

A model for classifying breast masses in ultrasound images

Shereen Ekhlas Morsy¹, Neveen Mahmoud Abd-Elsalam¹, Zaid Abdu Al-Saidy¹,
Ahmed Hisham Kandil¹, Ahmed Mohammed El-Bialy^{1,2}, Abou-Bakr Mohammed Youssef¹

¹Department of Biomedical Engineering and Systems, Faculty of Engineering, Cairo University, Giza, Egypt

²Department of Systems and Biomedical Engineering, El-Shorouk Higher Institute of Engineering, Cairo, Egypt

Article Info

Article history:

Received Jan 24, 2024

Revised May 6, 2024

Accepted May 17, 2024

Keywords:

Artificial intelligence

Breast ultrasound

Convolutional autoencoders

Convolutional neural networks

Deep learning-based CAD

ABSTRACT

The most frequent type of cancer among women is breast cancer. Artificial intelligence (AI) researchers are developing automated systems to assist in the detection and classification of breast cancer. This study explores machine learning (ML) and deep learning (DL) as two AI methods for identifying benign and malignant breast tumors in ultrasound images. The investigation assesses the performance of various computer-aided detection and diagnosis (CAD) systems, which utilize either handcrafted features or deep features extracted from DL models. Furthermore, three models for CAD deep learning-based systems were implemented using convolutional neural networks (CNN), convolutional autoencoders (CAE), and deep features with CNN models, and compared with three traditional ML models based on handcrafted (texture) features. The results indicate that the deep features of the CNN model are promising, achieving a mean accuracy of 95% with a standard deviation of 1.1%.

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



Corresponding Author:

Shereen Ekhlas Morsy

Department of Biomedical Engineering and Systems, Faculty of Engineering, Cairo University

1 Gamaa Street, Giza, Egypt. Postal Code: 12613

Email: shereen.morsy.m@eng-st.cu.edu.eg

1. INTRODUCTION

Breast cancer had a worldwide impact in 2022, with over 2 million new cases reported and 670,000 fatalities [1]. Increased awareness and regular screening are essential for early detection. This is why high-income countries have succeeded in reducing breast cancer mortality. They experienced a significant 40% reduction in the age-standardized breast cancer death rate from the 1980s to 2020. Thus, if an annual reduction in worldwide mortality of 2.5% is achieved, this could save 2.5 million lives from breast cancer between 2020 and 2040 [1]. Breast cancer screening procedures are beneficial in early tumor detection. Breast ultrasound (BUS) is a technique used in conjunction with mammography to detect and characterize tumors, especially when breast density is high. Moreover, this imaging method is low-cost and produces no ionizing radiation. Additionally, it can distinguish between cancerous and non-cancerous tumors through sonographic traits. Despite the benefits, cancer screening still has a risk of inaccurate results (false positives and false negatives).

Machine learning (ML) is crucial in helping medical professionals find cancerous lesions earlier. Since the 1990s, computer-aided detection and diagnosis (CAD) systems have been developed and deployed within clinical environments to assist radiologists during the screening procedure, enhance screening's predictive accuracy, reduce subjectivity, and avoid mistaken diagnoses caused by weariness, eyestrain, or inexperience [2]. The rapid growth of ML has raised interest in employing these techniques to improve cancer screening accuracy in the medical imaging field [3]. In the domain of BUS image analysis, various ML approaches have been explored such as [4]-[7] primarily focusing on feature engineering techniques.

However, in recent years, there has been a significant surge in the utilization of deep learning (DL) algorithms for the analysis of medical images, including tasks such as detecting, segmenting, and classifying abnormalities. This is due to their ability to automatically learn and compute both global and local texture and shape features from medical images through a series of convolution and pooling layers. Consequently, the necessity for a separate feature engineering phase is eliminated.

Despite these advances, accurate breast cancer classification using deep learning algorithms still faces challenges. For instance, ultrasound (US) images may have low resolution, poor quality due to artifacts, or vary in tissue appearance [8]. Additionally, a large, annotated dataset is typically necessary for training the DL model. Hence, many research studies use limited datasets since large datasets are often unavailable which may lead to poor generalizability [9]. An alternative method to address this challenge involves implementing the concept of transfer learning, which enables the sharing of parameters from a pre-trained model on large datasets [10]-[14]. Alternatively, semi-supervised techniques [15]-[17] can be utilized, leveraging autoencoders for feature extraction and subsequently connecting them to fully connected (FC) layers for classification.

The objective of this paper is to explore and implement a range of traditional ML and deep learning techniques to address the challenge of extracting features related to breast cancer in ultrasound images to improve classification performance. The primary aim is to develop a CAD system capable of efficiently extracting these features, with the overarching goal of enabling early detection of breast cancer. Six classification systems were developed in this study: three based on ML algorithms which are support vector machine (SVM), k-nearest neighbor (KNN), and decision tree (DT) classifiers utilizing a set of texture features. Additionally, three deep learning-based classification systems were implemented, including a self-designed convolutional neural network (CNN), convolutional autoencoders (CAE), and deep features integrated with CNN models. Additionally, the study involves a comparative analysis of the implemented techniques to determine their effectiveness in achieving the desired outcome.

The subsequent sections of this paper are structured as follows: In section 2, the proposed methodologies are expounded upon in detail. Section 3 encompasses the presentation and discussion of results. Finally, section 4 encapsulates the conclusions drawn from the study.

2. RESEARCH METHOD

2.1. Conventional machine learning (ML) based computer-aided detection and diagnosis (CAD) systems

The common architecture of ML-based CAD systems integrates four key stages: preprocessing, segmentation, feature extraction and selection, and classification. Typically, such systems leverage handcrafted features to effectively classify breast lesions, yielding commendable results in practical application. These stages are proposed for the classification of breast cancer.

2.1.1. Image pre-processing

In the image preprocessing stage, the imaging artifacts and inconsistencies caused by different imaging conditions are removed with image preprocessing techniques for better detection [18]. Many preprocessing techniques are available like linear filtering, wiener filtering, wavelet despeckling, and wavelet filtering. [18]. In this work, the median filter is used.

2.1.2. Image segmentation

The procedure of image segmentation entails partitioning an image into distinguishable segments to identify objects and delineate borders [19]. This is crucial in breast images to isolate relevant regions of interest (ROI) for subsequent detection methods. Various segmentation algorithms are available, categorized into thresholding, edge-based, seeded region growing, cluster-based, and artificial neural network methods. This research utilizes a combination of threshold-based and edge-based segmentation techniques. Specifically, it employs edge boundary tracing through Otsu's thresholding, followed by the application of morphological operations to eliminate undesired pixels and fill gaps. Subsequently, exterior boundaries are delineated for each identified boundary.

