



## OPEN Variational mode directed deep learning framework for breast lesion classification using ultrasound imaging

Manali Saini<sup>1,3</sup>, Sara Hassanzadeh<sup>1,3</sup>, Bushira Musa<sup>2</sup>, Mostafa Fatemi<sup>2,4</sup> & Azra Alizad<sup>1,2,4</sup>✉

Breast cancer is the most prevalent cancer and the second cause of cancer related death among women in the United States. Accurate and early detection of breast cancer can reduce the number of mortalities. Recent works explore deep learning techniques with ultrasound for detecting malignant breast lesions. However, the lack of explanatory features, need for segmentation, and high computational complexity limit their applicability in this detection. Therefore, we propose a novel ultrasound-based breast lesion classification framework that utilizes two-dimensional variational mode decomposition (2D-VMD) which provides self-explanatory features for guiding a convolutional neural network (CNN) with mixed pooling and attention mechanisms. The visual inspection of these features demonstrates their explainability in terms of discriminative lesion-specific boundary and texture in the decomposed modes of benign and malignant images, which further guide the deep learning network for enhanced classification. The proposed framework can classify the lesions with accuracies of 98% and 93% in two public breast ultrasound datasets and 89% in an in-house dataset without having to segment the lesions unlike existing techniques, along with an optimal trade-off between the sensitivity and specificity. 2D-VMD improves the areas under the receiver operating characteristics and precision-recall curves by 5% and 10% respectively. The proposed method achieves relative improvement of 14.47%(8.42%) (mean (SD)) in accuracy over state-of-the-art methods for one public dataset, and 5.75%(4.52%) for another public dataset with comparable performance to two existing methods. Further, it is computationally efficient with a reduction of 18 – 97% in floating point operations as compared to existing methods.

**Keywords** Ultrasound, Convolutional layers, Variational mode decomposition, Mixed pooling, Deep learning

Breast cancer is the most common cancer among females in the United States, with estimated over 3,00,000 of new cases in 2024<sup>1</sup>. It accounts for 31% of the new cases of all types of cancer diagnoses and is the second leading cause of cancer related death in the United States<sup>1</sup>. Ultrasonography is becoming increasingly important in diagnosing breast cancer, especially in low-resource settings where mammography is less feasible<sup>2</sup>. Ultrasound imaging increased the diagnostic sensitivity, but reduced the specificity, which necessitates the integration of computer-aided detection (CAD) systems to improve the diagnostic performance<sup>3,4</sup>. The effectiveness of ultrasound can be substantially influenced by the radiologist's experience, however, CAD systems are shown to be effective in augmenting the diagnostic performance of less experienced radiologists<sup>5,6</sup>. Although, these systems require additional research to be considered fully reliable<sup>7</sup>.

Recently, deep learning (DL) has shown significant advancements in breast cancer detection using mammograms<sup>8–11</sup>. With the increasing demand of resource-aware settings, researchers have integrated artificial intelligence-based techniques<sup>12</sup>, including DL with ultrasound systems, and shown promising results<sup>5,13–15</sup>. DL models perform automated feature extraction from the raw data, unlike the traditional CAD approaches<sup>16–19</sup>. The diagnostic accuracy of the DL models is demonstrated to be comparable with that of the ultrasound experts, which shows their clinical usefulness in minimizing the number of biopsies due to false positive outcomes<sup>20</sup>.

<sup>1</sup>Department of Radiology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA. <sup>2</sup>Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA. <sup>3</sup>Manali Saini and Sara Hassanzadeh contributed equally. <sup>4</sup>Mostafa Fatemi and Azra Alizad share senior authorship. ✉email: alizad.azra@mayo.edu

Further, DL is useful in improving the evaluation standards between experienced and less experienced clinicians<sup>20,21</sup>. It has substantially augmented the capabilities of B-mode ultrasound by providing a sophisticated method of image classification and feature extraction, outperforming the traditional CAD systems<sup>16</sup>. The integration of DL-based radiomics with B-mode ultrasound has shown improvement in the classification of breast masses<sup>22</sup>. Recent DL networks such as, convolutional neural networks have demonstrated remarkable performance in breast lesion classification showing comparable performance to radiologists<sup>23–25</sup>.

