

# A hybrid ensemble deep learning approach for reliable breast cancer detection



Mohamed Abdelmoneim Elshafey<sup>a,1,\*</sup>, Tarek Elsaid Ghoniemy<sup>a,2</sup>

<sup>a</sup> Military Technical College, Department of Computer Engineering and Artificial Intelligence, Cairo, Egypt

<sup>1</sup> [m.shafey@mtc.edu.eg](mailto:m.shafey@mtc.edu.eg); <sup>2</sup> [ghoniemy\\_t@mtc.edu.eg](mailto:ghoniemy_t@mtc.edu.eg)

\* corresponding author

## ARTICLE INFO

### Article history

Received January 18, 2021

Revised March 29, 2021

Accepted April 2, 2021

Available online April 20, 2021

### Keywords

Breast Cancer Detection

Transfer Learning

Xception

LSTM

Support Vector Machine

Precision

## ABSTRACT

Among the cancer diseases, breast cancer is considered one of the most prevalent threats requiring early detection for a higher recovery rate. Meanwhile, the manual evaluation of malignant tissue regions in histopathology images is a critical and challenging task. Nowadays, deep learning becomes a leading technology for automatic tumor feature extraction and classification as malignant or benign. This paper presents a proposed hybrid deep learning-based approach, for reliable breast cancer detection, in three consecutive stages: 1) fine-tuning the pre-trained Xception-based classification model, 2) merging the extracted features with the predictions of a two-layer stacked LSTM-based regression model, and finally, 3) applying the support vector machine, in the classification phase, to the merged features. For the three stages of the proposed approach, training and testing phases are performed on the BreakHis dataset with nine adopted different augmentation techniques to ensure generalization of the proposed approach. A comprehensive performance evaluation of the proposed approach, with diverse metrics, shows that employing the LSTM-based regression model improves accuracy and precision metrics of the fine-tuned Xception-based model by 10.65% and 11.6%, respectively. Additionally, as a classifier, implementing the support vector machine further boosts the model by 3.43% and 5.22% for both metrics, respectively. Experimental results exploit the proposed approach's efficiency with outstanding reliability in comparison with the recent state-of-the-art approaches.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## 1. Introduction

As per International Agency for Research on Cancer (IARC) [1], cancer is one of the top-ranked causes of death globally. Breast Cancer (BC), especially, is considered the second most prevalent one among women, making it crucial to conduct more studies on BC detection. The early-stage diagnosis, which the pathologists usually perform a thorough visual inspection of histopathological slides under the microscope, can significantly reduce the mortality rate. However, a successful pathological examination requires a professional background, predictive power, and rich experience.

A recent solution to tackle the abovementioned challenges is using automated methodology with intelligent diagnostic techniques, which can learn over time from previous experience. With the help of powerful computing hardware capability, the automatic algorithm can speed the manual diagnosing process and reduce the error rate. Deep Learning (DL) technology is a leading one for predicting tumors as malignant or benign from histopathology images [2]–[6]. Deep learning-based techniques involve two main models: Convolutional Neural Network (CNN) as a classification model and Long Short Term

Memory (LSTM) model as a regressive one. The former can exploit spatial correlation in data images, while the latter can make predictions in these data sequences. For BC researches, a publicly available benchmark dataset [7], with a new histopathological database of microscopic breast tumor images (BreakHis) is introduced. It is widely employed in evaluating state-of-the-art BC detection approaches [8]–[12].

The BC image recognition schemes can typically be classified into two main categories based on feature extraction methods: hand-crafted extraction and automatic extraction [13], [14]. Different research works have been published concerning cancer detection using machine learning techniques [15], [16]. However, such methods' applications are limited due to manual feature extraction that can be considered a critical step of BC detection. Traditionally, SIFT [17] and SURF [8] hand-crafted feature descriptors were being utilized to feature extraction until the advent of DL techniques that can extract more discriminative information from data with no need to design feature extractors by human experts.

In the second category of feature extraction, the DL techniques offer an automated, accurate, and reliable methodology for learning features from medical images in a way that avoids the constraints of such hand-crafted features [18]. CNN's, as a type of deep forward networking, have achieved empirical success in automatic diagnosis and analysis of the BC in histopathological images [3]–[6] to classify images into one of two classes benign (tumor-free) or malignant (tumored). Learning DL models from scratch in large data sets [19] is a tedious task due to computational complexity and convergence problems [20]. Furthermore, in the case of an insufficient amount of high-quality labeled samples, as in most common medical BC datasets [7], one can benefit from applying Transfer Learning (TL) [21], [22] to one of the top-ranked pre-trained models for faster convergence and outperforming training from scratch [20]–[23].

In [8], authors showed that using fine-tuning, the pre-trained VGG16, VGG19, and ResNet50 models achieved improved accuracy but with only 92.0% precision a reliability indicator. In [9], the two-step TL-based approach is proposed for feature extraction from histopathological images using Inception-v3 and SVM classifier that improved the classification accuracy by 3.7%. The use of multiple instances learning for histopathological BC detection is investigated in [10]. However, the presented average accuracy is only 88%. A TL-based model is proposed and trained, in [11], on stain normalized and augmented BreakHis dataset. Based on accuracy and precision metrics, the observed results are 81.25% and 91.79%, respectively. A TL on the pre-trained Xception model, in [12], is applied. However, an important evaluation metric such as precision to study the proposed approach's reliability is not presented. The combination of the pre-trained CNN activation features on SVMs has been investigated [24], while another combination of CNN and LSTM in [25] achieves an average precision over the four categories of the BreakHis dataset of only 90.25%. In [26], a compact CNN approach achieves accuracy and precision of 87.40% and 88.08%, respectively. In [27], a multi-layer feature fusion for BC image classification is proposed, in which the independence and partial dependence of all sub-layers are considered. A deep convolution generative adversarial network, in [28], is proposed to balance the BC data set class distribution by the augmentation of only minor classes to avoid the classifier bias toward the majority class. In [29], a proposed ensemble deep learning approach achieved 95.3% of accuracy with a lower precision value of 93.5%.

In this paper, a hybrid approach of the TL-based classification model and regression model, for more tuned and robust feature extraction, is suggested to comb with SVM classifier for highly accurate and reliable BC detection. This work investigates breast cancer detection using a combined Xception-based classification approach and LSTM-based regression one for highly tuned extracted features that feed a robust Support Virtual Machine (SVM) classifier. Combining both classification and regression approaches leads to a highly reliable efficiency of the proposed approach in accuracy, precision, and different false rates. Section 2 presents the overall methodology of the proposed approach in a clarified sequence of stages, while Section 3 discusses the experimental work in a detailed analysis of results against those of the recent competing state-of-the-art approaches. Then, Section 4 concludes the work by highlighting the main results followed by possible future works.