2.1.3. Features extraction and selection

A group of handcrafted features is used to represent the content of the segmented breast area during the features' extraction stage. Texture features are the most common features utilized to analyze and interpret images by considering intensity variation and quantifying different characteristics like coarseness, smoothness, and regularity. Texture feature extraction techniques are categorized into four distinct types: structural approaches, statistical approaches, transform-based approaches, and model-based approaches [20]. The statistical approach features are the most typically used in breast CAD systems, thus it has been utilized

to analyze and interpret images in this work. The extracted features are 149 grey-level statistics which include:

- i) First-order statistical or histogram-based features: they are based solely on pixel values and are independent of pixel spatial distribution; nine features are calculated from the extracted ROIs as follows (mean, standard deviation, third moment, smoothness, skewness, kurtosis, harmonic mean, variance, and percentile) [21].
- ii) Matrix of gray-level-co-occurrence (GLCM): high-order statistical features depend on pixel values and their spatial correlations. The GLCM matrix is built in four directions=0°, 45°, 90°, and 135° with a one-pixel distance ($d=1$). The following measures are based on the GLCM: auto-correlation, correlation, contrast, dissimilarity, energy, cluster-prominence, cluster-shade, entropy, homogeneity, maximum probability, sum of squares variance, sum-average, sum-variance, sum-entropy, difference variance, difference entropy, information measure of correlation, information of correlation 1, information of correlation 2, inversed difference normalized [22].
- iii) Matrix of gray-level-run-length (GLRLM) statistical features of high-order: after constructing the GLRLM matrix, its properties are calculated as features. About eleven features are calculated using the zigzag method in four directions (0°, 45°, 90°, and 135°). These 44 features include short-run emphasis, long-run emphasis, gray-level non-uniformity, run length non-uniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, short-run low gray-level emphasis, long run low gray-level emphasis, long run high gray-level emphasis [23].

The process of minimizing input variables in a classification model, known as feature selection, entails retaining only the most pertinent features while discarding extraneous noise. Therefore, principal component analysis (PCA) was employed to identify and select the most relevant features.

2.1.4. Classification

The stage in which the extracted features are utilized to train a classifier to discriminate between positive and negative ROIs. Once trained, the classifier can predict the class of new ROIs. Here, three supervised machine-learning algorithms were utilized: SVM [24], KNN [25], and DT [26].

2.2. Deep learning (DL)-based computer-aided detection and diagnosis (CAD) systems

DL is a component of ML techniques that uses representation learning and artificial neural networks. Using multiple network layers is referred to as “deep learning” in this context. These deep networks necessitate a substantial volume of training data to achieve optimal performance [27]. Unfortunately, gathering the necessary training data to build models in medical applications is challenging. These problems could be resolved using a variety of strategies, including data augmentation and transfer learning for artificially increasing the instances of images for enhancing precision in results. Typically, four steps are followed in implementing a DL model for classification purposes. These steps are displayed in Figure 1 and will be briefly explained next.

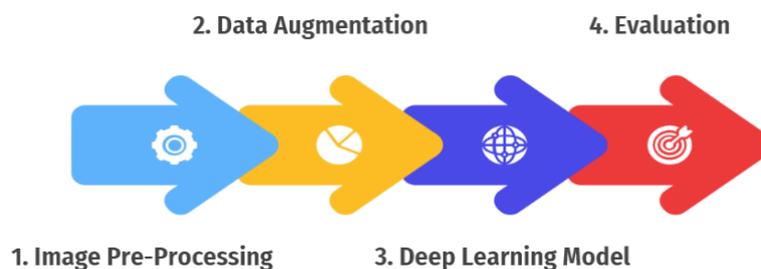


Figure 1. Implementing a DL model

2.2.1. Pre-processing

In an object or image detection/classification task, the dataset, specifically the images within it, stands as the cornerstone of a DL network model. Therefore, the pixel values of these images must be scaled correctly before creating this model. Typically, neural networks operate by processing inputs with weights that have relatively small values. However, inputs containing large integers, ranging from zero to 255, can potentially disrupt or slow down the learning process. As a result, pixel values need to be normalized by dividing each pixel's value by 255, resulting in pixels with values ranging from zero to one. This

normalization approach, along with standardization, is a standard technique in pixel value preprocessing. This entails converting the pixel values, so they are centered around a mean of zero and have a standard deviation of one. This transformation guarantees that the pixel values are normally distributed, facilitating comparison and analysis [28].

2.2.2. Data augmentation

It involves expanding a dataset by generating new data points from existing ones. This helps to enhance the data's diversity and size. Multiple transformations and alteration techniques are applied to the data to create new data points such as rotations, translation, zoom, shearing, and flipping. This procedure is frequently employed in ML to enhance the performance of models by making them more robust to noise and overfitting [29], [30].

2.2.3. Deep learning (DL) model

In this work, breast images were classified with three variants of deep architectures: CNN, CAE, and deep features with CNN models. The description of each architecture in detail as follows:

- i) CNN model: unlike traditional models, CNNs are self-sufficient, autonomously uncovering valuable patterns. CNNs are hierarchical neural networks, and the way the network is trained and the way the layers are designed affect how well the network performs. A typical CNN comprises three primary parts: convolutional layers, pooling layers, and activation functions [31]. Figure 2 illustrates the intricate architecture used in this model. CNN has four convolutional layers, 512 neurons, and a function for rectified linear unit (ReLU) activation. Batch normalization and the maximum pooling layer come after each convolutional layer. Following the convolutional layers are a flat layer, two FC layers with a total of 1024, 512 neurons each with ReLU activation. Finally, the output layer classifies data using the SoftMax layer.

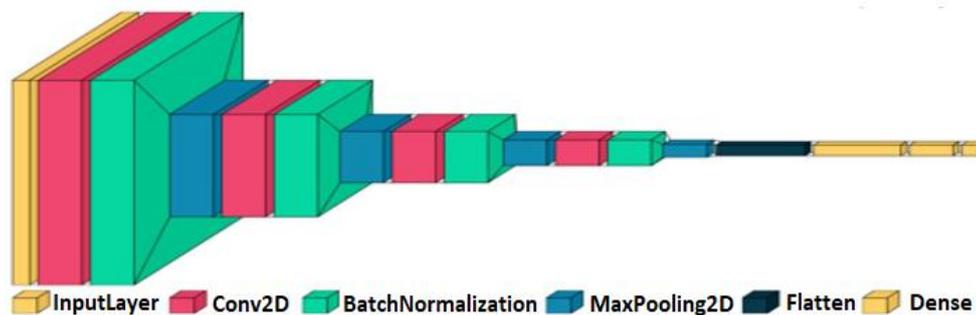


Figure 2. CNN model

- ii) CAE model: A CAE operates as a type of CNN designed to recreate its input data without being provided labels during training. This means that CAEs can be trained on unlabeled data, which is often easier and more affordable to obtain than labeled data. CAEs are recommended instead of supervised learning methods that require a large, labeled dataset, which can be difficult to obtain. A CAE has three main components: an encoder, a decoder, and an input layer. The encoder encodes the input into a smaller-sized version, and the decoder then utilizes this condensed version to recreate the original input. Figure 3 depicts the CAE architecture utilized in this work. The encoder comprises four convolutional layers, succeeded by batch normalization and max-pooling layers. The max-pooling layers down-sample the features. The decoder is composed of three convolutional layers, subsequently followed by batch normalization and up-sampling layers. Once the CAE is trained, only the encoder part is used for classification. The encoder is then followed by three FC layers with 1024, 512, and 2 neurons, respectively. The initial two FC layers utilize the ReLU activation function, while the final FC layer employs the SoftMax activation function, see Figure 4. The SoftMax layer converts the encoder output into a probability distribution over the different classes [32], [33].

