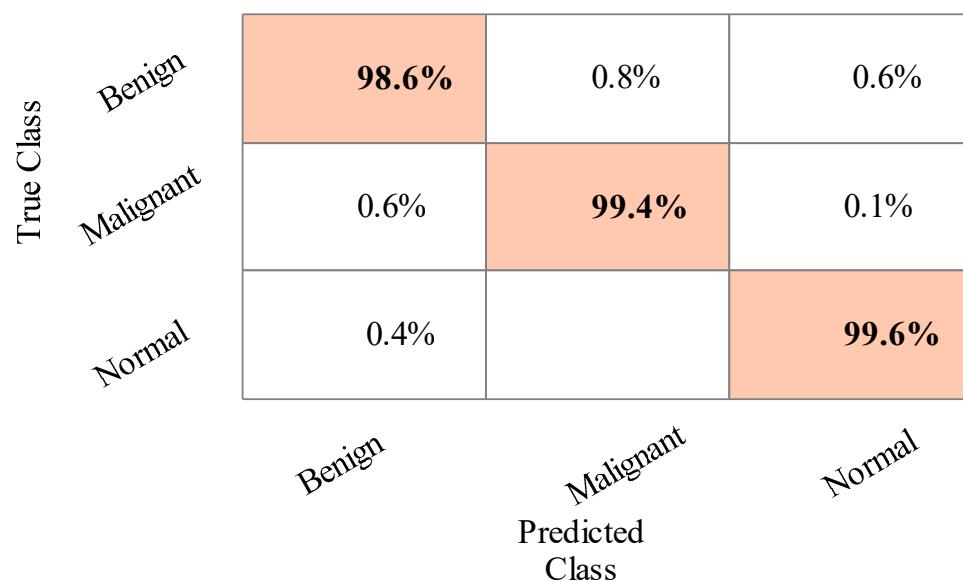
Figure 6. Confusion matrix of Cubic SVM for training/testing ratio of 50:50.

The results of the second experiment are given in Table 3. The best accuracy of 99.3% was obtained for Cubic SVM. A few other parameters are also computed, such as sensitivity rate, precision rate, F1 score, accuracy, FNR, and time complexity, and their values are 99.3%, 99.3%, 99.3%, 99.3%, 0.7%, and 11.112 (s), respectively. The MGSVM and Q-SVM classifiers obtained the second-best accuracy of 99.3% and 99.2%, respectively. The rest of the classifiers also achieved better performance. The confusion matrix of the Cubic SVM is illustrated in Figure 7. In addition, the computational time of each classifier is noted, and the minimum time is 111.112 (s) for the Cubic SVM, whereas the highest time is 167.126 (s) for ESD. When comparing the results of this experiment in Table 2, the classification accuracy is found to be consistent, but the computational time is minimized.

Table 3. Classification results of DarkNet53 using ultrasound images, where the training/testing ratio is 70:30.

Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR (%)	Classification Time (s)
CSVM	99.3	99.3	99.3	99.3	0.7	111.112
MGSVM	99.2	99.2	99.2	99.3	0.8	113.896
QSVM	99.16	99.2	99.17	99.2	0.84	125.304
ESD	99.0	99.03	99.01	99.0	1.0	167.126
LSVM	99.06	99.1	99.07	99.1	0.94	120.608
ESKNN	98.06	98.03	98.04	98.1	1.94	141.71
FKNN	97.73	97.76	97.74	97.7	2.27	124.324
LD	97.6	97.6	97.6	97.7	2.4	131.507
CGSVM	98.06	98.06	98.06	98.1	1.94	155.501
WKNN	96.03	96.13	96.07	96.0	3.97	127.675
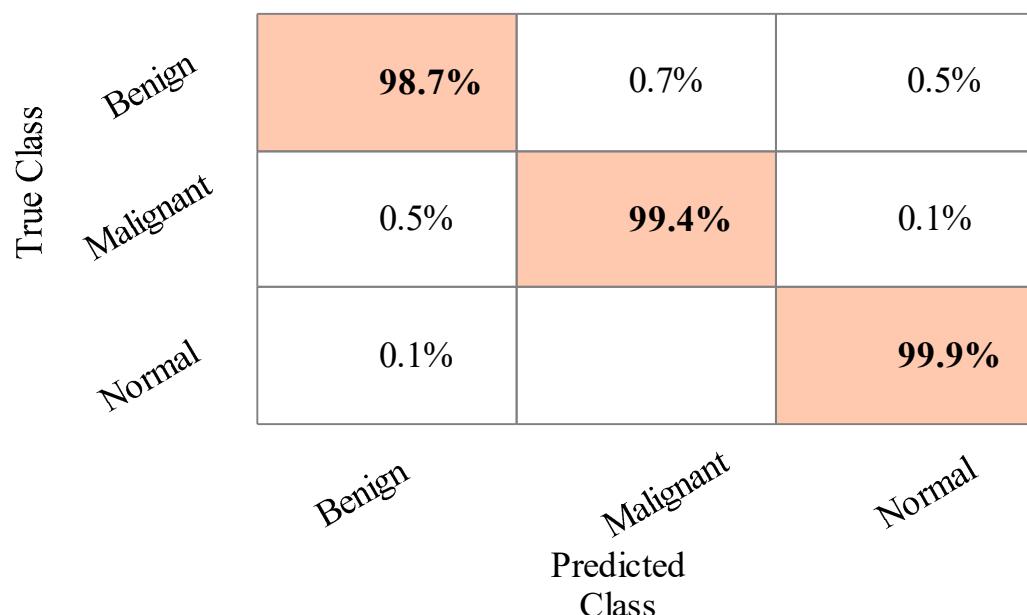
Figure 7. Confusion matrix of Cubic SVM for training/testing ratio of 70:30.