## 2. Method

### 2.1. Stage of implementing a convolutional-based classification

In the proposed approach, the pre-trained Xception-based model is applied in a fine-tuned manner. Xception model is a deep-CNN model in the form of a linear stack of depthwise separable convolution layers, with residual connections in a modified version. A modified depthwise separable convolution consists of  $1 \times 1$  pointwise convolution that maps cross-channel correlations, followed by  $n \times n$  depthwise convolution for separately mapping every channel's spatial correlations. Depthwise separable convolution provides greatly reduced parameter count, more efficient complexity, maintains cross-channel features. For  $n \times n$  convolutional layer on  $k$  input channels and  $m$  output channels, regular convolution generates  $(k \times n \times n \times m)$  parameters, but with depthwise separable convolution, count of (depthwise Conv. + spatial Conv.) =  $(k \times 1 \times 1 \times m + n \times n \times m)$ , parameters are generated as illustrated in Fig. 1.

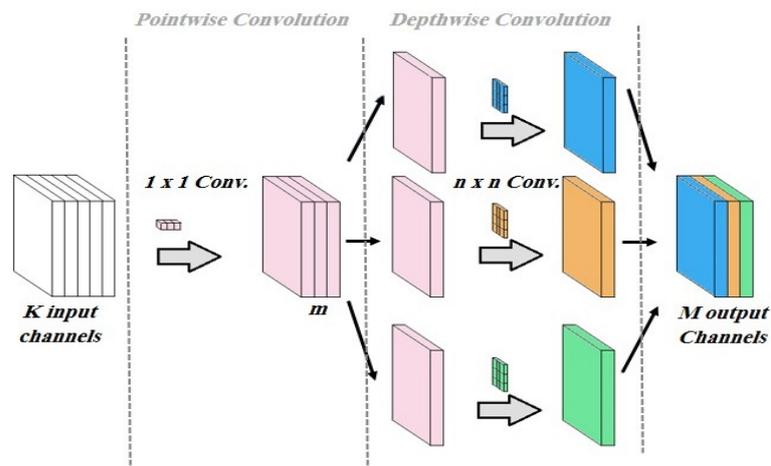


Fig. 1. Xception-based model structure.

Xception architecture has outperformed VGG16, ResNet, and Inception-V3 in most classical classification challenges [30]–[32]. Xception model comprises 36 convolutional layers forming the feature extraction base of the network structured into 14 modules, all of which have linear residual connections around them, except for the first and last modules. It is previously trained on a 1000-class single-label classification task on the ImageNet dataset [19] of more than 14 million images.

### 2.2. Applying Transfer Learning

In this step, the TL is applied to the Xception-based model for the BC detection task. Recent implementations of DL-based models, as in Fig. 2(a), adopt one of two different main methods: the first method is by learning the model from scratch on the large dataset for achieving better accuracy, while the second method incorporates TL, in which the parameters of a pre-trained model for a specific task with high accuracy, are used to initialize the new model with the necessary modification towards a required task. TL is mainly useful for tasks where enough training samples are not available to train a model from scratch, such as medical image classification [21]–[33] as in Fig. 2(b).

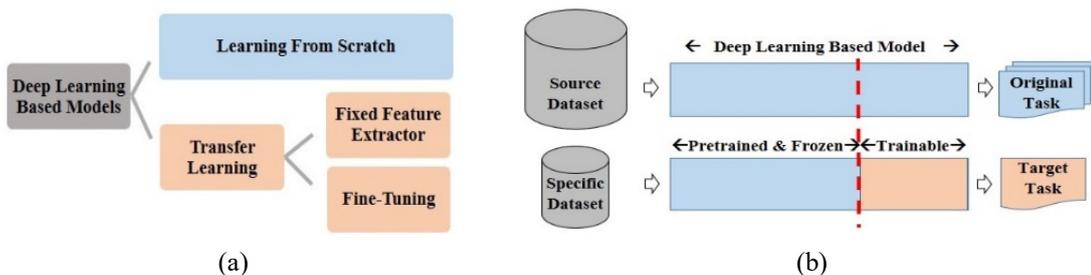


Fig. 2. Describing (a) implementation of deep learning models, (b) transfer learning-based DL approach.

Generally, the low levels of DL models provide generic features, while the higher achieved the specific features. The learned features are related to the task of the pre-trained model. Therefore, there are two main factors in transfer learning, upon which the pre-trained model can be used towards a new task. These two main factors are 1) the size of the targeted dataset and 2) the similarity of the new task to that of the pre-trained model. These considered factors lead to four different cases, as shown in Fig. 3.

- **Case1** is for a small data set and similar task, in which the high-level features, i.e., from top layers, are specific for the same and can be used. Hence, the original model is applied as a feature extractor with no modification, and just the classifier on top of it can be retrained.
- **Case2** is for small data set and different tasks for which high-level features cannot be used. Hence, the original pre-trained model can be applied as a feature extractor but should be retrained from a low level to the end of the model, i.e., fully connected layers that provide more generic features than those from a higher layer. From the start of the pre-trained model to a selected lower level, it is kept frozen.
- **Case3** is for large data set and similar task, such that the pre-trained weights, from a low level of the model, should be fine-tuned. The pre-trained model should be relearned, starting from a lower level, i.e., high-level convolutional layers, for some new learned features.
- **Case4** is for large data set and different tasks, allowing the whole base model to be fine-tuned and relearned with that amount of data.

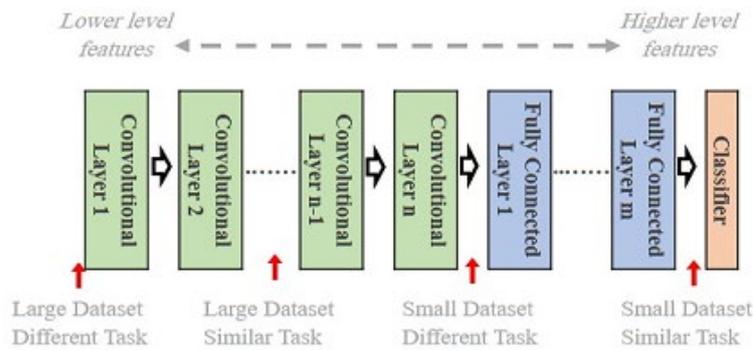


Fig. 3. Implementation of different cases of TL.