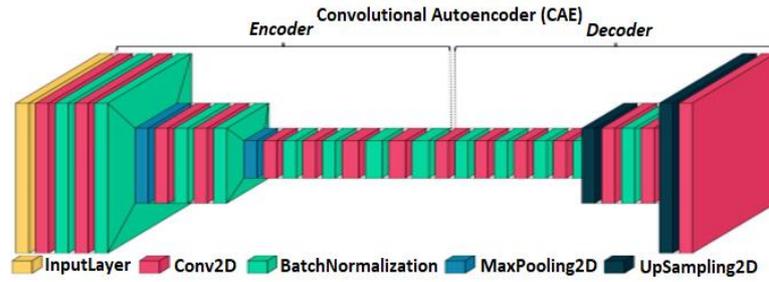


Figure 3. The architecture of autoencoder

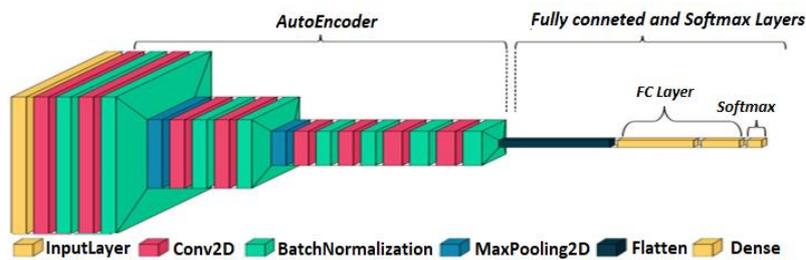


Figure 4. Model of CAE, only the encoder part is used after the CAE has been trained.

iii) Deep features with CNN model: in this part, as shown in Figure 5, the CAE algorithm was utilized for obtaining deep features and then feeding them to the CNN model. The same auto-encoder model structure explained in part CAE model is used for the extraction of deep features and then fed to the same CNN model structure explained in part CNN model. Once the CAE is trained, only the encoder component is utilized for feature extraction, which is then fed into a CNN for classification task [33], [34]. Figure 6 shows the final model.

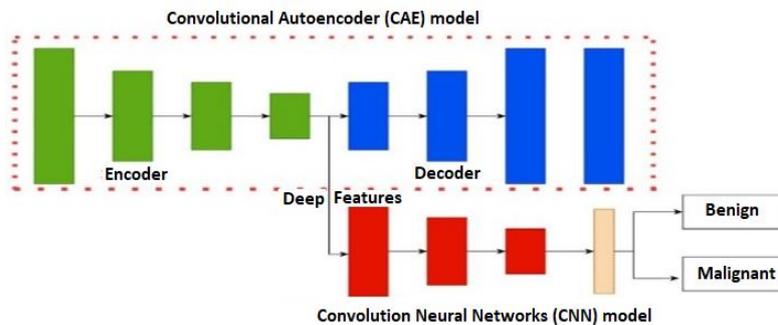


Figure 5. Deep features with CNN model

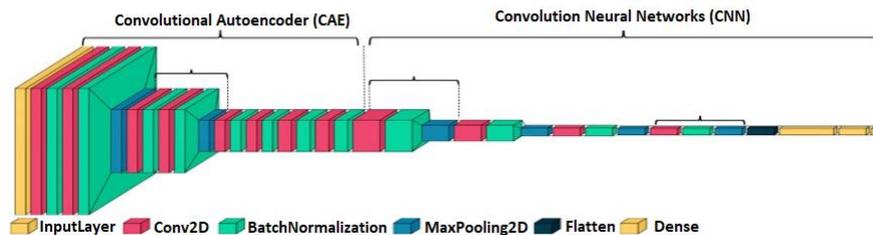


Figure 6. The final model of deep features with the CNN model

2.2.4. Evaluation

To assess the classification process, the dataset is partitioned into three subsets: the training dataset, the validation dataset, and the testing dataset. While the network evaluation is conducted on the training dataset, this alone does not provide a robust indication of the network's performance as a "predictive model" since it has processed this data before. Instead, the performance is evaluated on a separate dataset (unseen during training) that is called the validation dataset. This would be the power of the estimation for the network performance to make predictions for new data in the future. Keeping in mind that training a CNN may show some complex issues like over-fitting and convergence problems, whose resolution often requires repetitive adjustments in the network architecture or changing the learning parameters of the network. The last step is making predictions. Upon completion of the former steps and confirming the model's satisfactory performance, it is then employed to make predictions on new data, referred to as the test dataset. Consequently, the evaluation hinges on the accuracy of classification on the testing dataset. As outlined in (1), P denotes the entire positive cases or malignant cases, while N represents the entire negative cases or benign. TP signifies the true positive count, encompassing all malignant cases correctly predicted by the classifier as malignant, whereas TN denotes the true negatives, indicating all benign cases accurately predicted as benign.

$$Accuracy = (TP + TN)/(P + N) \quad (1)$$

3. RESULTS AND DISCUSSION

3.1. Dataset

The breast ultrasound images (BUSI) dataset [35] is a widely recognized publicly available dataset of breast ultrasound images that is categorized into three classes: normal, benign, and malignant images, accompanied by their corresponding ground truth data. From this dataset, a subset comprising 168 benign and 168 malignant images was selected for training and testing in our experiments. The code was written in Python and run-on Google Colab [36] which is a web-based service that offers access to Google's robust hardware infrastructure, encompassing graphics processing units (GPUs) and tensor processing units (TPUs), to facilitate the execution of ML experiments.