### Related works

Based on the recent research works for breast cancer detection using DL, we categorize them into two groups, with each group having its own distinct methodologies and distinctions. The first category uses DL models trained from scratch, based on mostly, convolutional neural network (CNN) architectures. For example, in Refs.<sup>26–30</sup>, different combinations of VGGNet, DenseNet, and ResNet, AlexNet, GoogleNet, and MobileNetV2 were used for classifying benign and malignant breast lesions. The study using ResNet50 was applied to a private dataset<sup>29</sup>, and the method used in Ref.<sup>30</sup> concatenated deep features extracted from AlexNet, MobileNetV2, and ResNet50 to feed to a support vector machine classifier using a public dataset. Another work in Ref.<sup>31</sup> explored InceptionV3 and Xception along with MobileNet and DenseNet121 models to categorize the malignant and benign lesions in a private dataset, and demonstrated superior performance as compared to MobileNet. The performance was further improved by using a modified VGG-like architecture, RepVGG on a large private dataset<sup>32</sup>. Huang et al. proposed a multi-class categorization of breast tumors by developing ROI-CNN based on VGGNet for tumor localization and graded CNN based on encoding and concatenation scheme for tumor classification using two public datasets<sup>33</sup>. However, the overall method was complex in terms of model parameters. Researchers proposed two cascaded CNNs based on Inceptionv3 architectures, i.e., Mt-Net to detect malignant lesions, and Sn-Net to identify the solid masses using a private dataset and demonstrated the robustness with respect to variations in shapes and sizes of breast lesions<sup>34</sup>. Recently, Ragab et al. employed an ensemble of VGG-16, 19, and SqueezeNet for diagnostic feature extraction and multi-layer perceptron for classification of benign and malignant lesions with a high accuracy on a public dataset<sup>35</sup>. However, the performance is subject to the effectiveness of image segmentation technique, which was used prior to classification in this work<sup>35</sup>.

The second category of DL models is increasingly emerging for breast cancer classification, i.e. transfer learning-based or pre-trained models with or without modifications. For example, literary works utilize the pre-trained ResNet-101, AlexNet, VGG16, VGG19, and DarkNet-53 models directly in Ref.<sup>36–39</sup> for different cases of breast cancer classification. Kaplan et al. in Ref.<sup>37</sup> proposed pyramid triple deep feature generator for image decomposition prior to feature extraction using pre-trained models, and applied it to a private dataset for identifying the benign and malignant lesions. Other research studies introduced different spatial attention mechanisms in the pre-trained ResNet-18 and VGG16 models<sup>40,41</sup>, and achieved comparable performances. Further, Sin et al. proposed a joint weakly and semi-supervised model based on R-CNN and ImageNet pretrained VGG-16 for localizing and classifying breast masses using a private and a public dataset<sup>42</sup>. The research works show that transfer learning can offer efficiency and a good starting point for learning domain-specific features. However, they may not always capture nuanced characteristics of images obtained from ultrasonography as effectively as models that are trained specifically on such data<sup>43</sup>.