In the first stage of the proposed approach, the TL technique, as in Case 4, is applied to the pre-trained Xception model for the BC detection task. The step of applying TL incorporates the pre-trained Xception-based model with three randomly initialized Fully Connected (FC) layers of dimensions  $D_{FC1}$ ,  $D_{FC2}$ , and  $D_{FC3}$ , respectively, an LR layer in the form of a two-node dense layer, and a binary cross-entropy activation function, as shown in Fig. 4.

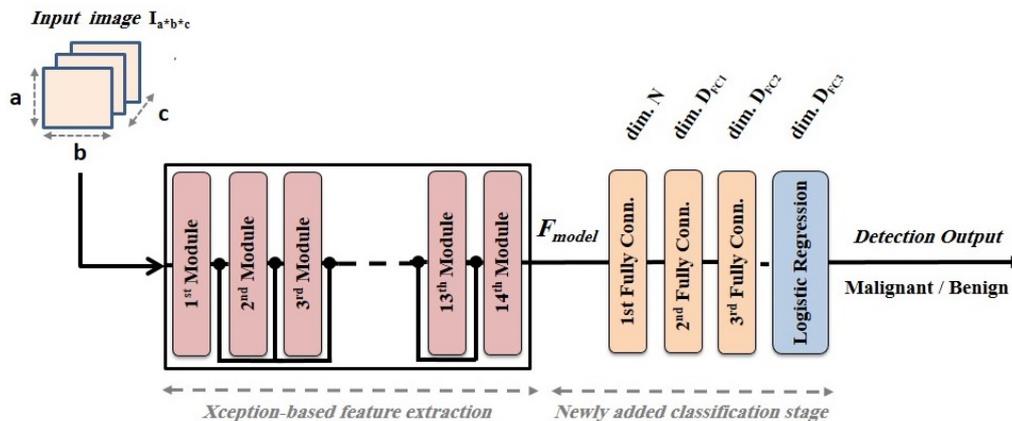


Fig. 4. Applying TL for Xception-based BC detection.

### 2.3. Stage of merging Xception-based classification with LSTM-based regression

In the second stage of the proposed approach, a Recurrent Neural Network (RNN) model, as the regressive branch, is suggested to provide predictions for sequences of data in images that can be applied as multiplicative values for the features extracted from the fine-tuned Xception-based model for more enhanced extracted features. In the regressive branch, the LSTM network, as a special type of RNN, is implemented to learn long-term dependencies and overcome the previously vanishing and exploding gradients of typical RNN [34].

LSTM model consists of multiple looped networks. Each network, in the loop, takes input information from the preceding network and produces output besides passing the information to the next network. The repeating module of the LSTM is the memory cell that consists of various gates: an input gate for controlling the amount of previous information to pass, forget gate for selecting allowable values to be updated, and an output gate for deciding information carried by the hidden state [34], [35]. A stacked version of the LSTM architecture is such an LSTM model with multiple LSTM layers to enhance the prediction efficiency, making the model deeper. For the stacked LSTM-based approach, two layers are recommended to avoid the degradation problem, in which the model becomes more difficult to train; hence the prediction accuracy will be saturated [35], [36].

The second stage of the proposed approach combines the Xception-based model with a two-layer stacked LSTM-based model in two separate branches, as shown in Fig. 5. The input image  $I_{a*b*c}$  to the Xception-based branch is of dimension  $a*b*c$ , representing the width, height, and depth of image  $I$ , respectively. For the LSTM-based branch, the input image  $I_{a*b*c}$  is transformed into a grayscale version  $I_{a*b}^g$  and fed in the form of time  $t$  series-based chunks each of size  $m$  such that  $t*m = a*b$ . At any time step  $i = 1 \dots t$ , the input to the LSTM-based branch is pixels  $I_{a*b}^g(1+(i-1)*m : m+(i-1)*m)$ . The output feature vector from the Xception-based branch  $f_{Xception}^N$  and predictions from the LSTM-based branch  $f_{LSTM}^N$  are of the same size  $N$ . The merging layer merges  $f_{Xception}^N$  and  $f_{LSTM}^N$  using element-wise multiplication to form the resultant extracted features according to (1).

$$F_{model} = \{f_{LSTM}^1 * f_{Xception}^1, \dots, f_{LSTM}^N * f_{Xception}^N\} \tag{1}$$

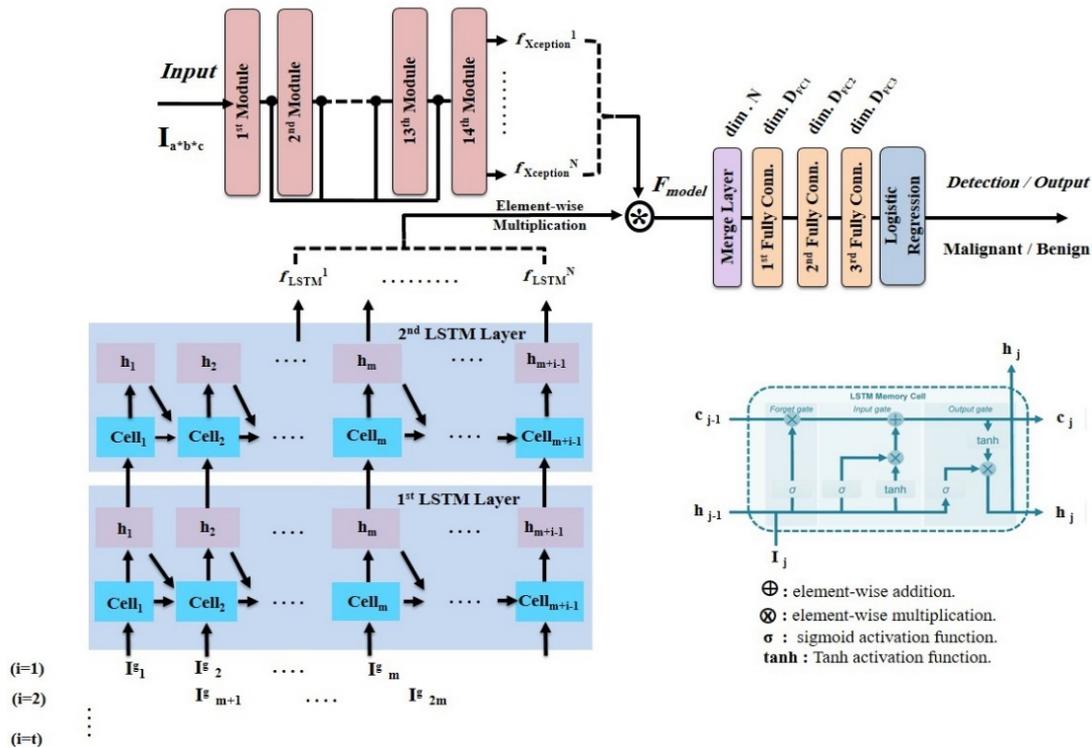


Fig. 5. Merging Xception-based model and two-layer stacked LSTM-based model for BC detection.