3.2. Conventional machine learning (ML)-based computer-aided detection and diagnosis (CAD) systems' results

Researchers exert significant efforts in developing ML-based CAD systems for breast cancer classification utilizing various techniques. Many recent studies such as Li *et al.* [37] developed a CAD system for breast tumor classification using radiomic features extracted from multimodal BUS images. Their method achieved an accuracy of 84.12%, with an area under curve (AUC) of 91.90%. Also, Hsu *et al.* [38] utilized quantitative features from ultrasound parametric images, achieving an accuracy of 89.40%, with an AUC of 96.00%. Chang and Chen [39] introduced an XGBoost classifier with a remarkable testing accuracy of 94.00%. However, there remains ample scope for further research and innovation in this area to enhance the accuracy and robustness of such systems. In this study, the initial stage involves preprocessing, where ultrasound images undergo enhancement and preparation for subsequent analysis. The enhanced images are then segmented, as demonstrated in Figure 7. Figure 7(a) displays an original benign image, while Figure 7(b) illustrates the resulting segmented image. Conversely, Figure 7(c) exhibits an original malignant image, and Figure 7(d) showcases the resultant segmented image. In the third stage, 149 gray-level statistics features are extracted and the PCA is then used to choose the most essential features. In the final stage, SVM (kernel=linear, gamma=auto), KNN, and DT (random state=0) classifiers are employed, and Tenfold cross-validation is used to evaluate the system. Experiments run 10 times, then the mean and standard deviation are calculated to check the overfitting. The classification accuracy after ten times running is calculated and shown in Table 1.

The results revealed that, for SVM, the mean accuracy across the ten runs is 89.6%, with a standard deviation of 4.35. This indicates a relatively consistent performance of the SVM classifier in accurately classifying the data, with minimal variability. In contrast, both KNN and DT classifiers show lower mean accuracies of 72.4% and 72.1% respectively, with higher standard deviations of 7.57 and 7.58. This suggests more variability in performance across runs and generally lower accuracy compared to SVM.

Figure 8 shows the receiver operating characteristic curve (ROC) and AUC for the three used classifiers. The SVM classifier achieved an AUC of 0.938, indicating a high discriminatory ability in distinguishing between classes as shown in Figure 8(a). On the other hand, both KNN (Figure 8(b)) and DT (Figure 8(c)) classifiers achieved an AUC of 0.741. These AUC values further support the superiority of the SVM classifier in accurately classifying the data, highlighting its effectiveness in this classification task. Overall, these results indicate that the SVM classifier outperforms KNN and DT in terms of mean

accuracy and consistency across multiple runs, making it a more reliable choice for classification in this context. While the current accuracy is noteworthy, further enhancements could be achieved by incorporating additional features, an avenue for exploration in future research.

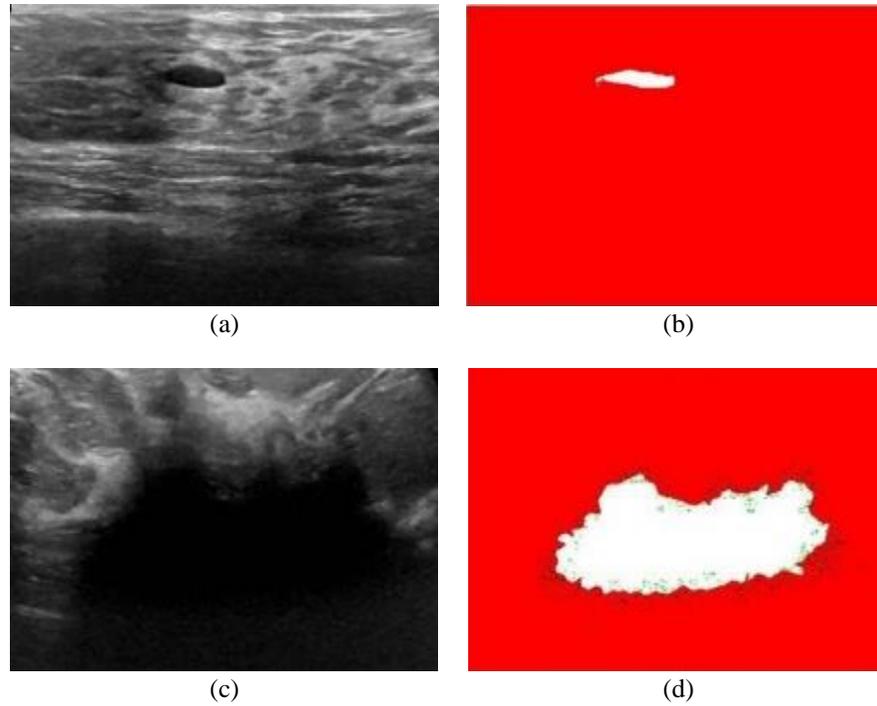


Figure 7. Segmentation results in (a) the original image benign, (b) the segmented image benign, (c) original image malignant, and (d) segmented image malignant

Table 1 Accuracy of different machine-learning classifiers

Classifier	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	STD
SVM	83.3	95.8	87.5	91.6	91.6	82.6	91.3	91.3	95.6	86.9	89.6	4.35
KNN	79.1	75	66.6	66.6	58.3	78.2	65.2	82.6	78.2	78.2	72.4	7.57
DT	79.1	70.8	70.8	66.6	62.5	69.5	78.2	56.5	82.6	78.2	72.1	7.58

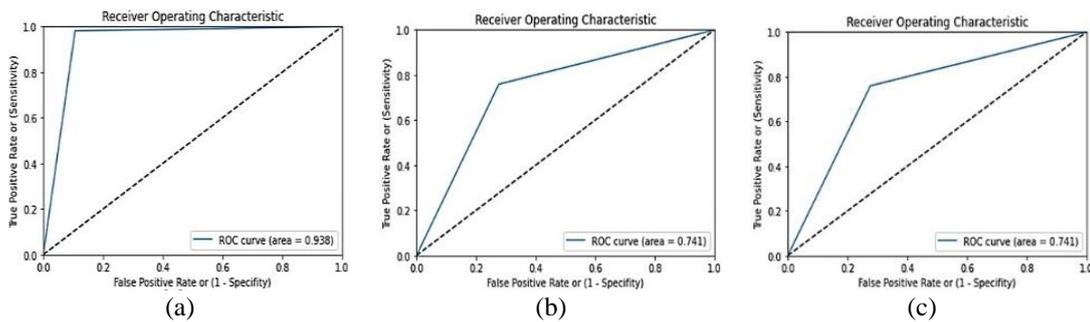


Figure 8. The ROC and AUC curves for the used classifiers in (a) SVM, (b) KNN, and (c) DT

3.3. Deep learning (DL)-based computer-aided detection and diagnosis (CAD) systems' results

To underscore the effectiveness of DL in breast tumor classification, experiments were conducted using the three described DNN models. The initial stage in developing our suggested model is to prepare the datasets for training and evaluation. This involves resizing the datasets using cubic interpolation to match the input requirements of the model and then preprocessing. For augmentation, rotations of up to 30

degrees, shearing within a range of 0.2, zooming with a factor of 0.2, horizontal flipping, and shifting in width and height by 0.1 were applied. Subsequently, the images were randomly shuffled, and the dataset was divided into 70% for the training set, 10% for the validation set, and 20% for the test set, as depicted in Figure 9.