### Objectives

The aforementioned scholarly studies utilize the standard CNN-based models for image classification-based feature extraction and their alterations for breast lesion classification. These studies establish effective discrimination between benign and malignant breast lesions in different datasets by extracting the corresponding pertinent features. However, these studies lack in terms of sufficient elucidation, need for segmenting the lesions prior, and monitoring the overall computational complexity involved in detecting breast cancer, which is crucial for timely diagnosis along with a reduction in the number of biopsies for benign lesions. In view of these aspects, we propose a novel variational mode directed low-complex deep learning framework to classify lesions in breast ultrasound images without segmenting them prior to classification. The proposed framework is built in two stages. The first stage includes elucidatory low level textural and boundary-based feature extraction using two-dimensional variational mode decomposition (2D-VMD) of the B-mode images of the scanned area, and the second stage includes high level feature extraction and classification of the images using a simplified sequential CNN-based deep learning network. To the best of our knowledge, this is the first work which uses 2D-VMD for deriving the discriminatory modes for feature extraction to assist a straightforward deep learning architecture towards breast cancer classification. The proposed DL architecture is guided by the 2D-VMD modes along with the input B-mode images and uses depth-wise separable convolutions along with mixed pooling and squeeze-and-excite-attention mechanisms for the first time to effectively classify the breast lesions with low computational complexity. The mixed-pooling and squeeze-and-excite-attention mechanisms help to improve the accuracy of breast lesion classification by focusing on the relevant features of the input images without having to extract the lesions or the regions of interest manually as done in many of the existing works which first segment the lesions and then classify them. Furthermore, the classification performance is evaluated using two publicly available datasets, and an in-hospital recorded dataset. The key contributions of this work are outlined as follows.

- Introduction of 2D-VMD for extracting discriminatory low level features of breast lesions to feed a deep learning network.
- Utilization of a computationally efficient deep learning network with depth-wise separable convolutions for extracting high level features to classify the lesions.
- An extensive performance evaluation and comparison of the proposed framework using two public datasets and an in-clinic dataset.

- Comparison of the computational complexity of the proposed framework with respect to existing methods for establishing its utility in real-time detection of breast lesion malignancy.
- An analysis of interpretability of the layers of the proposed deep learning network for discriminating benign and malignant breast lesions.

This work is presented and organized as follows. Section **Methods** describes the methods utilized in this work to build the proposed framework. Section **Results** presents the datasets used in this work and performance assessment results along with the comparison studies. Section **Discussion** discusses the results and future directions of the proposed work. Finally, section **Conclusion** concludes the work.

## Methods

In this section, we present the proposed framework which uses two-dimensional variational mode decomposition and a sequential CNN-based deep learning model for breast cancer classification, as shown in Fig. 1.

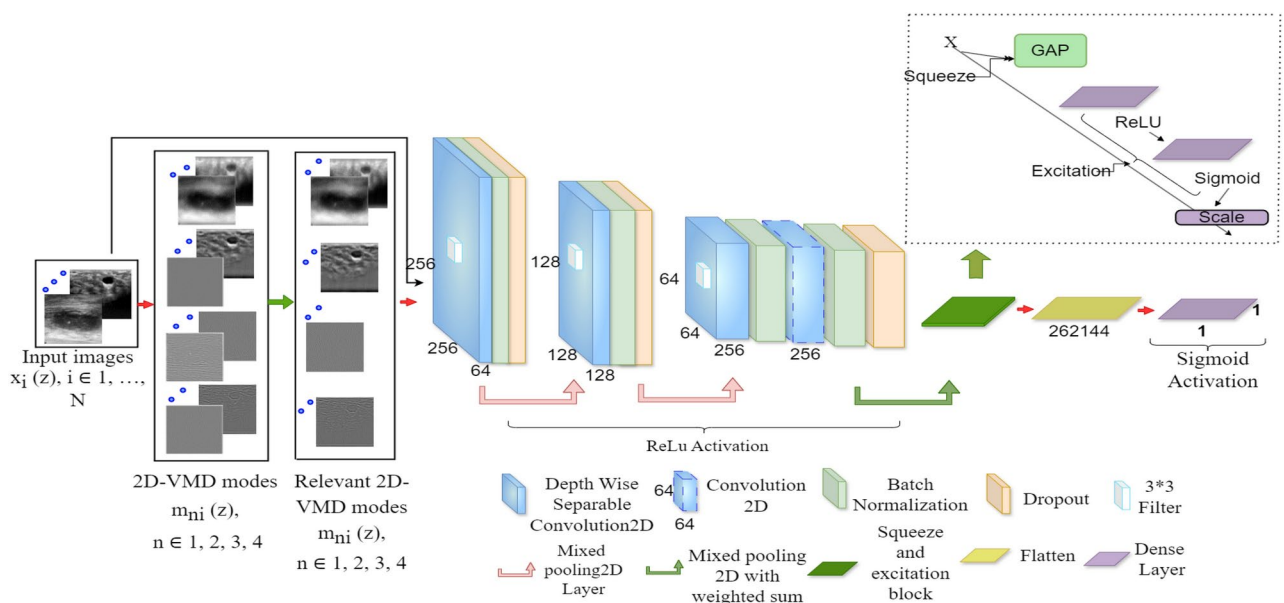
### Two-dimensional variational mode decomposition for ultrasound images