The model feature vector  $F_{model}$  is then applied to three FC layers and the LR layer that are mentioned in the first stage. Training of the second stage of the proposed approach passes two sequential steps. In the first training step, the fine-tuned Xception-based branch, from the first stage, is frozen and the LSTM-based branch combined with the three randomly initialized FC layers are trained from scratch, then in the second training step, the whole system, including the two branches of Xception model and LSTM model, are fine-tuned.

#### 2.4. Stage of implementing SVM classifier for the merged Xception and LSTM features

In the third stage, i.e., the final one of the proposed approach, the three FC layers and the LR layer in the second stage are replaced by an SVM classifier. SVM becomes a powerful machine learning tool for binary as well as multi-class labeling scenarios [37]. It is extensively used in computer vision applications especially, medical ones [38]–[40].

Given a sample test image as a feature vector  $F_{model}$  of dimension  $N$  that is closest to the hyperplane  $H$ , it forms an orthogonal vector  $d$  that stems from it in the same direction as  $w$ . Any point  $X^0 \in H$  (corresponding  $y^0 = 0$ ) will form a vector  $r$  with  $F_{model}$  in which  $d$  is the projection of  $r$  on  $w$  given by  $d = 1/||W||$ . Therefore, one can easily find the optimal margin by maximizing  $1/||W^2||$ , S.T.  $\min_n |(W^T F_{model} + b)| = 1$ . However, due to the presence of the reciprocal terms, it is more convenient to turn the problem into minimization as  $d = \min ||W^2||/2$  in order to avoid the derivation issue. So, it is a quadratic programming problem to classify the label  $L$  of  $F_{model}$  as  $L(w, b, \alpha) = 1/2 * W^T W - \sum_{n=1}^K \alpha^i [y^i (W^T F_{model} + b) - 1]$ , where  $K$  is the number of support vectors,  $\alpha^i$  is the Lagrange multipliers that are always positive for those points that represent support vectors since they influence the behavior of  $H$ . By calculation of the derivative of  $L$  w.r.t  $w$  and  $b$  then substituting with  $\sum_{n=1}^K \alpha^i y^i F_{model} = 0$  and  $\sum_{n=1}^K \alpha^i y^i = 0$  in leads the next equation, in which  $\mathcal{K}$  is a kernel function that is used to express the product of  $F_{model}^m$  and  $F_{model}^i$  inputs.

$$L(\alpha) = \sum_{n=1}^K \alpha^i - \frac{1}{2} \sum_{m=1}^K \sum_{n=1}^K \alpha^m \alpha^i y^m y^i \mathcal{K}(F_{model}^m F_{model}^i) \quad (2)$$

Kernel functions allow the transformation from non-linearly separable spaces to linearly-separable ones and are considered a useful tool for solving diverse classification tasks [41]. For each feature map  $F_{model}$ , a slack variable  $\varepsilon$  is defined, which is zero for points on the margin and increases as going further from the correct boundary, till the point on the wrong side, in which  $\varepsilon$  is expected to be greater than the value 1. By substitution of  $\varepsilon$  into  $L$ , it is needed to minimize  $(||W^T W||/2 + c \sum_{n=1}^K \varepsilon^i)$ .

In the final stage of the proposed approach, the three FC layers and the LR layer, as mentioned in the second stage shown in Fig. 5, are replaced by the SVM classifier as shown in Fig. 6.

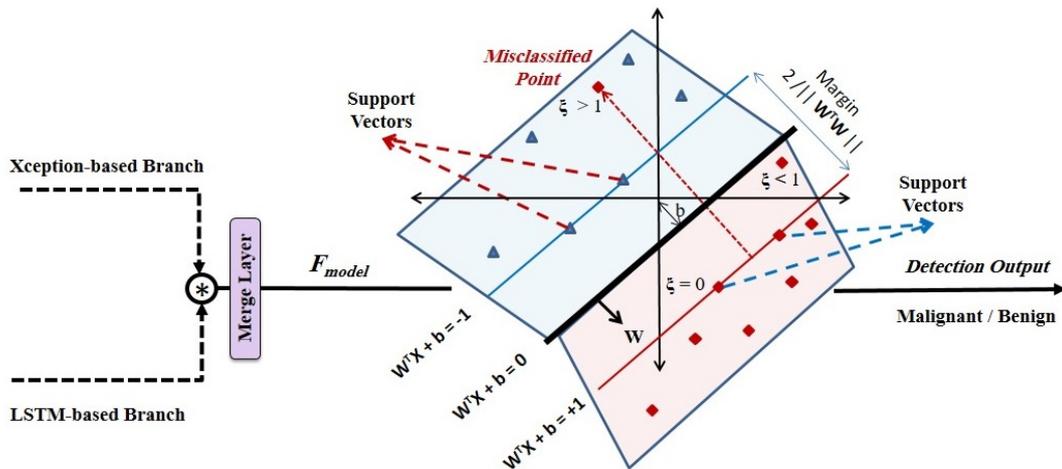


Fig. 6. Combining the merged Xception-based classifier and two-layer stacked LSTM-based model with SVM classifier for BC detection.

### 3. Results and Discussion

#### 3.1. Experimental Work Setup

The system configuration used for learning and testing the proposed approach, during its three stages, has the following specifications: Intel CPU core I7-10700T 10<sup>th</sup> generation (35 MB L3 Cache, 2.6 GHz), RAM: 16 GB RAM DDR4 2133 MHz, GPU: ASUS GeForce GTX 1650, 1733 MHz, 2560 CUDA Cores, 6GB GDDR5 under Windows 64-bit operating system. Python 3.6.3, with GPU-enabled Tensor Flow version 1.7.0, CUDA 9.1 Toolkit, and cuDNN v7.1.2, are used.