The results of the third experiment are given in Table 4. This table presented the best accuracy obtained at 98.9% for Cubic SVM. The MGSVM and Q-SVM obtained the second-best accuracy of 98.7%. The rest of the classifiers such as ESD, LSVM, ESKNN, FKNN, LD, CGSVM, and WKNN, and their accuracy values are 98.7%, 98.6%, 98%, 97.8%, 98.1%, 97.9% and 97.2%, respectively. The confusion matrix of Cubic SVM is illustrated in Figure 8. In addition, the computational time of each classifier is also noted, and the best time is 76.2 (s) for Cubic SVM and the worst time is 107.679 (s) for the ESKNN classifier. The accuracy of classifiers from experiments (i)–(iii) using different training/testing ratios is summarized in Figure 9. This figure illustrated that the performance at 50:50 is overall better than the rest of the selected ratios.

Table 4. Classification results of DarkNet53 using ultrasound images, where the training/testing ratio is 60:40.

| Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR (%)	Classification Time (s)
CSVM	98.9	98.9	98.9	98.9	1.1	107.697
MGSVM	98.7	98.7	98.7	98.7	1.3	103.149
QSVM	98.7	98.7	98.7	98.7	1.3	89.049
ESD	98.6	98.7	98.65	98.7	1.4	94.31
LSVM	98.5	98.5	98.5	98.6	1.5	68.827
ESKNN	97.9	98	97.95	98.0	2.1	79.34
FKNN	97.8	97.8	98.24	97.8	2.2	84.537
LD	98.1	98.1	98.1	98.1	1.9	85.317
CGSVM	97.8	97.9	98.34	97.9	2.2	76.2
WKNN	97.2	97.2	97.2	97.2	2.8	81.191
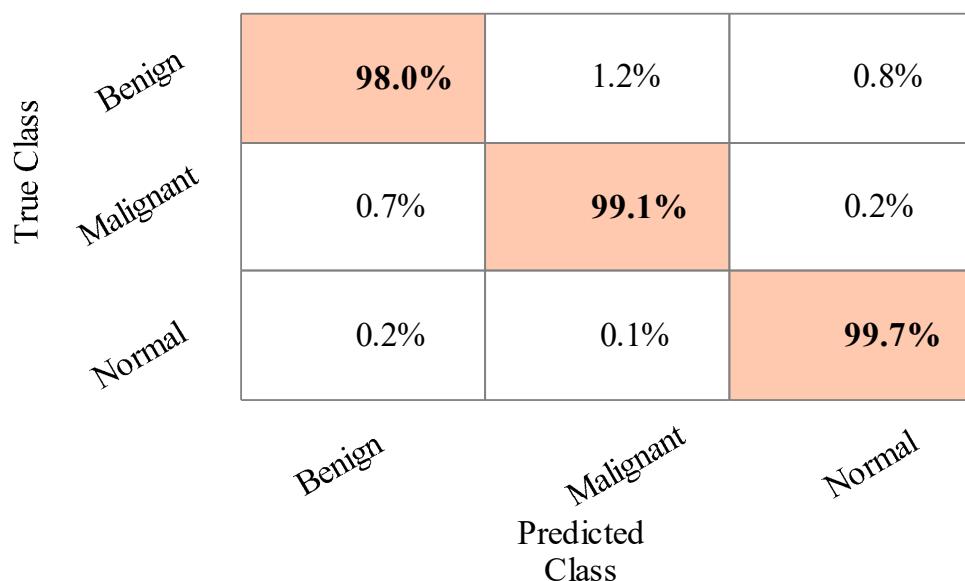
Figure 8. Confusion matrix of Cubic SVM for the training/testing ratio of 60:40.
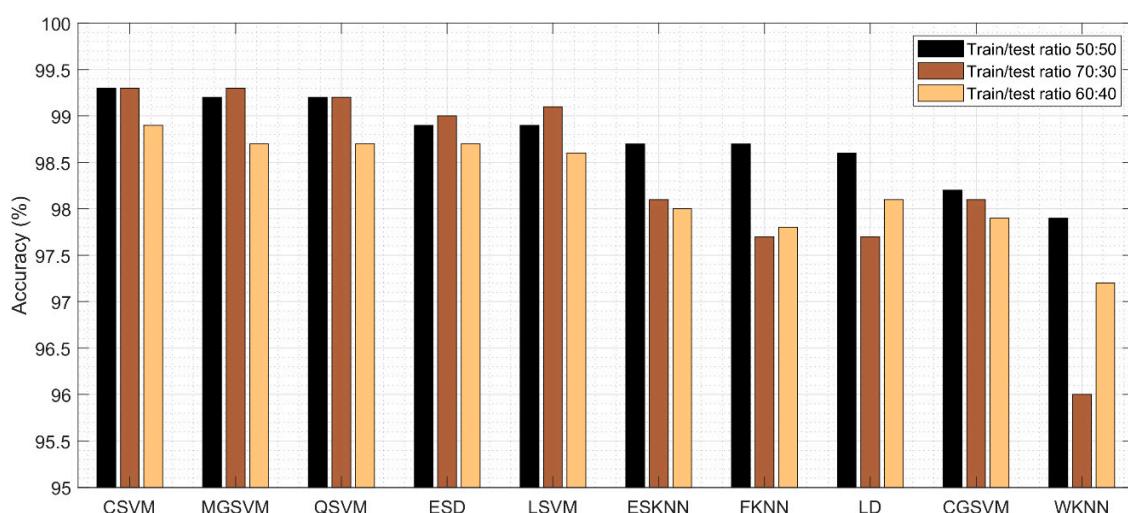
Figure 9. Summary of DarkNet53 classification accuracy using different training/testing ratios.

Table 5 presents the results of the fourth experiment. In this experiment, a 50:50 training/testing ratio is used. The best features are selected using the binary DE method.

The 99.1% accuracy is achieved by Cubic SVM after feature selection. A few other parameters are also computed for this classifier, such as sensitivity rate, precision rate, F1 score accuracy, FNR, and time complexity, and their values are 99.1%, 99.06%, 99.08%, 1, 0.9 and 16.082, respectively. The confusion matrix of Cubic SVM is illustrated in Figure 10. The computational time of each classifier is also noted, and the best time is 28.082 (s) for the CSVM classifier, and the worst time is 42.829 (s) for the WKNN classifier. This shows that the computational time after the selection process is significantly minimized compared with the time given in Tables 2 and 3.