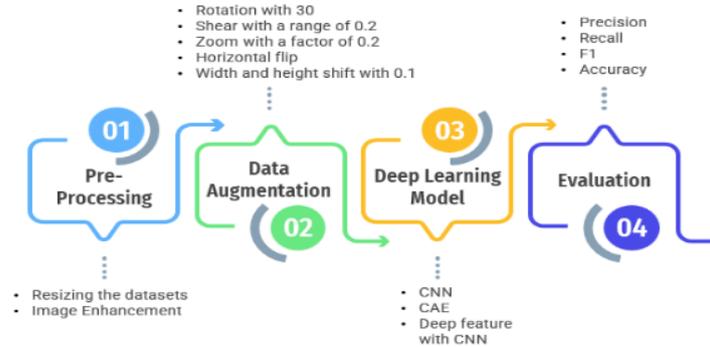


Figure 9. The proposed algorithm

For CNN, first, we resized each image to 96×96 to be the model's input. Also, the Adam optimization algorithm was employed during the training process. The model underwent training for 100 epochs to attain its optimal parameters. The activation function of convolutional layers is ReLU, and the SoftMax activation function is employed in the final layer. While the CAE was trained without labels for 100 epochs using the mean square error (MSE) loss function. The root mean square propagation (RMSprop) optimization algorithm was employed with a learning rate (lr) of 0.0001. After training, the CAE's weights were saved for the next stage. The decoder was discarded, and the encoder was used to extract features, which were then fed into three FC layers and a SoftMax layer to transform the encoder into a classifier.

The deep features with the CNN model were trained using the same parameters of the CAE and CNN models. It differs from the CAE model in that the part of the three FC layers was replaced with the whole CNN model. The autoencoder part used the RMSprop optimizer and the Adam optimizer was used for the CNN part, and both were trained for 100 epochs. The used loss function was categorical cross-entropy. To effectively assess the models' performance, we leveraged a callback function. This function closely tracks the training process, specifically monitoring performance on the validation set. When a model outperforms its predecessors on the validation set, the weights of the model are saved. This meticulous approach allows us to harness the best-performing model for testing against fresh, unseen images, ensuring a comprehensive evaluation. During the 100 epochs training, the validation accuracy is computed after each epoch, the callback function checks the highest value achieved and then saves that model. Figures 10-12 show the accuracy and loss results of training and validation of CNN (Figures 10(a) and (b)), CAE (Figures 11(a) and (b)), and deep features with CNN models (Figures 12(a) and (b)).

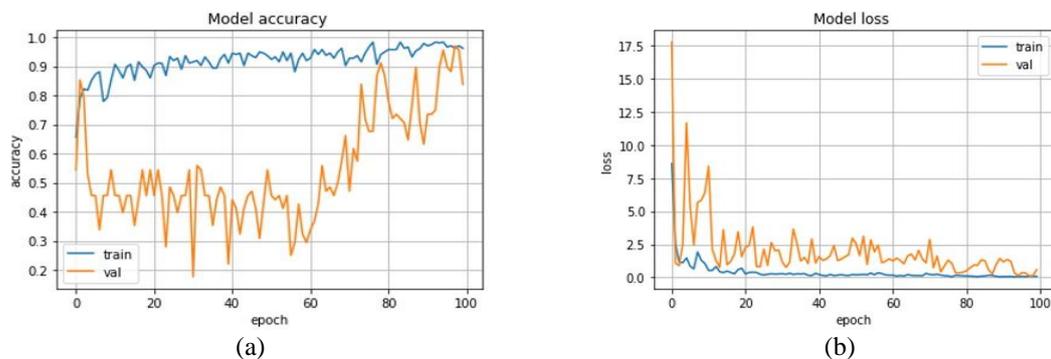


Figure 10. The performance of the CNN model throughout the training process in (a) accuracy and (b) loss

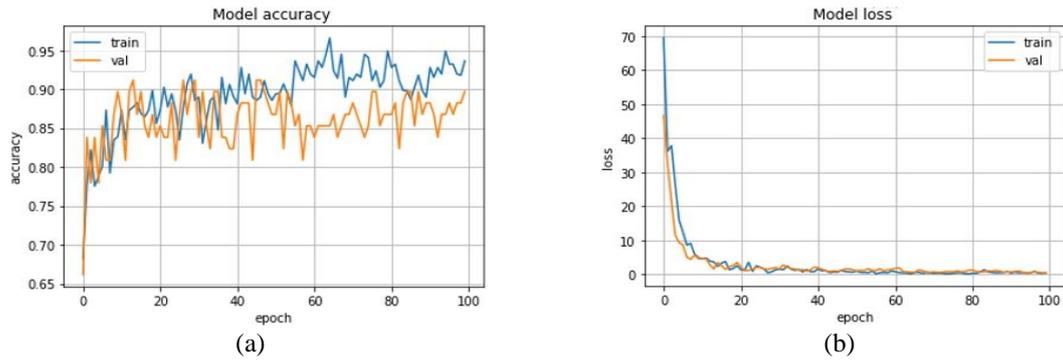


Figure 11. The performance of the CAE model throughout the training process in (a) accuracy and (b) loss