#### 3.2. Dataset Description

In experiments, the BreakHis dataset [7] is used to validate the proposed approach during its three stages. It contains a total of 9109 sample images, each categorized as either benign (2450 samples) or malignant (5429 samples). The samples were collected from 82 patients with different magnification factors (40x, 100x, 200x, 400x) in an RGB format with a resolution of 700\*460\*3.

#### 3.3. Pre-processing and Data Augmentation

In the pre-processing phase, all histopathological images in the dataset are normalized to reduce the color variation that enhances the color consistency. Training a DL model on the larger dataset is the best way to generalize it and to minimize overfitting probability in the obtained results. Besides that, to avoid the degradation that may affect the state-of-the-art deep predictive models due to the data scarcity problem, data augmentation is recently used to artificially expand the labeled training dataset, which is essential for combating such data scarcity problem [42][43]. For these considerations, nine data augmentation techniques 1) horizontal shift, 2) vertical shift, 3) horizontal-vertical shift, 4) horizontal flip, 5) vertical flip, 6) random rotation, 7) random brightness, 8) random zoom, and 9) Gaussian noise, are applied to the dataset before training that enlarges the dataset to 10-times the original size and in turn improves the model generalization.

#### 3.4. Performance Evaluation Metrics

In experiments, a list of different metrics [24] that targets the proposed approach's accuracy and reliability is considered. The targeted metrics in (3)-(8) rely on the actual values predicted by the proposed model. True Positive (TP) is the count at which the actual value was positive, and the model predicted a positive. True Negative (TN) is the number of times the actual value was negative, and the model predicted as a negative value. False Positive (FP) is the count of the actual negative values predicted as positive values. False Negative (FN) is the opposite to FP, i.e., count of actual positive values predicted as negative values.

- Accuracy, the higher is, the better, measures how often the classification model predicts the output correctly.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

- Precision, the higher is, the better, measures the model's performance in terms of positive example classification, i.e., percentage of the Positively Predicted Values (PPV) that were truly positive. It indicates the model reliability in cases where FP is a higher concern than FN (4).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- The Negative Predicted Values (NPV), the higher is, the better, can be defined as (5).

$$NPV = \frac{TN}{TN + FN} \quad (5)$$

- The False Discovery Rate (FDR), the lower is better. It can be defined as (6).

$$FDR = \frac{FP}{TP + FP} \quad (6)$$

- The False Positive Rate (FPR) and False Negative Rate (FNR), the lower values of each are the better, can be calculated as:

$$FPR = \frac{FP}{FP+TN} \quad (7)$$

$$FNR = \frac{FN}{FN+TP} \quad (8)$$

Besides the considered metrics, the Receiver Operating Characteristic (ROC) graph is extracted for the proposed approach's three stages. ROC curve shows the True Positive Rate (TPR) as a False Positive Rate (FPR) function. Additionally, The Area Under the Curve (AUC) of the ROC curve, the higher and better, is the classification model's capability to discriminate between different classes. While the ROC is a two-dimensional representation of the model's performance, the AUC provides this information in a single scalar representation form. It is a commonly used evaluation method for binary classification tasks as it provides a better assessment of the model's ability to discriminate between the two classes.

### 3.5. Model Setting and Training Hyper-parameters

In experiments of all the three stages towards the proposed model, the trainable three FC layers have dimensions:  $D_{FC1} = 2048$ ,  $D_{FC2} = 512$  and  $D_{FC3} = 128$ . For the LSTM-based branch, as in Fig. 5 and Fig. 6, each chunk is of size  $m = 3500$ , and the time series  $t = 92$  steps. The output of both the Xception-branch and the LSTM-based branch is of size  $N = 2048$ , which also corresponds to the merge layer's dimension. In the third stage, the SVM module is supported by linear kernel function and hyperplane parameter  $C = 10$ .

The training dataset, i.e., the BreakHis, is partitioned into 70%, 10%, and 20% for training, validation, and testing phases, respectively. The suggested nine augmentation techniques are applied to only 80% of the dataset associated for the training and validation phases in the proposed approach. The rest 20% of the dataset is left without augmentation for testing to achieve high reliability with the obtained results. K-fold cross-validation is adopted to the training phase of the dataset where  $K = 10$  to get a less biased model, avoid the overfitting, and lead to better generalization of the predictive model [44]. In the training phase, the learning rate = 0.0001 with 0.3 dropout rate and Adam optimizer. Batch size = 64 and associated number of epochs = 100.

### 3.6. Comparative Results and Analysis

Classification results of the three consecutive stages of the proposed approach are presented to entirely evaluate each stage's contributions. The proposed approach's performance is further analyzed by the associated ROC curves, as shown in Fig. 7. The average AUC values for the first, second, and third stages of the proposed approach are 0.84, 0.92, and 0.95, respectively. The resulted AUC values show an enhancement of 9.52% for implementing the LSTM-based model in the second stage, and an additional one of 3% by applying the SVM classifier at the final stage.

The obtained performance metrics of the three stages: S1, S2, and S3, respectively, of the proposed approach, are shown in Table 1, including accuracy, precision, FPR, FDR, and FNR, associated with the four subcategories of the BreakHis dataset: 40x, 100x, 200x, and 400x. Experimental results in Table 1 show significant enhancements of the second stage, in all evaluation metrics for all subcategories of the dataset, and additional enhancements provided by the third stage, in almost all cases except some cases, in which the second stage outperforms the third one with small values in comparison with those between the second and the first stage.

Moreover, as shown in Fig. 8, the overall incremental enhancements of the second stage over the first one are 10.65%, 11.6%, 48.26%, 49.38%, and 45.04% for metrics: accuracy, precision, FDR, FPR, and FNR, relatively, while the corresponding enhancements of the third stage over the second stage are: 3.43%, 5.22%, 46.88%, 48.78%, and 13.89%, respectively. These significant enhancements in FDR, FPR, and FNR metrics demonstrate the proposed approach's high prediction reliability.