Table 5. Classification results of binary differential evolution selector using ultrasound images, where the training/testing ratio is 50:50.

Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR	Classification Time (s)
CSVM	99.10	99.06	99.08	99.1	0.9	28.082
MGSVM	99.13	99.13	99.13	99.1	0.87	41.781
QSVM	99.10	99.10	99.1	99.1	0.9	35.448
ESD	98.70	98.70	98.7	98.7	1.3	42.74
LSVM	98.90	98.86	98.88	98.9	1.1	33.073
ESKNN	98.40	98.36	98.38	98.4	1.6	37.555
FKNN	98.26	98.30	98.28	98.3	1.74	35.349
LD	98.50	98.50	98.5	98.5	1.5	35.213
CGSVM	98.43	98.43	98.43	98.4	1.57	39.686
WKNN	97.00	97.10	97.05	97.0	3.0	42.829
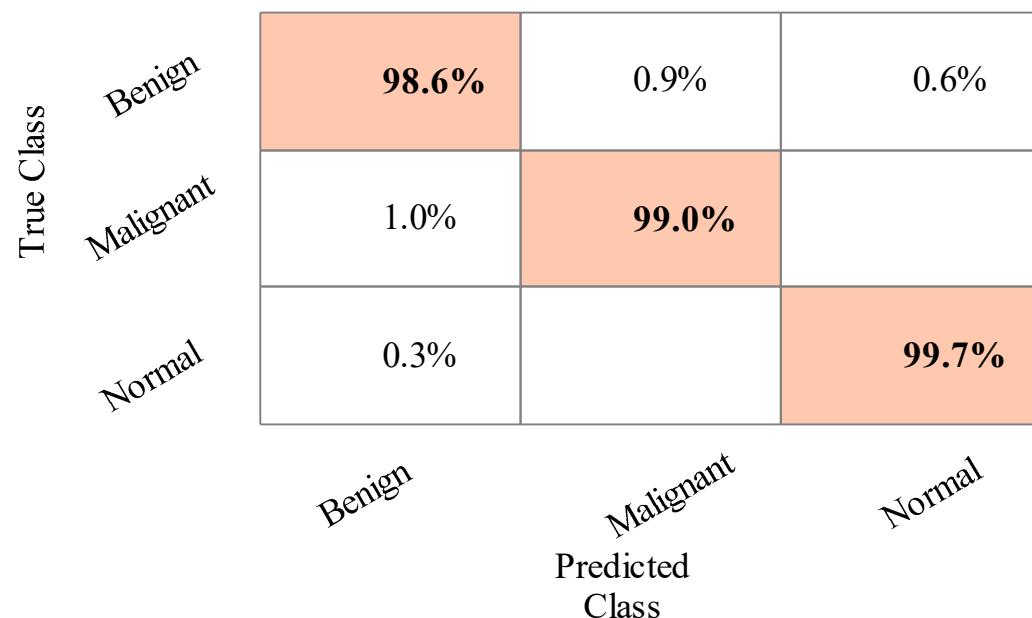
Figure 10. Confusion matrix of Cubic SVM for the features selected using DE and the train/test ratio of 50:50.

The results of the fifth experiment are given in Table 6. In this experiment, the binary gray wolf optimization algorithm is implemented and selects the best features for the final classification. This table presents the best accuracy obtained of 99.1% for Cubic SVM. A few other parameters are also computed for this classifier, such as sensitivity rate, precision rate, F1 score, accuracy, FNR, and time complexity, and their values are 99.06%, 99.1%, 99.08%, 1, 0.94, and 15.239, respectively. The confusion matrix of Cubic SVM is illustrated in Figure 11. The computational time of each classifier is also noted, and the best time is 25.239 (s) for CSVM. This table shows that the overall time is minimized, and the accuracy is consistent with Tables 2 and 3.

Table 6. Classification results of binary gray wolf optimization selector using ultrasound images, where the training/testing ratio is 50:50.

Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR	Classification Time (s)
CSVM	99.06	99.1	99.08	99.1	0.94	25.239
MGSVM	98.96	98.96	98.96	99.0	1.04	29.732
QSVM	98.96	98.96	98.96	99.0	1.04	33.632
ESD	98.5	98.5	98.5	98.5	1.5	30.823
LSVM	98.7	98.7	98.7	98.7	1.3	35.774
ESKNN	98.5	98.5	98.5	98.5	1.5	31.585
FKNN	98.36	98.36	98.36	98.4	1.64	40.854
LD	98.2	98.2	98.2	98.3	1.8	38.5073
CGSVM	98.06	98.1	98.08	98.1	1.94	32.396
WKNN	97.2	97.2	97.2	97.2	2.8	30.698
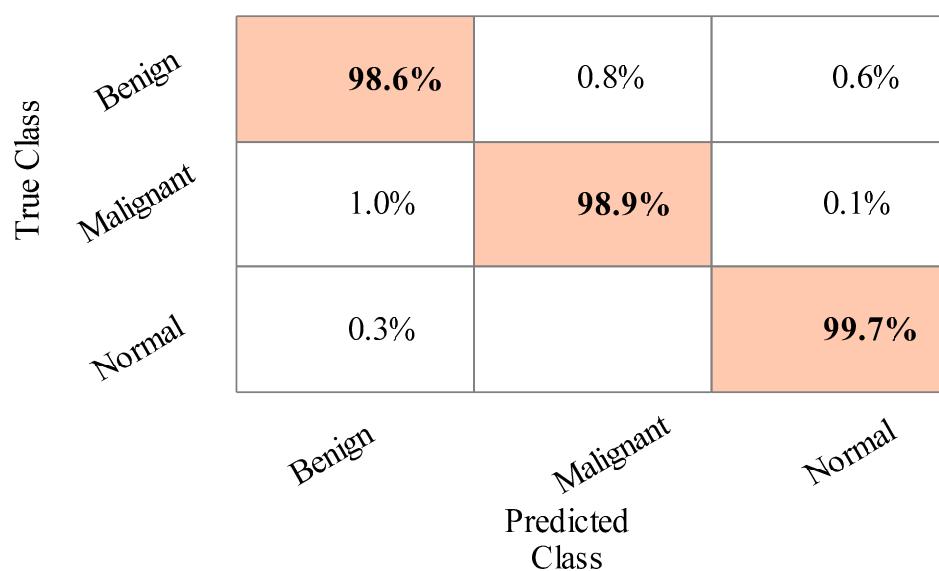
Figure 11. Confusion matrix of Cubic SVM for BGWO-based best feature selection.