Zosso et al. in<sup>45</sup> proposed a two-dimensional (2D) extension of the non-recursive and fully adaptive one dimensional variational mode decomposition (VMD) technique, following the same analogy. 2D-VMD decomposes an ultrasound image in a sparse manner, into constituent sub-images or modes with each mode having a very limited bandwidth around a distinctive center frequency. 2D-VMD composes the image decomposition process into a constrained variational problem of minimizing an objective function derived from the definition of 2D analytical signal with one sided spectrum, shifted to baseband by combining with a complex exponential of the current center frequency estimate<sup>45</sup>. The corresponding function is represented as:

$$\min_{m_n, \omega_n} \left\{ \sum_n \alpha_n \left\| \nabla \left[ m_{AS,n}(z) e^{-j(\omega_n, z)} \right] \right\|_2^2 \right\} \quad \text{s.t.} \quad \forall z : \sum_n m_n(z) = I(z), \quad (1)$$

where an  $n^{th}$  2D band-limited mode/sub-image of an ultrasound image  $I(z)$  to be decomposed is denoted as  $m_n(z)$  with the center frequency  $\omega_n$ ,  $\nabla$  and  $\alpha_n$  denote the gradient with respect to  $z$  and data fidelity constant respectively, and  $m_{AS,n}(z)$  denotes the 2D analytical image obtained after Hilbert transformation<sup>45</sup>. Further explanation of above constrained minimization problem and its unconstrained formulation by adding a quadratic penalty term and Lagrangian multiplier along with optimization using alternate direct method of multipliers (ADMM) can be found in<sup>45</sup>. The solution to this minimization problem is obtained in iterations which produces a number of decomposed modes in frequency domain with distinctive center frequencies. The complete iterations-based process of 2D-VMD can be summarized in three steps as follows<sup>45</sup>.

1. Minimization with respect to modes  $m_n$  : The frequency domain of modes is updated using Weiner filter with the center as the current center frequency, and the corresponding mode update is obtained as:



**Fig. 1.** Proposed framework for breast cancer classification.

$$\hat{m}_n^{j+1}(\omega) = \left( \hat{I}(\omega) - \sum_{i \neq n} \hat{m}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right) \frac{1}{1 + 2\alpha_n |\omega - \omega_n|^2} \quad (2)$$

$$\forall \omega \in \Omega_n : \Omega_n = \{\omega \mid \langle \omega, \omega_n \rangle \geq 0\},$$

Here,  $\hat{m}_n(\omega)$  represents the frequency domain sub-image/mode,  $\hat{I}(\omega)$  represents image in frequency domain,  $\alpha$  is data fidelity constant,  $\lambda$  denotes Lagrangian multiplier, and  $j$  is the iteration number. The first few modes are updated with lower frequencies capturing the overall structure and outline of the ultrasound B-mode image with slow changes in intensity and texture. While the higher order modes are updated with higher frequencies that represent the finer details with rapid fluctuations in intensity and texture. In the context of breast lesions, the first few modes provide softer boundaries of the lesion and the higher order modes provide the sharp edges and details of the area surrounding the lesion.