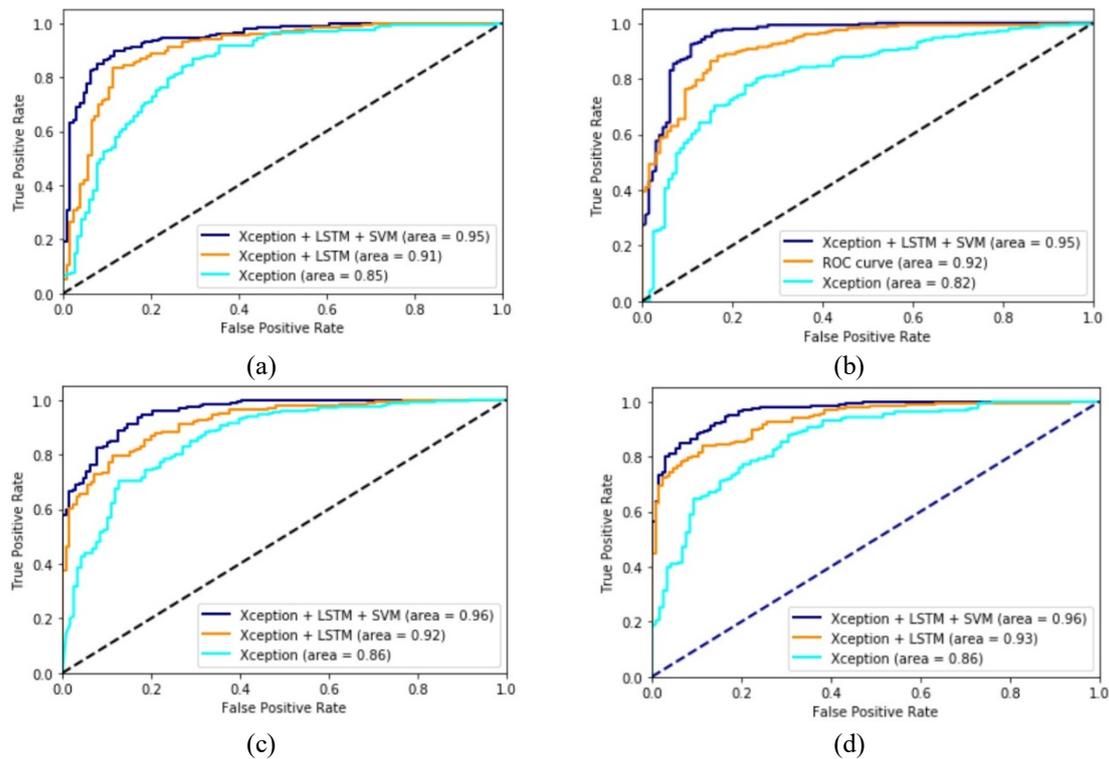


Fig. 7. The ROC curves and the AUCs values of the three stages for magnification level of (a) 40x, (b) 100x, (c) 200x, and (d) 400x.

Table 1. Performance metrics in (%) of the three stages: S1, S2, and S3, of the proposed approach on the dataset subcategories: 40x, 100x, 200x, and 400x.

Performance Metrics	40x			100x			200x			400x		
	S1	S2	S3									
Accuracy	83.3	90.3	93.3	81.3	89.3	92.5	83.3	92.3	94.8	79	89.9	93.5
Precision	84.5	91.3	94.4	79.9	88.7	95.2	81.5	93.3	92.8	67.6	86.6	96.3
FDR	15.5	8.7	5.6	20.1	11.3	4.8	18.5	6.7	7.2	23.4	13.4	3.7
FPR	15	8.5	5.5	21	11.5	4.5	19.5	6.5	7.5	25.5	14.5	3.5
FNR	18.5	11	8	16.5	10	10.5	14	9	3	16.5	6	9.5

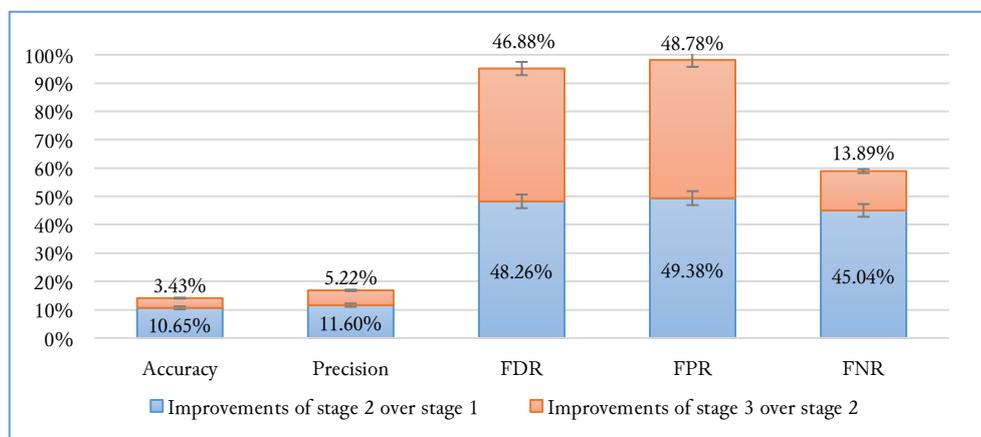


Fig. 8. Incremental enhancement ratios (%) of the second and the third stages, over the first one, of the proposed approach.

For evaluating the proposed approach's effectiveness, a comparative analysis with the results of recent state-of-the-art related approaches, using the same BreakHis dataset, is presented in Table 2. Among various metrics [45] that can be considered for evaluating classification models, the accuracy metric is the most frequently used in the related state-of-the-art approaches. A related point to consider is that even the model is of high accuracy. It may not predict the actual cancer patients reliably, leading to severe consequences, especially if there is a significant disparity between the number of positive and negative labels in BreakHis.

According to Table 2, the results show significant improvements of the proposed approach and outstanding reliability from a precision perspective against the competing state-of-the-art ones, evaluated on the same benchmark dataset, i.e., the BreakHis. It is worth mentioning that some of these approaches present their results only from the accuracy perspective, which can not exploit the model reliability. Hence, approaches as in [8]–[10], [12], and [27][28] cannot be verified as a reliable BC detection system.

**Table 2.** Comparative results in (%) of performance metrics between the proposed approach and recent state-of-the-art approaches.