Finally, the best selected features are fused using the proposed approach. The results are given in Table 7. This table presented the best accuracy obtained with 99.1% for Cubic SVM. The confusion matrix of Cubic SVM is illustrated in Figure 12. In this figure, the diagonal values show the correct predicted values. In addition, the computational time of each classifier is noted, and the best time is 13.599 (s) for the CSVM classifier.

Table 7. Classification results using the feature fusion of DE and BGWO using ultrasound images, where the training/testing ratio is 50:50.

Classifier	Sensitivity (%)	Precision (%)	F1 Score (%)	Accuracy (%)	FNR (%)	Classification Time (s)
CSVM	99.06	99.06	99.06	99.18	0.94	13.599
MGSVM	99.10	99.10	99.10	99.16	0.9	15.659
QSVM	98.96	98.96	98.96	99.30	1.04	17.601
ESD	98.76	98.80	98.78	98.90	1.24	26.240
LSVM	98.93	98.90	98.91	99.00	1.07	19.185
ESKNN	98.56	98.60	98.58	98.90	1.44	22.425
FKNN	98.36	98.36	98.36	98.74	1.64	24.508
LD	98.40	98.40	98.40	98.40	1.60	21.045
CGSVM	98.20	98.20	98.20	98.30	1.80	20.627
WKNN	97.46	97.53	97.49	98.10	2.54	18.734
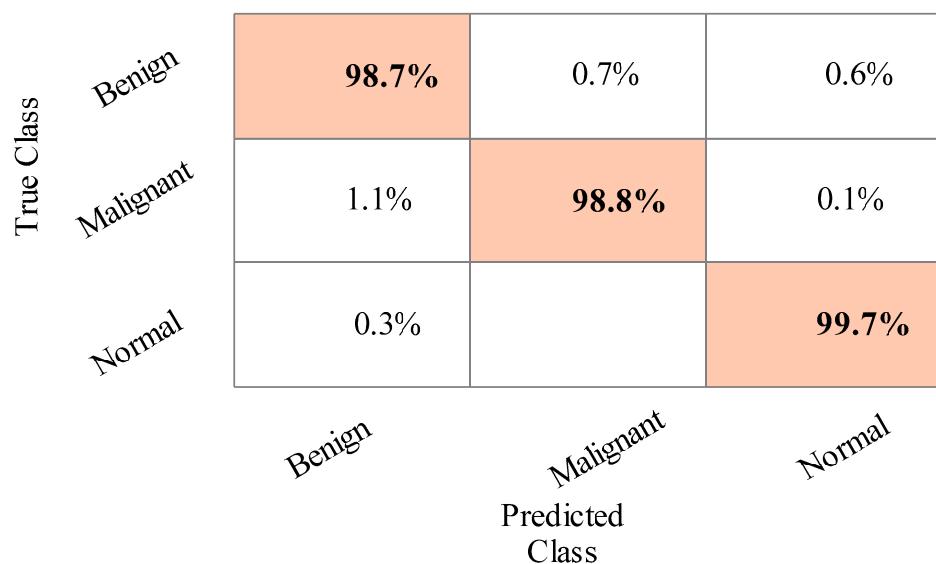
Figure 12. Confusion matrix of Cubic SVM after the proposed feature fusion approach.

Figure 13 compares the computational time while using the original features, the selected features based on DE, the feature selection based on BGWO, and feature fusion. In this figure, it is illustrated that the computational time of the original features is high, which was decreased after the feature selection step. Further, the proposed fusion process improves the performance in terms of computational time and consistency with the accuracy.


	CSVM	MGSVM	QSVM	ESD	LSVM	ESKNN	FKNN	LD	CGSVM	WKNN
Original	200.697	207.879	159.21	198.053	122.98	189.79	130.664	120.909	133.085	129.357
DE	28.082	41.781	35.448	42.74	33.073	37.555	35.349	35.213	39.686	42.829
BGWO	25.239	29.732	33.632	30.823	35.774	31.585	40.854	38.5073	32.396	30.698
Proposed	13.599	15.659	17.601	26.24	19.185	22.425	24.508	21.045	20.627	18.734
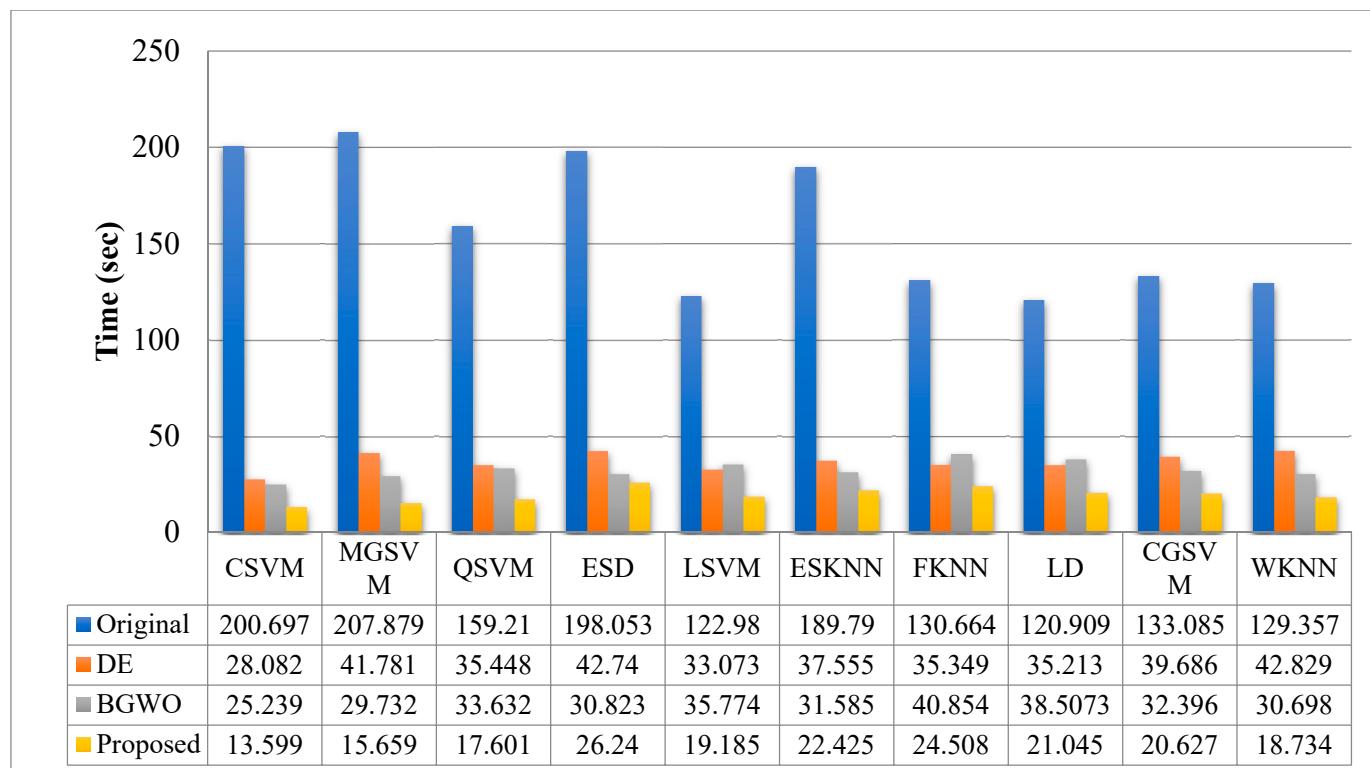
Figure 13. Computational time-based comparison of each step using the proposed framework.