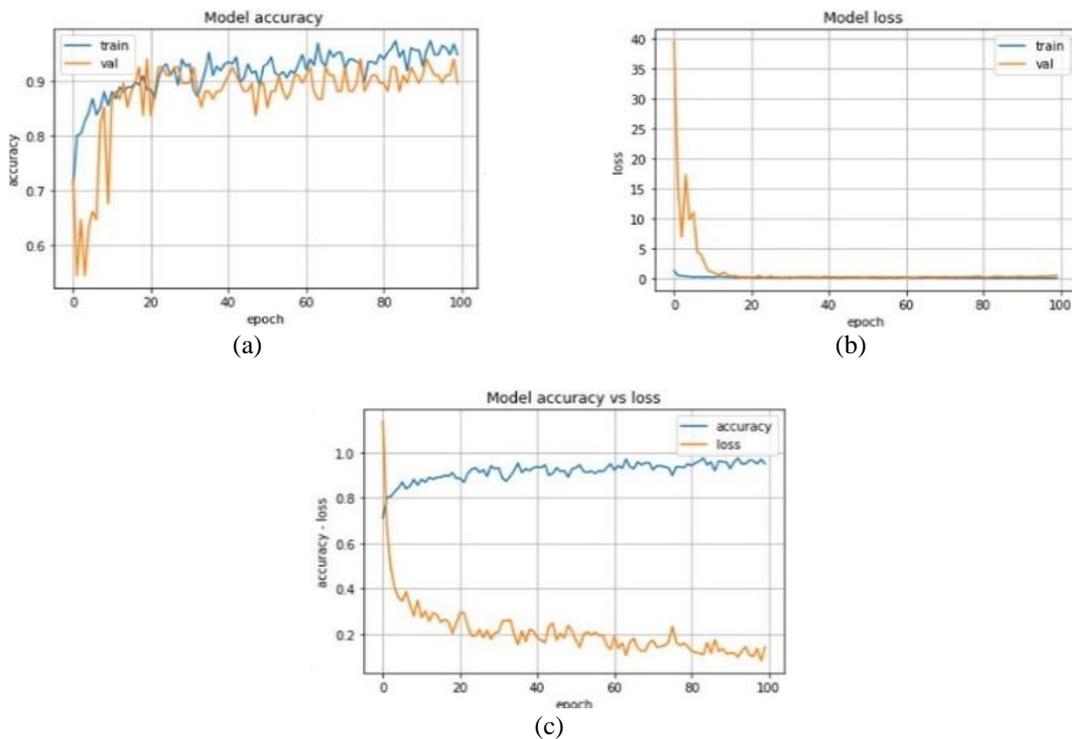


Figure 12. Deep features with CNN model performance during the training process in (a) accuracy, (b) loss, and (c) accuracy vs loss

To assess the performance of the three executed models, several evaluation metrics were employed, including precision (Pre.), recall (Rec.), F1 score (F1), and accuracy (Acc.). To verify result consistency, the experiments were run ten times, and both the mean and standard deviation were calculated to check the potential for overfitting. Table 2 summarizes the outcomes obtained from the testing dataset after conducting ten runs. The results indicate that all models exhibit consistent performance, as evidenced by low standard deviation values across all iterations (#1 to #10). Notably, the deep feature with the CNN model consistently surpasses both CNN and CAE in precision, recall, F1-score, and accuracy across all iterations. Furthermore, the mean values for all evaluation metrics significantly exceed those of both CNN and CAE, underscoring the superior effectiveness of the deep feature with the CNN model.

Additionally, the proposed model exhibits superior performance compared to our previous study [40], which employed transfer learning with various pre-trained models on the same dataset (BUSI). Our earlier investigation identified the EfficientNetB7 model as the most accurate, achieving an 88% accuracy rate on segmented input images. A recent study [41], that also employed the BUSI dataset, introduced a hybrid methodology merging a pre-trained CNN with optimization techniques and ML for breast tumor

diagnosis. In this approach, the pre-trained CNN model ResNet-50 was utilized for feature extraction, binary gray wolf optimization was applied for feature selection, and classification was conducted using SVM. The study reported an accuracy rate of 84.9% as a result of this approach.

Table 2. Results of the testing dataset for ten times runs

Model	Measures	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	STD
CNN	Pre.	0.92	0.9	0.94	0.87	0.93	0.87	0.89	0.91	0.91	0.9	0.904	0.022
	Rec.	0.91	0.89	0.94	0.88	0.93	0.85	0.88	0.92	0.92	0.89	0.901	0.026
	F1	0.92	0.9	0.94	0.87	0.93	0.84	0.88	0.91	0.91	0.9	0.9	0.028
	Acc.	0.91	0.9	0.94	0.88	0.93	0.84	0.88	0.91	0.91	0.9	0.9	0.026
CAE	Pre.	0.92	0.9	0.93	0.9	0.88	0.87	0.87	0.88	0.93	0.9	0.898	0.021
	Rec.	0.92	0.9	0.93	0.9	0.88	0.87	0.87	0.88	0.93	0.9	0.898	0.021
	F1	0.91	0.9	0.93	0.9	0.88	0.87	0.87	0.88	0.93	0.9	0.897	0.021
	Acc.	0.91	0.9	0.93	0.9	0.88	0.87	0.87	0.88	0.93	0.9	0.897	0.021
Deep feature with CNN	Pre.	0.95	0.96	0.94	0.95	0.94	0.95	0.97	0.94	0.96	0.96	0.952	0.009
	Rec.	0.95	0.96	0.94	0.96	0.94	0.95	0.97	0.94	0.96	0.96	0.953	0.01
	F1	0.94	0.96	0.94	0.96	0.94	0.94	0.97	0.94	0.96	0.96	0.951	0.011
	Acc.	0.94	0.96	0.94	0.96	0.94	0.94	0.97	0.94	0.96	0.96	0.951	0.011

In summary, our proposed deep feature with the CNN model demonstrates encouraging outcomes, attaining a mean accuracy of 95% with a standard deviation of 1.1%, which aligns well with existing research in the field. Nonetheless, while these results have been achieved using a widely utilized dataset, there remains a pressing need for the system to undergo training with an ample quantity of BUS images.

4. CONCLUSION

This study investigates the performance of various CAD systems employing both handcrafted and deep features. Handcrafted features are integrated into traditional ML systems, while deep features are extracted from within DL models. We developed three CAD models utilizing SVM, KNN, and DTs for ML-based systems, alongside three DL-based systems employing a convolutional neural network, a convolutional autoencoder, and a model utilizing deep features with CNN. The performance of these models is evaluated using a publicly available BUS image dataset, with each model trained independently on the dataset to classify tumors within the images. Our findings indicate that the deep features of the CNN model exhibit promising performance potential, even when trained on a limited dataset, achieving a mean accuracy of 95%. The results suggest that our proposed model can reduce reliance on handcrafted feature selection methods. Future endeavors will involve expanding our analysis to encompass larger datasets, ensuring the model's robustness and generalizability across diverse scenarios.

ACKNOWLEDGEMENTS

The authors thank the Egyptian Academy of Scientific Research and Technology (ASRT) for the financial support.

REFERENCES

- [1] World Health Organization, "Breast cancer." Accessed: Mar. 20, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [2] M. Giger, "Computer-aided diagnosis in diagnostic mammography and multimodality breast imaging," *RSNA Categorical Course in Diagnostic*, vol. 60637, pp. 205–217, 2004.
- [3] D. Painuli, S. Bhardwaj, and U. Köse, "Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review," *Computers in Biology and Medicine*, vol. 146, p. 105580, Jul. 2022, doi: 10.1016/j.cmpbiomed.2022.105580.
- [4] Kriti, J. Virmani, and R. Agarwal, "Effect of despeckle filtering on classification of breast tumors using ultrasound images," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 2, pp. 536–560, Apr. 2019, doi: 10.1016/j.bbe.2019.02.004.
- [5] W. K. Moon, I.-L. Chen, A. Yi, M. S. Bae, S. U. Shin, and R.-F. Chang, "Computer-aided prediction model for axillary lymph node metastasis in breast cancer using tumor morphological and textural features on ultrasound," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 129–137, Aug. 2018, doi: 10.1016/j.cmpb.2018.05.011.
- [6] Y. Liu, L. Ren, X. Cao, and Y. Tong, "Breast tumors recognition based on edge feature extraction using support vector machine," *Biomedical Signal Processing and Control*, vol. 58, p. 101825, Apr. 2020, doi: 10.1016/j.bspc.2019.101825.
- [7] Q. Huang, Y. Huang, Y. Luo, F. Yuan, and X. Li, "Segmentation of breast ultrasound image with semantic classification of superpixels," *Medical Image Analysis*, vol. 61, p. 101657, Apr. 2020, doi: 10.1016/j.media.2020.101657.
- [8] X. Qi *et al.*, "Automated diagnosis of breast ultrasonography images using deep neural networks," *Medical Image Analysis*, vol. 52, pp. 185–198, Feb. 2019, doi: 10.1016/j.media.2018.12.006.
- [9] M. H. Yap *et al.*, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018, doi: 10.1109/JBHI.2017.2731873.