Performance metrics	Competing approaches										
	[8]	[9]	[10]	[11]	[12]	[25]	[26]	[27]	[28]	[29]	<i>Proposed</i>
Accuracy	92.60	89.00	88.00	81.25	92.50	89.00	87.40	94.66	94.7	95.3	<b>94.00</b>
Precision	-	-	-	91.79	-	90.25	88.08	-	-	93.5	<b>95.00</b>

#### 4. Conclusion

This paper presents a hybrid ensemble deep learning approach for reliable breast cancer detection. The presented approach combines the pre-trained Xception model and two-layer stacked LSTM model for enhanced extracted features, upon which the SVM classifier employs breast cancer detection. BreakHis dataset is implemented in training and testing phases, with an additional nine applied different data augmentation techniques to boost the performance and the reliability of the proposed approach. Experimental results demonstrate that incorporating the regression-based LSTM branch improves the fine-tuned Xception-based model, especially in accuracy, precision, FDR, FPR, and FNR by ratios 10.65%, 11.6%, 48.26%, 49.38%, and 45.04%, respectively. An additional improvement of 3.43%, 5.22%, 46.88%, 48.78%, and 13.89% for the same metrics are provided by applying the SVM classifier on the merged extracted features. Comparative results, among the proposed approach and a recent list of state-of-the-art approaches, show a significant outperforming of the proposed approach by values of 94% and 95% for both accuracy and precision metrics, respectively, which proves its high reliable efficiency in BC detection. As future work, the proposed approach can be implemented in detecting different cancer types in histopathology images and even in the detection of COVID-19 in X-Rays. Moreover, sequence-based DL models as attention-based ones, can be implemented in a new hybrid DL model and compared with the proposed one.

#### Acknowledgment

The authors thank the anonymous reviewers for their valuable support and effort to improve the manuscript quality greatly. The authors also thank the Department of Computer Engineering and Artificial Intelligence at the Military Technical College (Cairo, Egypt) for their appreciated encouragement.

#### Declarations

**Author contribution.** All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

**Funding statement.** None of the authors have received any funding or grants from any institution or funding body for the research.

**Conflict of interest.** The authors declare no conflict of interest.

**Additional information.** No additional information is available for this paper.

### References

- [1] L. A. Torre, R. L. Siegel, E. M. Ward, and A. Jemal, "Global Cancer Incidence and Mortality Rates and Trends—An Update," *Cancer Epidemiol. Biomarkers Prev.*, vol. 25, no. 1, pp. 16–27, Jan. 2016, doi: [10.1158/1055-9965.EPI-15-0578](https://doi.org/10.1158/1055-9965.EPI-15-0578).
- [2] K. G. Kim, "Book Review: Deep Learning," *Healthc. Inform. Res.*, vol. 22, no. 4, p. 351, 2016, doi: [10.4258/hir.2016.22.4.351](https://doi.org/10.4258/hir.2016.22.4.351).
- [3] T. G. Debelee, F. Schwenker, A. Ibenthal, and D. Yohannes, "Survey of deep learning in breast cancer image analysis," *Evol. Syst.*, vol. 11, no. 1, pp. 143–163, Mar. 2020, doi: [10.1007/s12530-019-09297-2](https://doi.org/10.1007/s12530-019-09297-2).
- [4] T. G. Debelee, S. R. Kebede, F. Schwenker, and Z. M. Shewarega, "Deep Learning in Selected Cancers' Image Analysis—A Survey," *J. Imaging*, vol. 6, no. 11, p. 121, Nov. 2020, doi: [10.3390/jimaging6110121](https://doi.org/10.3390/jimaging6110121).
- [5] Y. Xue, S. Chen, J. Qin, Y. Liu, B. Huang, and H. Chen, "Application of Deep Learning in Automated Analysis of Molecular Images in Cancer: A Survey," *Contrast Media Mol. Imaging*, vol. 2017, pp. 1–10, 2017, doi: [10.1155/2017/9512370](https://doi.org/10.1155/2017/9512370).
- [6] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," *J. Healthc. Eng.*, vol. 2019, pp. 1–11, Nov. 2019, doi: [10.1155/2019/4253641](https://doi.org/10.1155/2019/4253641).
- [7] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2560–2567, doi: [10.1109/IJCNN.2016.7727519](https://doi.org/10.1109/IJCNN.2016.7727519).
- [8] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, Dec. 2018, doi: [10.1016/j.ict.2018.10.007](https://doi.org/10.1016/j.ict.2018.10.007).
- [9] J. de Matos, A. de S. Britto, L. E. S. Oliveira, and A. L. Koerich, "Double Transfer Learning for Breast Cancer Histopathologic Image Classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8, doi: [10.1109/IJCNN.2019.8852092](https://doi.org/10.1109/IJCNN.2019.8852092).
- [10] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, "Multiple instance learning for histopathological breast cancer image classification," *Expert Syst. Appl.*, vol. 117, pp. 103–111, Mar. 2019, doi: [10.1016/j.eswa.2018.09.049](https://doi.org/10.1016/j.eswa.2018.09.049).
- [11] G. Murtaza *et al.*, "Breast cancer classification using digital biopsy histopathology images through transfer learning," *J. Phys. Conf. Ser.*, vol. 1339, p. 012035, Dec. 2019, doi: [10.1088/1742-6596/1339/1/012035](https://doi.org/10.1088/1742-6596/1339/1/012035).
- [12] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Breast Cancer Diagnosis with Transfer Learning and Global Pooling," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 519–524, doi: [10.1109/ICTC46691.2019.8939878](https://doi.org/10.1109/ICTC46691.2019.8939878).
- [13] Y. J. Suh, J. Jung, and B.-J. Cho, "Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning," *J. Pers. Med.*, vol. 10, no. 4, p. 211, Nov. 2020, doi: [10.3390/jpm10040211](https://doi.org/10.3390/jpm10040211).
- [14] M. U. Dalmiş, S. Vreemann, T. Kooi, R. M. Mann, N. Karssemeijer, and A. Gubern-Mérida, "Fully automated detection of breast cancer in screening MRI using convolutional neural networks," *J. Med. Imaging*, vol. 5, no. 01, p. 1, Jan. 2018, doi: [10.1117/1.JMI.5.1.014502](https://doi.org/10.1117/1.JMI.5.1.014502).
- [15] Z. M. Zain *et al.*, "Predicting breast cancer recurrence using principal component analysis as feature extraction: an unbiased comparative analysis," *Int. J. Adv. Intell. Informatics*, vol. 6, no. 3, p. 313, Nov. 2020, doi: [10.26555/ijain.v6i3.462](https://doi.org/10.26555/ijain.v6i3.462).
- [16] K.-H. Yu *et al.*, "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features," *Nat. Commun.*, vol. 7, no. 1, p. 12474, Nov. 2016, doi: [10.1038/ncomms12474](https://doi.org/10.1038/ncomms12474).
- [17] G. Csurka, C. R. Dance, and M. Humenberger, "From hand-crafted to deep local features," *arXiv Prepr. arXiv1807.10254*, 2018. Available at: [Google Scholar](https://scholar.google.com/).