4.2. Statistical Analysis

For statistical analysis and comparison of the results, we used the post-hoc Nemenyi test. Demšar [62] has suggested using the Nemenyi test to compare techniques in a paired manner. The test determines a critical difference (CD) value for a given degree of confidence

α. If the difference in the average ranks of two techniques exceeds the CD value, the null hypothesis, H0, that both methods perform equally well, is rejected.

The results of statistical analysis are summarized in Figure 14 (mean ranks of classifiers) and Figure 15 (mean ranks of feature selection methods). The best classifier is CSVM, but MGSVM and QSVM also show very good results, in terms of accuracy, which are not significantly different from CSVM. The best feature selection method among the four methods analyzed is the proposed feature fusion approach, which is significantly better than other approaches (DE, BGWO, and original).
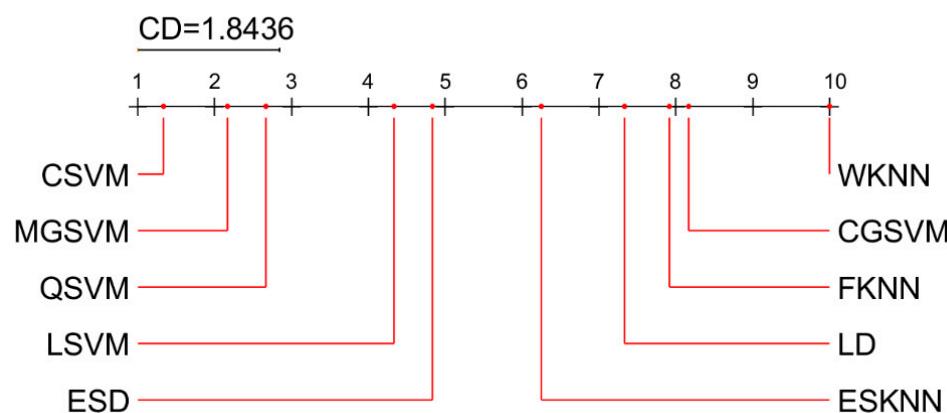

Figure 14. Critical difference diagram from the Nemenyi test: a comparison of classifiers (α = 0.05).
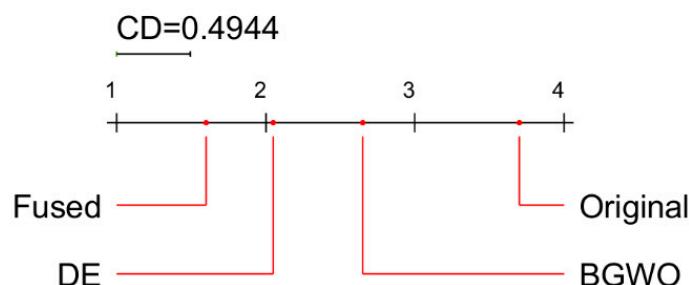

Figure 15. Critical difference diagram from the Nemenyi test: a comparison of feature selection methods (α = 0.05).

4.3. Comparison with the State of the Art

The proposed method is compared with the state-of-the-art techniques, as given in Table 8. In [63], the authors used ultrasound images and achieved an accuracy of 73%. In [64], the adaptive histogram equalization method was used to enhance ultrasound images and obtained an accuracy of 89.73%. In [52], a CAD system was presented for tumor identification that combines an imaging fusion method with various formats of image content and ensembles of multiple CNN architectures. The accuracy achieved for this data set was 94.62%. In [65], the source breast ultrasound image was first processed using bilateral filtering and fuzzy enhancement methods. The accuracy achieved was 95.48%. In [66], authors implemented a semi-supervised generative adversarial network (GAN) model and achieved an accuracy of 90.41%. The proposed method achieved an accuracy of 99.1% using a BUSi augmented dataset, where the computational time is 13.599 (s).

Table 8. Comparison with the state-of-the-art techniques.