2. Minimization with respect to center frequencies  $\omega_n$ : Center frequencies are obtained based on the first moment of the corresponding mode's power spectrum  $|\hat{m}_n(\omega)|^2$  on the spectral half plane  $\Omega_n$  as:

$$\omega_n^{j+1} = \frac{\int_{\Omega_n} \omega |\hat{m}_n(\omega)|^2 d\omega}{\int_{\Omega_n} |\hat{m}_n(\omega)|^2 d\omega} \quad (3)$$

The center frequencies of the sub-images vary from low to high values as the number of sub-images increase.

3. Maximization with respect to the Lagrangian Multiplier  $\lambda$ : Lagrangian multiplier in the unconstrained formulation is updated using a conventional gradient ascent with fixed time step  $\tau$  as:

$$\lambda^{j+1}(z) = \lambda^j(z) + \tau \left( f(z) - \sum_n m_n^{j+1}(z) \right), \quad (4)$$

where the term in the parentheses denotes the data residual.

The all-inclusive algorithm is provided as the ADMM optimization of 2D-VMD in algorithm 1<sup>45</sup>.

---

```

1: Initialize  $\{\hat{m}_n^0\}, \{\hat{\omega}_n^0\}, \hat{\lambda}^0, \leftarrow 0$ 
2: repeat
3:    $j \leftarrow j + 1$ 
4:   for  $n = 1 : N$  do
5:     Update  $\omega_n: \omega_n^{j+1} \leftarrow \frac{\int_{\mathbb{R}^2} \omega |\hat{m}_{AS,n}^{j+1}(\omega)|^2 d\omega}{\int_{\mathbb{R}^2} |\hat{m}_{AS,n}^{j+1}(\omega)|^2 d\omega}$ 
6:     Update  $m_n: m_n^{j+1}(z) \leftarrow \Re \left\{ \mathcal{F}^{-1} \hat{m}_{AS,n}^{j+1}(\omega) \right\}$ 
7:   end for
8:   Dual Ascent:
      $\hat{\lambda}^{j+1}(\omega) \leftarrow \hat{\lambda}^j(\omega) + \tau \left( \hat{I}(\omega) - \sum_n \hat{m}_n^{j+1}(\omega) \right)$ 
9: until convergence:  $\sum_n \left\| \hat{m}_n^{j+1} - \hat{m}_n^j \right\|_2^2 / \left\| \hat{m}_n^j \right\|_2^2 < N\epsilon$ 

```

---

**Algorithm 1.** Pseudocode for 2D-VMD.

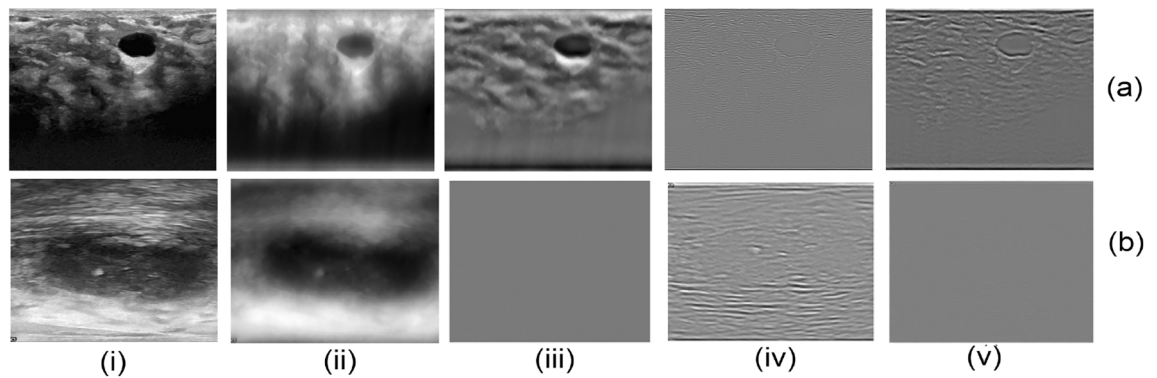
---

#### Decomposition of breast ultrasound images using 2D-VMD

In this section, we introduce the use of 2D-VMD with selective design parameters for the decomposition of ultrasound images with breast lesions such that the distinctive features are captured in one or more modes. Let  $I(z)$  denote an input B-mode image to be decomposed. The decomposed modes using 2D-VMD can be expressed as:

$$m_n\{z, f, \omega\} = VMD\{I(z)\} \quad (5)$$

The two method parameters i.e., data fidelity constant ( $\alpha$ ) and number of modes ( $N$ ) need to be selected exhaustively for obtaining the required features in the decomposed modes.  $\alpha$  determines the admissible bandwidth in each mode, and the adequate number of modes helps in maintaining the appropriate distribution



**Fig. 2.** Illustration of 2D-VMD modes of example images with benign and malignant lesions in (a) and (b) respectively taken from BUSI database: (i) Original image, (ii)-(v) Decomposed modes.

of energy of the image. Based on an extensive experimental evaluation,  $\alpha$  and  $N$  are selected as 500 and 4 respectively for capturing the distinctive features of benign and malignant lesions in breast ultrasound images. Figure 2a and b illustrate the decomposed modes of two example images with benign and malignant lesions respectively, with these method parameters. It can be observed from the figure that for both images, the first mode is able to capture the complete structure of image with lesion. Further, for the benign lesion, modes 2 and 4 display the definite lesion boundaries, whereas, the malignant lesion is captured in the third mode with improper boundary. This aligns with the fact that benign lesions have more definite boundaries as compared to malignant lesions. Further, the textural differences between the modes in benign and malignant lesions are notable. Therefore, it can be demonstrated that 2D-VMD modes are discriminatory for benign and malignant lesions. Further, it can also be observed that there are modes which have regular texture throughout the image, and thereby, do not contain relevant information, such as, mode 3 for benign, and modes 2 and 4 for malignant case. We discard these modes from the analysis for smooth learning process of the proposed deep learning network. For this purpose, we propose an entropy-based exclusion criteria, which first estimates the entropy of each mode of an image, and then rejects the mode with minimum entropy value. The selection of entropy is based on the fact that the mode which captures lesion, will have irregularities in the overall textural structure, while the one with uniform texture does not contain relevant features of the lesion area. The mathematical description of this criteria is provided as follows.

- Firstly, Entropy estimation is performed for each mode of the mode/ sub-image, as:

$$Et_{m_{ni}(z)} = -\sum \left( h_{m_{ni}(z)}^p \cdot \log 2 \left( h_{m_{ni}(z)}^p \right) \right) \quad (6)$$

where  $h_{m_{ni}(z)}^p$  denotes the histogram counts representing the distribution of pixel values of the  $n^{th}$  mode of  $i^{th}$  image, i.e.,  $m_{ni}$ .

- Reject the mode  $m_{ni}(z)$ , if:

$$Et_{m_{ni}(z)} = \min \left( Et_{m_i(z)} \right); \quad \text{where, } Et_{m_i(z)} = [Et_{m_{ni}(z)} \forall n] \quad (7)$$

where  $[]$  denotes set.

In this work, the images and the corresponding selected 2D-VMD modes based on above criteria are used as the inputs to train the proposed deep learning model, as shown in Fig. 1. The proposed model is described in the next sub-section.

### Proposed deep learning model

The proposed deep learning model for breast lesion classification is shown in Fig. 1. The layered structure comprises majorly three two-dimensional (2D) depth-wise separable convolutional layers and a 2D convolutional layer along with batch normalization, dropout, and 2D mixed pooling layers with and without weighted sum, a squeeze and excite block, flatten layer, and an output dense layer for classification, as illustrated in Fig. 1. In this sub-section, we describe the functioning of the proposed model layers for classification.