- [10] M. Byra *et al.*, "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion," *Medical Physics*, vol. 46, no. 2, pp. 746–755, Feb. 2019, doi: 10.1002/mp.13361.
- [11] L. G. Falconi, M. Perez, and W. G. Aguilar, "Transfer learning in breast mammogram abnormalities classification with mobilenet and nasnet," in *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, Jun. 2019, pp. 109–114, doi: 10.1109/IWSSIP.2019.8787295.
- [12] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, p. e6201, Jan. 2019, doi: 10.7717/peerj.6201.
- [13] N. M. Abd-Elsalam, S. A. Fawzi, and A. H. Kandil, "Comparing different pre-trained models based on transfer learning technique in classifying mammogram masses," in *2020 30th International Conference on Computer Theory and Applications (ICCTA)*, IEEE, Dec. 2020, pp. 54–59, doi: 10.1109/ICCTA52020.2020.9477663.
- [14] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, "Comparison of Transferred deep neural networks in ultrasonic breast masses discrimination," *BioMed Research International*, vol. 2018, pp. 1–9, Jun. 2018, doi: 10.1155/2018/4605191.
- [15] E. Zhang, S. Seiler, M. Chen, W. Lu, and X. Gu, "BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis," *Physics in Medicine & Biology*, vol. 65, no. 12, p. 125005, Jun. 2020, doi: 10.1088/1361-6560/ab7e7d.
- [16] C.-Y. Lee, G.-L. Chen, Z.-X. Zhang, Y.-H. Chou, and C.-C. Hsu, "Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network?," *Journal of Healthcare Engineering*, vol. 2018, pp. 1–10, Dec. 2018, doi: 10.1155/2018/8413403.
- [17] M. Song and Y. Kim, "Deep representation for the classification of ultrasound breast tumors," in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, IEEE, Jan. 2022, pp. 1–6, doi: 10.1109/IMCOM53663.2022.9721796.
- [18] Y. Lu, J.-Y. Li, Y.-T. Su, and A.-A. Liu, "A review of breast cancer detection in medical images," in *2018 IEEE Visual Communications and Image Processing (VCIP)*, IEEE, Dec. 2018, pp. 1–4, doi: 10.1109/VCIP.2018.8698732.
- [19] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 2nd ed. New Jersey: Prentice Hall, Inc., 2002.
- [20] E. Ehrlich and D. Fanelli, "Segmentation," in *The Financial Services Marketing Handbook*, Wiley, 2012, pp. 17–32, doi: 10.1002/9781118531716.ch1.
- [21] M. I. Daoud, S. Abdel-Rahman, T. M. Bdair, M. S. Al-Najar, F. H. Al-Hawari, and R. Alazrai, "Breast Tumor classification in ultrasound images using combined deep and handcrafted features," *Sensors*, vol. 20, no. 23, p. 6838, Nov. 2020, doi: 10.3390/s20236838.
- [22] X. Huang, X. Liu, and L. Zhang, "A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation," *Remote Sensing*, vol. 6, no. 9, pp. 8424–8445, Sep. 2014, doi: 10.3390/rs6098424.
- [23] F. Albreghsen and B. Nielsen, "Texture classification based on cooccurrence of gray level run length matrices," *Australian Journal of Intelligent Information Processing Systems*, vol. 6, no. 1, pp. 38–45, 2000.
- [24] V. Jakkula, "Tutorial on support vector machine (SVM)," *School of EECS, Washington State University*. pp. 1–13, 2011. [Online]. Available: <http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf>
- [25] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of Translational Medicine*, vol. 4, no. 11, pp. 218–218, Jun. 2016, doi: 10.21037/atm.2016.03.37.
- [26] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [27] S. Sakib, N. Ahmed, A. J. Kabir, and H. Ahmed, "An overview of convolutional neural network: its architecture and applications," *Preprints*, p. 2018110546, 2018, doi: 10.20944/preprints201811.0546.v1.
- [28] S. J. H. Ma, "Image Normalization, a Basic Requirement for Computer-based Automatic Diagnostic Applications," *Image (Rochester, N.Y.)*, vol. 44, no. 0, pp. 0–26, 2001.
- [29] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- [30] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: a survey," Apr. 2022, [Online]. Available: <http://arxiv.org/abs/2204.08610>.
- [31] G. Latif, M. O. Butt, F. Y. A. Anezi, and J. Alghazo, "Ultrasound image despeckling and detection of breast cancer using deep CNN," in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, IEEE, Oct. 2020, pp. 1–5, doi: 10.1109/RIVF48685.2020.9140767.
- [32] V. E. Balas, R. Kumar, and R. Srivastava, Eds., *Recent Trends and advances in artificial intelligence and internet of things*, vol. 172. in Intelligent Systems Reference Library, vol. 172. Cham: Springer International Publishing, 2020, doi: 10.1007/978-3-030-32644-9.
- [33] K. Rai, F. Hojatpanah, F. Badrkhani Ajaei, and K. Grolinger, "Deep learning for high-impedance fault detection: convolutional autoencoders," *Energies*, vol. 14, no. 12, p. 3623, Jun. 2021, doi: 10.3390/en14123623.
- [34] J. D. Bodapati and N. Veeranjanyulu, "Feature extraction and classification using deep convolutional neural networks," *Journal of Cyber Security and Mobility*, vol. 8, no. 2, pp. 261–276, 2019, doi: 10.13052/jcsm2245-1439.825.
- [35] W. Al-Dhabyani, M. Gomma, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.
- [36] "Google Colaboratory." Accessed: Dec. 01, 2023. [Online]. Available: <https://colab.research.google.com>
- [37] Y. Li, Y. Liu, M. Zhang, G. Zhang, Z. Wang, and J. Luo, "Radiomics With attribute bagging for breast tumor classification using multimodal ultrasound images," *Journal of Ultrasound in Medicine*, vol. 39, no. 2, pp. 361–371, Feb. 2020, doi: 10.1002/jum.15115.
- [38] S.-M. Hsu, W.-H. Kuo, F.-C. Kuo, and Y.-Y. Liao, "Breast tumor classification using different features of quantitative ultrasound parametric images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 4, pp. 623–633, Apr. 2019, doi: 10.1007/s11548-018-01908-8.
- [39] C.-C. Chang and S.-H. Chen, "Developing a Novel machine learning-based classification scheme for predicting spcs in breast cancer survivors," *Frontiers in Genetics*, vol. 10, Sep. 2019, doi: 10.3389/fgene.2019.00848.
- [40] S. Ekhlash, N. M. Abd-Elsalam, Z. A. AlSaidy, A. H. Kandil, A. Al-bialy, and A. B. M. Youssef, "Comparing different deep-learning models for classifying masses in ultrasound images," in *Proceedings of 2023 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2023)*, 2024, pp. 318–328, doi: 10.1007/978-981-97-1335-6_28.
- [41] P. Khanna, M. Sahu, and B. Kumar Singh, "Improving the classification performance of breast ultrasound image using deep learning and optimization algorithm," in *2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES)*, IEEE, Dec. 2021, pp. 1–6, doi: 10.1109/TRIBES52498.2021.9751677.