Reference	Year	Accuracy (%)	Time (s)
Cao et al. [63]	2020	73.0	-
Illesnami et al. [64]	2021	89.73	-
Pang et al. [66]	2021	90.41	-
Moon et al. [52]	2020	94.62	-
Zhuang et al. [65]	2021	95.48	-
Proposed		99.1	13.599

5. Conclusions

We proposed an automated system for breast cancer classification using ultrasound images. The proposed method is based on a few sequential steps. Initially, the breast ultrasound data are augmented and then retrained using a DarkNet-53 deep learning model. Next, the features were extracted from the pooling layer and then the best feature was selected using two different optimization algorithms such as the reformed BGWO and the reformed DE. The selected features are finally fused using a proposed approach that is later classified using machine learning algorithms. Several experiments were performed, and the proposed method achieved the best accuracy of 99.1% (using feature fusion and CSVM classifier). The comparison with recent techniques shows improvement in the results using the proposed framework. The strength of this work is: (i) augmentation of the dataset improved the training strength, (ii) the selection of best features reduced the irrelevant features, and (iii) the fusion method further reduced the computational time and consistency of accuracy.

In future, we will focus on two key steps: (i) increasing the size of the database, and (ii) designing a CNN model from scratch for breast tumor classification. We will discuss our proposed model with ultrasound imaging specialists and medical doctors, aiming for practical implementation at hospitals.

Author Contributions: Data curation, K.J. and M.A.K.; Formal analysis, K.J., M.A., Y.-D.Z. and R.D.; Funding acquisition, A.M.; Investigation, K.J., M.A., U.T. and A.H.; Methodology, M.A.K.; Resources, K.J.; Software, K.J.; Validation, M.A.K., M.A., U.T., Y.-D.Z., A.H., A.M. and R.D.; Writing—original draft, K.J., M.A.K., M.A., Y.-D.Z. and A.H.; Writing—review & editing, A.M. and R.D. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this paper is available from https://scholar.cu.edu.eg/?q=afahmy/pages/dataset (accessed on 20 November 2021).

Abbreviations

CNN	Convolutional neural network
RDE	Reformed differential evaluation
RGW	Reformed differential evaluation
MRI	Magnetic resonance imaging
CAD	Computer-aided diagnosis
AI	Artificial intelligence
GA	Genetic algorithm
PSO	Particle swarm optimization

HT	Hilbert transform
KNN	K-Nearest neighbor
ML	Machine learning
ROI	Region of interest
TL	Transfer learning
DRS	Deep representation scaling
Di-CNN	Dilated semantic segmentation network
LUPI	Learning using privileged information
MMD	Maximum mean discrepancy
DDSTN	Doubly supervised TL network
BTS	Breast tumor section
ELM	Extreme learning machine
DT	Decision tree
SVM	Support vector machine
WKNN	Weighted KNN
QSVM	Quadratic SVM
CGSVM	Cubic gaussian SVM
LD	Linear discriminant
ESKNN	Ensemble subspace KNN
ESD	Ensemble subspace discriminant
FNR	False negative rate

References

1. Yu, K.; Chen, S.; Chen, Y. Tumor Segmentation in Breast Ultrasound Image by Means of Res Path Combined with Dense Connection Neural Network. *Diagnostics* **2021**, *11*, 1565. [CrossRef] [PubMed]
2. Feng, Y.; Spezia, M.; Huang, S.; Yuan, C.; Zeng, Z.; Zhang, L.; Ji, X.; Liu, W.; Huang, B.; Luo, W. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis.* **2018**, *5*, 77–106. [CrossRef] [PubMed]
3. Badawy, S.M.; Mohamed, A.E.-N.A.; Hefnawy, A.A.; Zidan, H.E.; GadAllah, M.T.; El-Banby, G.M. Automatic semantic segmentation of breast tumors in ultrasound images based on combining fuzzy logic and deep learning—A feasibility study. *PLoS ONE* **2021**, *16*, e0251899. [CrossRef] [PubMed]
4. Zhang, S.-C.; Hu, Z.-Q.; Long, J.-H.; Zhu, G.-M.; Wang, Y.; Jia, Y.; Zhou, J.; Ouyang, Y.; Zeng, Z. Clinical implications of tumor-infiltrating immune cells in breast cancer. *J. Cancer* **2019**, *10*, 6175. [CrossRef] [PubMed]
5. Irfan, R.; Almazroi, A.A.; Rauf, H.T.; Damaševičius, R.; Nasr, E.; Abdelgawad, A. Dilated Semantic Segmentation for Breast Ultrasonic Lesion Detection Using Parallel Feature Fusion. *Diagnostics* **2021**, *11*, 1212. [CrossRef]
6. Faust, O.; Acharya, U.R.; Meiburger, K.M.; Molinari, F.; Koh, J.E.W.; Yeong, C.H.; Ng, K.H. Comparative assessment of texture features for the identification of cancer in ultrasound images: A review. *Biocybern. Biomed. Eng.* **2018**, *38*, 275–296. [CrossRef]
7. Pourasad, Y.; Zarouri, E.; Salemizadeh Parizi, M.; Salih Mohammed, A. Presentation of Novel Architecture for Diagnosis and Identifying Breast Cancer Location Based on Ultrasound Images Using Machine Learning. *Diagnostics* **2021**, *11*, 1870. [CrossRef]
8. Sainsbury, J.; Anderson, T.; Morgan, D. Breast cancer. *BMJ* **2000**, *321*, 745–750. [CrossRef]
9. Sun, Q.; Lin, X.; Zhao, Y.; Li, L.; Yan, K.; Liang, D.; Sun, D.; Li, Z.-C. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: Don't forget the peritumoral region. *Front. Oncol.* **2020**, *10*, 53. [CrossRef]
10. Almajalid, R.; Shan, J.; Du, Y.; Zhang, M. Development of a deep-learning-based method for breast ultrasound image segmentation. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; IEEE: New York, NY, USA, 2018; pp. 1103–1108.
11. Ouahabi, A. *Signal and Image Multiresolution Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
12. Ahmed, S.S.; Messali, Z.; Ouahabi, A.; Trepout, S.; Messaoudi, C.; Marco, S. Nonparametric denoising methods based on contourlet transform with sharp frequency localization: Application to low exposure time electron microscopy images. *Entropy* **2015**, *17*, 3461–3478. [CrossRef]
13. Sood, R.; Rositch, A.F.; Shakoor, D.; Ambinder, E.; Pool, K.-L.; Pollack, E.; Mollura, D.J.; Mullen, L.A.; Harvey, S.C. Ultrasound for breast cancer detection globally: A systematic review and meta-analysis. *J. Glob. Oncol.* **2019**, *5*, 1–17. [CrossRef] [PubMed]
14. Byra, M. Breast mass classification with transfer learning based on scaling of deep representations. *Biomed. Signal Process. Control* **2021**, *69*, 102828. [CrossRef]
15. Chen, D.-R.; Hsiao, Y.-H. Computer-aided diagnosis in breast ultrasound. *J. Med. Ultrasound* **2008**, *16*, 46–56. [CrossRef]
16. Moustafa, A.F.; Cary, T.W.; Sultan, L.R.; Schultz, S.M.; Conant, E.F.; Venkatesh, S.S.; Sehgal, C.M. Color doppler ultrasound improves machine learning diagnosis of breast cancer. *Diagnostics* **2020**, *10*, 631. [CrossRef] [PubMed]
17. Shen, W.-C.; Chang, R.-F.; Moon, W.K.; Chou, Y.-H.; Huang, C.-S. Breast ultrasound computer-aided diagnosis using BI-RADS features. *Acad. Radiol.* **2007**, *14*, 928–939. [CrossRef]