- The input data, consisting of B-mode images and the corresponding selected 2D-VMD modes are fed to the first layer of the proposed model, i.e., depth-wise separable convolutional layer with 64 filters. This layer performs a two-step convolution operation unlike the standard convolution<sup>48,49</sup>. Supplementary Fig. S1 illustrates the operation of both standard and depth-wise separable convolutional layers, for an  $N$  channel input image with dimensions  $I \times I$  and  $P k \times k$  filters<sup>48</sup>. It can be observed that for the standard convolution, for



each input channel is convolved with a particular filter, and all channels are convolved at the same time. For  $P$  filters, the corresponding output dimension is  $I \times I \times P$ . While, for the depth-wise separable convolution, it can be observed that the first step involves convolution operation with each channel at a time by dividing each of the prior filters into  $N \times k \times 1$  filters. The second step involves the point-wise convolution operation which is the standard convolution using  $1 \times 1$  filters, as shown in the figure. As a result, the overall operation of depth-wise separable convolution helps in reducing the total number of weights and parameters as required by the standard convolution. More details can be found in<sup>50</sup>. Batch normalization and dropout with a factor of 0.3 is used after this layer to reduce over-fitting.

- This work proposes the use of mixed pooling operation for down-sampling in the context of breast cancer classification, as shown in the figure. It is a combination of max-pooling and average pooling operation as shown by the dotted block in Supplementary Fig. 2. Mixed pooling helps in capturing the spatial information by adding the average pooling ( $Avg()$ ) after max-pooling ( $Max()$ ) and can reduce over-fitting. The feature map after using this operation is denoted as:

$$l'_i = Max(l_{i-1}) + Avg(l_{i-1})$$

where  $l_{i-1}$  represents the output feature map of the preceding layer  $i - 1$ . The feature map is more localized as compared to that obtained by using only the max-pooling operation. This further helps in extracting more discriminant features between the benign and malignant lesions. Previous related works utilized the conventional max-pooling mechanisms for down-sampling the convolved outputs, while we introduce the use of mixed-pooling for downsampling, which provides more robust and accurate features capturing the global as well as local information for enhancing the breast lesions discrimination. The output of this layer is again fed to a similar block as described previously. Further, it can be observed from Fig. 1 that the corresponding output is fed to a third block, including depth-wise separable convolution2D with 256 filters, batch normalization, convolution2D with 256 filters, batch normalization and dropout followed by a mixed-pooling2D layer with weighted sum. It uses an attention mechanism with sigmoid activation for assigning the weights to the max-pooling and average pooling outputs before adding them together, as shown in Supplementary Fig. 2. The corresponding output is obtained as:

$$l_i = z \times Max(l_{i-1}) + (1 - z) \times Avg(l_{i-1})$$

It helps in regulating the outputs of both the down-sampling layers by using the weighted attention mechanism, which further helps in keeping the relevant amount of local and global information about the lesions<sup>51</sup>. Further, it provides an attention mechanism to select an appropriate combination of max-pooling and average pooling mechanisms for down-sampling the deeper features.

- The attention-based mixed pooling output is further fed to a second attention mechanism, which uses a squeeze-and-excite-block. This block consists of initial global average pooling layer for squeezing each input feature map to a single scalar value<sup>52</sup>. These single valued feature maps are further presented as inputs to two dense/fully connected layers with ReLu and sigmoid activation respectively, to estimate the learned weights in the excite phase. Finally, these learned weights are used to scale/multiply the corresponding original input feature maps from the preceding mixed pooling layer, which adds the second attention mechanism and further enhances the relevant features for classification. It boosts the attention mechanism by recalibrating the channel-level features. Finally, it can be observed from Fig. 1 that the output is flattened for converting to a suitable shape, i.e., a vector of dimension 262144 ( $512 \times 512$ ) to be given as input to the output dense layer with sigmoid activation for binary classification of breast lesions, i.e., benign (assigned as '0') or malignant (assigned as '1').