BIOGRAPHIES OF AUTHORS



Shereen Ekhlis Morsy    is a Ph.D. student in systems and biomedical engineering Cairo University, M.Sc. degree in systems and biomedical engineering, Cairo University 2018. Her research interests include ultrasound imaging, computer vision, software development, pattern recognition, deep learning, and machine learning. Through her research, she focused on computational techniques to address challenges in healthcare, particularly in the areas of medical imaging analysis, image processing, and diagnostic systems. She can be contacted at email: shereen.morsy.m@eng-st.cu.edu.eg.



Neeven Mahmoud Abd-Elsalam    is a Ph.D. degree in biomedical engineering and systems from Cairo University, Egypt in 2021, an M.Sc. degree in computer engineering and control systems, from Mansoura University, Egypt, and a B.Eng. in computers and control systems, Mansoura University, Egypt. Her research interests include computer vision, artificial intelligence, and data science. Through her research, she focused on leveraging advanced computational techniques to address challenges in healthcare, particularly in the areas of medical imaging analysis, and diagnostic systems. She can be contacted at email: neveen-mahmoud@hotmail.com.



Zaid Abdu Al-Saidy    had a Ph.D. in biomedical engineering from the faculty of engineering at Cairo University, Egypt in 2019. Master in biomedical engineering from the faculty of engineering- Cairo University, Egypt in 2014. B.sc. in biomedical engineering from the faculty of engineering at Hashimite University, Jordan in 2003. His research interests include medical imaging, ultrasound imaging, image processing, and artificial intelligence. He can be contacted at email: zaabduh@gmail.com.



Ahmed Hisham Kandil    is a professor with a history of working in Systems and Biomedical Engineering since 1991, at Cairo University. He studied systems engineering at Case Western Reserve University in Cleveland, Ohio, where he earned his Ph.D. in 1991 in Systems Engineering Department. He earned his M.Sc. in systems and biomedical engineering from Cairo University in 1984. His thesis focused on a speech recognition system for single Arabic words. From 2004 to 2008, He was the SysNet lab manager in the Systems Engineering Department. He has been a consultant at the Center of Advanced Software and Biomedical Consultation (CASBEC), Faculty of Engineering, Cairo University, from 1991 to 2020. He served as the Executive Manager of CASBEC for 6 years. He held the position of General Manager at Union Aire Group's Top Technology factory for electronic assembly during its establishment in 2000. He worked in the Sakhr Group's speech processing division in 1995. His areas of interest include control engineering, mobile applications, hospital information systems, clinical engineering, speech processing, image processing, natural language processing, and pattern recognition. He can be contacted at email: ahkandil@eng1.cu.edu.eg, ahkandil_1@yahoo.com.



Ahmed Mohammed El-Bialy    is the head of the Biomedical and Systems Engineering department at El-Shorouk Higher Institute of Engineering. He is an accomplished professor with a track record of working in the fields of systems and biomedical engineering. In 1990, he obtained his Ph.D. from Case Western Reserve University in Cleveland, Ohio, where he studied systems engineering. In 1984, the M.Sc. in Systems and Biomedical Engineering from Cairo University was obtained. For three years, from 1991 to 1994, he served as the Executive Manager of the Cairo University Centre for Advanced Software Development and Applications' Multimedia Lab. He worked at the Engineering Office for Integrated Projects (EOIP) as an executive manager and consultant for software IT engineering from 1994 until 2003. He worked as a visiting professor at Louisville, Kentucky's University of Louisville., USA in 1998. His areas of interest in the study are artificial intelligence, computer vision, pattern recognition, and computers. He can be contacted at email: abialy_86@yahoo.com.



Abou-Bakr Mohammed Youssef    is an experienced professor with a demonstrated history of working in biomedical Engineering. He is one of the principal Founders of Cairo University's Department of Systems and Biomedical Engineering. He is double qualified to have a bachelor of electronics and communication from the Faculty of Engineering at Cairo University with honors and received his medical degree from Cairo University with excellence and honors. He became a Fellow of the US Society for Clinical Ultrasound in 1994 and a Registered Diagnostic Medical Sonographer in the USA in 1995, he has a keen interest in ultrasound imaging and strong teaching capabilities. The American Institute of Ultrasound in Medicine (AIUM) admitted him as a member in 1991 and a senior member in 1996. He received his Master of Radiation Therapy and Nuclear Medicine from Cairo University in 1987 and his Doctor of Medicine from Heidelberg University, he worked in The German Cancer Research Centre (DKFZ) Germany from 1986 to 1989 as a visiting professor during his work in DKFZ. He worked on non-invasive temperature estimation in biological tissues during hyperthermia He participated in the Diagnostic Medical Sonography program at both Drexel University and the University of Zagreb. Additionally, he earned a doctorate in philosophy from the Department of Systems and Biomedical Engineering with a focus on liver sonography in schistosomiasis cases in cooperation with Harvard Medical School and MIT and received his degree In 1982, he developed the first liver blood circulation simulator for schistosomiasis cases, becoming the first to establish the foundation for detecting the disease by ultrasound imaging. He graduated with honors from Cairo University's Faculty of Medicine in 1980 with excellence and Honor and from the same institution in 1987 with a Master's in Radiotherapy and nuclear medicine, he received a Master's in Electronics and Telecommunication Engineering in 1978 studying the effect of microwaves on the Human Eye. He established a business in the biomedical field regarding research and development and established and manufacturing facility for ultrasound imaging as well as a medical engineering service facility, he developed a company in medical insurance and health care facilities. He can be contacted at email: aboubakryoussef51@gmail.com.