18. Lee, J.-H.; Seong, Y.K.; Chang, C.-H.; Park, J.; Park, M.; Woo, K.-G.; Ko, E.Y. Fourier-based shape feature extraction technique for computer-aided b-mode ultrasound diagnosis of breast tumor. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; IEEE: New York, NY, USA, 2012; pp. 6551–6554.

19. Ding, J.; Cheng, H.-D.; Huang, J.; Liu, J.; Zhang, Y. Breast ultrasound image classification based on multiple-instance learning. *J. Digit. Imaging* **2012**, *25*, 620–627. [CrossRef] [PubMed]

20. Bing, L.; Wang, W. Sparse representation based multi-instance learning for breast ultrasound image classification. *Comput. Math. Methods Med.* **2017**, *2017*, 7894705. [CrossRef]

21. Prabhakar, T.; Poonguzhali, S. Automatic detection and classification of benign and malignant lesions in breast ultrasound images using texture morphological and fractal features. In Proceedings of the 2017 10th Biomedical Engineering International Conference (BMEiCON), Hokkaido, Japan, 31 August–2 September 2017; IEEE: New York, NY, USA, 2017; pp. 1–5.

22. Zhang, Q.; Suo, J.; Chang, W.; Shi, J.; Chen, M. Dual-modal computer-assisted evaluation of axillary lymph node metastasis in breast cancer patients on both real-time elastography and B-mode ultrasound. *Eur. J. Radiol.* **2017**, *95*, 66–74. [CrossRef]

23. Gao, Y.; Geras, K.J.; Lewin, A.A.; Moy, L. New frontiers: An update on computer-aided diagnosis for breast imaging in the age of artificial intelligence. *Am. J. Roentgenol.* **2019**, *212*, 300–307. [CrossRef]

24. Geras, K.J.; Mann, R.M.; Moy, L. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. *Radiology* **2019**, *293*, 246–259. [CrossRef]

25. Fujioka, T.; Mori, M.; Kubota, K.; Oyama, J.; Yamaga, E.; Yashima, Y.; Katsuta, L.; Nomura, K.; Nara, M.; Oda, G. The utility of deep learning in breast ultrasonic imaging: A review. *Diagnostics* **2020**, *10*, 1055. [CrossRef] [PubMed]

26. Zahoor, S.; Lali, I.U.; Javed, K.; Mehmood, W. Breast cancer detection and classification using traditional computer vision techniques: A comprehensive review. *Curr. Med. Imaging* **2020**, *16*, 1187–1200. [CrossRef] [PubMed]

27. Kadry, S.; Rajinikanth, V.; Taniar, D.; Damaševičius, R.; Valencia, X.P.B. Automated segmentation of leukocyte from hematological images—A study using various CNN schemes. *J. Supercomput.* **2021**, 1–21. [CrossRef]

28. Abayomi-Alli, O.O.; Damaševičius, R.; Misra, S.; Maskeliūnas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. *Turk. J. Electr. Eng. Comput. Sci.* **2021**, *29*, 2600–2614. [CrossRef]

29. Maqsood, S.; Damaševičius, R.; Maskeliūnas, R. Hemorrhage detection based on 3d cnn deep learning framework and feature fusion for evaluating retinal abnormality in diabetic patients. *Sensors* **2021**, *21*, 3865. [CrossRef]

30. Hussain, N.; Kadry, S.; Tariq, U.; Mostafa, R.R.; Choi, J.-I.; Nam, Y. Intelligent Deep Learning and Improved Whale Optimization Algorithm Based Framework for Object Recognition. *Hum. Cent. Comput. Inf. Sci.* **2021**, *11*, 34.

31. Kadry, S.; Parwekar, P.; Damaševičius, R.; Mehmood, A.; Khan, J.A.; Naqvi, S.R.; Khan, M.A. Human gait analysis for osteoarthritis prediction: A framework of deep learning and kernel extreme learning machine. *Complex Intell. Syst.* **2021**, 1–19. [CrossRef]

32. Dhungel, N.; Carneiro, G.; Bradley, A.P. The automated learning of deep features for breast mass classification from mammograms. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016*; Springer: Cham, Switzerland, 2016; pp. 106–114.

33. Alhaisoni, M.; Tariq, U.; Hussain, N.; Majid, A.; Damaševičius, R.; Maskeliūnas, R. COVID-19 Case Recognition from Chest CT Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel Feature Fusion. *Sensors* **2021**, *21*, 7286.