- [18] R. Zemouri, N. Zerhouni, and D. Racocanu, "Deep Learning in the Biomedical Applications: Recent and Future Status," *Appl. Sci.*, vol. 9, no. 8, p. 1526, Apr. 2019, doi: [10.3390/app9081526](https://doi.org/10.3390/app9081526).
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [20] M. Kohler and S. Langer, "On the rate of convergence of fully connected very deep neural network regression estimates," *arXiv Prepr. arXiv1908.11133*, 2019. Available at: [Google Scholar](https://scholar.google.com/).
- [21] L. Gui, R. Xu, Q. Lu, J. Du, and Y. Zhou, "Negative transfer detection in transductive transfer learning," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 2, pp. 185–197, Feb. 2018, doi: [10.1007/s13042-016-0634-8](https://doi.org/10.1007/s13042-016-0634-8).
- [22] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: [10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555).
- [23] T. Han, C. Liu, W. Yang, and D. Jiang, "Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions," *ISA Trans.*, vol. 93, pp. 341–353, Oct. 2019, doi: [10.1016/j.isatra.2019.03.017](https://doi.org/10.1016/j.isatra.2019.03.017).
- [24] K. Gupta and N. Chawla, "Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN," *Procedia Comput. Sci.*, vol. 167, pp. 878–889, 2020, doi: [10.1016/j.procs.2020.03.427](https://doi.org/10.1016/j.procs.2020.03.427).
- [25] A.-A. Nahid, M. A. Mehrabi, and Y. Kong, "Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering," *Biomed Res. Int.*, vol. 2018, pp. 1–20, 2018, doi: [10.1155/2018/2362108](https://doi.org/10.1155/2018/2362108).
- [26] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 198, Dec. 2019, doi: [10.1186/s12911-019-0913-x](https://doi.org/10.1186/s12911-019-0913-x).
- [27] V. Gupta and A. Bhavsar, "Partially-Independent Framework for Breast Cancer Histopathological Image Classification," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1123–1130, doi: [10.1109/CVPRW.2019.00146](https://doi.org/10.1109/CVPRW.2019.00146).
- [28] M. Saini and S. Susan, "Deep transfer with minority data augmentation for imbalanced breast cancer dataset," *Appl. Soft Comput.*, vol. 97, p. 106759, Dec. 2020, doi: [10.1016/j.asoc.2020.106759](https://doi.org/10.1016/j.asoc.2020.106759).
- [29] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. María Vanegas, "Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models," *Sensors*, vol. 20, no. 16, p. 4373, Aug. 2020, doi: [10.3390/s20164373](https://doi.org/10.3390/s20164373).
- [30] C. Wang *et al.*, "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," *IEEE Access*, vol. 7, pp. 146533–146541, 2019, doi: [10.1109/ACCESS.2019.2946000](https://doi.org/10.1109/ACCESS.2019.2946000).
- [31] M. Habibzadeh Motlagh, M. Jannesari, Z. Rezaei, M. Totonchi, and H. Baharvand, "Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception," in *Tenth International Conference on Machine Vision (ICMV 2017)*, 2018, p. 105, doi: [10.1117/12.2311282](https://doi.org/10.1117/12.2311282).
- [32] V. Alex, M. Khened, S. Ayyachamy, and G. Krishnamurthi, "Medical image retrieval using Resnet-18 for clinical diagnosis," in *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, 2019, p. 35, doi: [10.1117/12.2515588](https://doi.org/10.1117/12.2515588).
- [33] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," 2018, pp. 270–279. Available at: [Google Scholar](https://scholar.google.com/).
- [34] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, Mar. 2020, doi: [10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306).
- [35] T. A. Mahmoud, A. F. Shehab, and M. A. Elshafey, "Different Long Short-Term Memory Approaches to Enhance Prediction-Based Satellite Telemetry Compression," *J. Aerosp. Inf. Syst.*, vol. 18, no. 2, pp. 50–57, Feb. 2021, doi: [10.2514/1.1010906](https://doi.org/10.2514/1.1010906).
- [36] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM--a tutorial into Long Short-Term Memory Recurrent Neural Networks," *arXiv Prepr. arXiv1909.09586*, 2019. Available at: [Google Scholar](https://scholar.google.com/).

- [37] H. Fabelo *et al.*, "Surgical aid visualization system for glioblastoma tumor identification based on deep learning and in-vivo hyperspectral images of human patients," in *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, 2019, p. 35, doi: [10.1117/12.2512569](https://doi.org/10.1117/12.2512569).
- [38] T. Nadira and Z. Rustam, "Classification of cancer data using support vector machines with features selection method based on global artificial bee colony," 2018, p. 020205, doi: [10.1063/1.5064202](https://doi.org/10.1063/1.5064202).
- [39] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement*, vol. 146, pp. 800–805, Nov. 2019, doi: [10.1016/j.measurement.2019.05.083](https://doi.org/10.1016/j.measurement.2019.05.083).
- [40] I. Vidić *et al.*, "Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: Preliminary study," *J. Magn. Reson. Imaging*, vol. 47, no. 5, pp. 1205–1216, May 2018, doi: [10.1002/jmri.25873](https://doi.org/10.1002/jmri.25873).
- [41] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: [10.1016/j.neucom.2019.10.118](https://doi.org/10.1016/j.neucom.2019.10.118).
- [42] Y. Tang, S. Oh, J. Xiao, R. M. Summers, and Y. Tang, "CT-realistic data augmentation using generative adversarial network for robust lymph node segmentation," in *Medical Imaging 2019: Computer-Aided Diagnosis*, 2019, p. 139, doi: [10.1117/12.2512004](https://doi.org/10.1117/12.2512004).
- [43] M. Moradi, A. Madani, A. Karargyris, and T. F. Syeda-Mahmood, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," in *Medical Imaging 2018: Image Processing*, 2018, p. 57, doi: [10.1117/12.2293971](https://doi.org/10.1117/12.2293971).
- [44] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC Bioinformatics*, vol. 20, no. S11, p. 281, Jun. 2019, doi: [10.1186/s12859-019-2823-4](https://doi.org/10.1186/s12859-019-2823-4).
- [45] I. Silva and J. Eugenio Naranjo, "A Systematic Methodology to Evaluate Prediction Models for Driving Style Classification," *Sensors*, vol. 20, no. 6, p. 1692, Mar. 2020, doi: [10.3390/s20061692](https://doi.org/10.3390/s20061692).