34. Odusami, M.; Maskeliūnas, R.; Damaševičius, R.; Krilavičius, T. Analysis of features of alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network. *Diagnostics* **2021**, *11*, 1071. [CrossRef]

35. Nawaz, M.; Nazir, T.; Masood, M.; Mehmood, A.; Mahum, R.; Kadry, S.; Thinnukool, O. Analysis of Brain MRI Images Using Improved CornerNet Approach. *Diagnostics* **2021**, *11*, 1856. [CrossRef]

36. Farzaneh, N.; Williamson, C.A.; Jiang, C.; Srinivasan, A.; Bapuraj, J.R.; Gryak, J.; Najarian, K.; Soroushmehr, S. Automated segmentation and severity analysis of subdural hematoma for patients with traumatic brain injuries. *Diagnostics* **2020**, *10*, 773. [CrossRef] [PubMed]

37. Meng, L.; Zhang, Q.; Bu, S. Two-Stage Liver and Tumor Segmentation Algorithm Based on Convolutional Neural Network. *Diagnostics* **2021**, *11*, 1806. [CrossRef] [PubMed]

38. Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. *IEEE Sens. J.* **2021**, *21*, 20704–20713. [CrossRef]

39. Majid, A.; Nam, Y.; Tariq, U.; Roy, S.; Mostafa, R.R.; Sakr, R.H. COVID19 classification using CT images via ensembles of deep learning models. *Comput. Mater. Contin.* **2021**, *69*, 319–337. [CrossRef]

40. Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. *Complex Intell. Syst.* **2021**, 1–14. [CrossRef]

41. Liu, D.; Liu, Y.; Li, S.; Li, W.; Wang, L. Fusion of handcrafted and deep features for medical image classification. *J. Phys. Conf. Ser.* **2019**, *1345*, 022052. [CrossRef]

42. Alinsaif, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. *Comput. Biol. Med.* **2021**, *138*, 104879. [CrossRef]

43. Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 4267–4275. [CrossRef]

44. Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models for diagnosis of breast cancer. *Neural Comput. Appl.* **2020**, 1–12. [CrossRef]

45. Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. *Appl. Sci.* **2020**, *10*, 8298. [CrossRef]

46. Zeebaree, D.Q. A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 78–91.

47. Huang, K.; Zhang, Y.; Cheng, H.; Xing, P. Shape-adaptive convolutional operator for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

48. Sadad, T.; Hussain, A.; Munir, A.; Habib, M.; Ali Khan, S.; Hussain, S.; Yang, S.; Alawairdhi, M. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. *Appl. Sci.* **2020**, *10*, 1900. [CrossRef]

49. Mishra, A.K.; Roy, P.; Bandyopadhyay, S.; Das, S.K. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. *Expert Syst.* **2021**, *38*, e12713. [CrossRef]

50. Hussain, S.; Xi, X.; Ullah, I.; Wu, Y.; Ren, C.; Lianzheng, Z.; Tian, C.; Yin, Y. Contextual level-set method for breast tumor segmentation. *IEEE Access* **2020**, *8*, 189343–189353. [CrossRef]

51. Xiangmin, H.; Jun, W.; Weijun, Z.; Cai, C.; Shihui, Y.; Jun, S. Deep Doubly Supervised Transfer Network for Diagnosis of Breast Cancer with Imbalanced Ultrasound Imaging Modalities. *arXiv* **2020**, arXiv:2007.066342020.

52. Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [CrossRef]

53. Byra, M.; Jarosik, P.; Szubert, A.; Galperin, M.; Ojeda-Fournier, H.; Olson, L.; O'Boyle, M.; Comstock, C.; Andre, M. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomed. Signal Process. Control* **2020**, *61*, 102027. [CrossRef]

54. Kadry, S.; Damaševičius, R.; Taniar, D.; Rajinikanth, V.; Lawal, I.A. Extraction of tumour in breast MRI using joint thresholding and segmentation–A study. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.

55. Lahoura, V.; Singh, H.; Aggarwal, A.; Sharma, B.; Mohammed, M.; Damaševičius, R.; Kadry, S.; Cengiz, K. Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. *Diagnostics* **2021**, *11*, 241. [CrossRef]

56. Maqsood, S.; Damasevicius, R.; Shah, F.M. *An Efficient Approach for the Detection of Brain Tumor Using Fuzzy Logic and U-NET CNN Classification, International Conference on Computational Science and Its Applications*; Springer: Cham, Switzerland, 2021; pp. 105–118.

57. Rajinikanth, V.; Kadry, S.; Taniar, D.; Damasevicius, R.; Rauf, H.T. Breast-cancer detection using thermal images with marine-predators-algorithm selected features. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Noida, India, 26–27 August 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

58. Ouahabi, A.; Taleb-Ahmed, A. Deep learning for real-time semantic segmentation: Application in ultrasound imaging. *Pattern Recognit. Lett.* **2021**, *144*, 27–34. [CrossRef]

59. Al-Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **2020**, *28*, 104863. [CrossRef]

60. Khan, M.A.; Kadry, S.; Zhang, Y.-D.; Akram, T.; Sharif, M.; Rehman, A.; Saba, T. Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Comput. Electr. Eng.* **2021**, *90*, 106960. [CrossRef] [PubMed]

61. Khan, M.A.; Sharif, M.I.; Raza, M.; Anjum, A.; Saba, T.; Shad, S.A. Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection. *Expert Syst.* **2019**, e12497. [CrossRef]

62. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

63. Cao, Z.; Yang, G.; Chen, Q.; Chen, X.; Lv, F. Breast tumor classification through learning from noisy labeled ultrasound images. *Med. Phys.* **2020**, *47*, 1048–1057. [CrossRef] [PubMed]

64. Ilesanmi, A.E.; Chaumrattanakul, U.; Makhanov, S.S. A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. *Biocybern. Biomed. Eng.* **2021**, *41*, 802–818. [CrossRef]

65. Zhuang, Z.; Yang, Z.; Raj, A.N.J.; Wei, C.; Jin, P.; Zhuang, S. Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion. *Comput. Methods Programs Biomed.* **2021**, *208*, 106221. [CrossRef]

66. Pang, T.; Wong, J.H.D.; Ng, W.L.; Chan, C.S. Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification. *Comput. Methods Programs Biomed.* **2021**, *203*, 106018. [CrossRef]