The aforementioned layers are used to build the proposed network in TensorFlow Keras 2.2.1 deep learning framework in Python. Keras Tuner is utilized to select the optimal number of layers and combinations based on the optimal performance. The data is split into 80% training, 10% for validation, and 10% for testing. The training hyper-parameters such as batch size, epochs, steps per epoch, learning rate and algorithm, are discussed in the later section.

## Results

This section discusses about the performance assessment of the proposed framework for classification of breast lesions in ultrasound images. Firstly, we describe the different datasets used in this work. Secondly, we present the different learning parameters and classification metrics for analysis of the proposed model. Thirdly, we analyze the proposed framework without using 2D-VMD for demonstrating the effectiveness of 2D-VMD. Finally, we compare the framework with some of the commonly used recent methods for breast lesion classification.

### Datasets

This section provides a description of the two public datasets and an in-clinic dataset utilized in this study.

#### UDIAT dataset

This is the first public breast ultrasound dataset used in this work. It consists of 163 B-mode images with 110 benign and 53 malignant lesions, and only one lesion per image<sup>53</sup>. The data was acquired using Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer with a frequency of 8.5 MHz, at the UDIAT

Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain). All images were  $760 \times 570$  pixels in size and pre-processed for noise removal. The ground truth segmented masks were obtained from the manually segmented lesions by experienced radiologists in this dataset<sup>53</sup>. This data is available for research purposes only after request from the authors in<sup>53</sup>.

*Breast ultrasound images (BUSI) dataset*

The second public breast ultrasound dataset consists of 780 B-mode images including 437 benign and 237 malignant lesions, and 133 normal cases<sup>54</sup>. The data was acquired from 600 female volunteers with ages ranged between 25 and 75 years, using LOGIQ E9 and LOGIQ E9 Agile ultrasound system with 1 – 5 MHz transducers on ML6-15-D matrix linear probe, at Baheya hospital, Cairo, Egypt<sup>54</sup>. It is available at <https://scholar.cu.edu.eg/?=afahmy/pages/dataset>. All images were  $500 \times 500$  pxels in size, and pre-processed for artifact removal. For maintaining uniformity across all datasets used in this work, we analyzed the images containing only one lesion per image, and discarded the ones with multiple lesions. Therefore, the number of images was 421 for benign and 209 for malignant cases, leading to total 630 images.

*In-hospital recorded (IHR) dataset*

The third dataset is a private breast ultrasound imaging dataset consisting of 2205 B-mode images with 1315 malignant and 890 benign lesions. The data was acquired from 655 female volunteers on multiple days, with each volunteer having a pre-biopsy scan, and three post-biopsy scans. The data was collected using the Supersonic Imagine Aixplorer (SSI, Aix-en-Provence, France) system with a 4-15 MHz array transducer, at Mayo Clinic, USA. The institutional review board approval (IRB:12-003329 and IRB: 19-003028)) was received in compliance with the Health Insurance Portability and Accountability Act along with signed written informed consent with permission for publication from each enrollee prior to this study. Four B-mode image acquisitions were performed with two each in the longitudinal and transverse orientations. Some images were duplicated, and thereby needed to be removed from the analysis, resulting in 2066 images in total. All images were stored in PNG format. Each image was  $1400 \times 1050$  in size, and were kept without prior processing for any type of artifact removal, for the purpose of analyzing the robustness of the proposed framework with respect to artifacts.

**Training hyper-parameters**

For adequate training of the proposed network, different learning hyper-parameters and number of layers are used in this work. For this purpose, Keras tuner is used to automatically select the parameters based on optimal performance. These include dropout rate of 0.3, filter width of  $3 \times 3$ , stride of 1, batch sizes of 8 and 16, epochs between 150 – 250, and squeeze-and-excite ratio of 2. Besides these, Adam is used for optimization of the network, with a learning rate of 0.00001.

**Classification performance metrics**

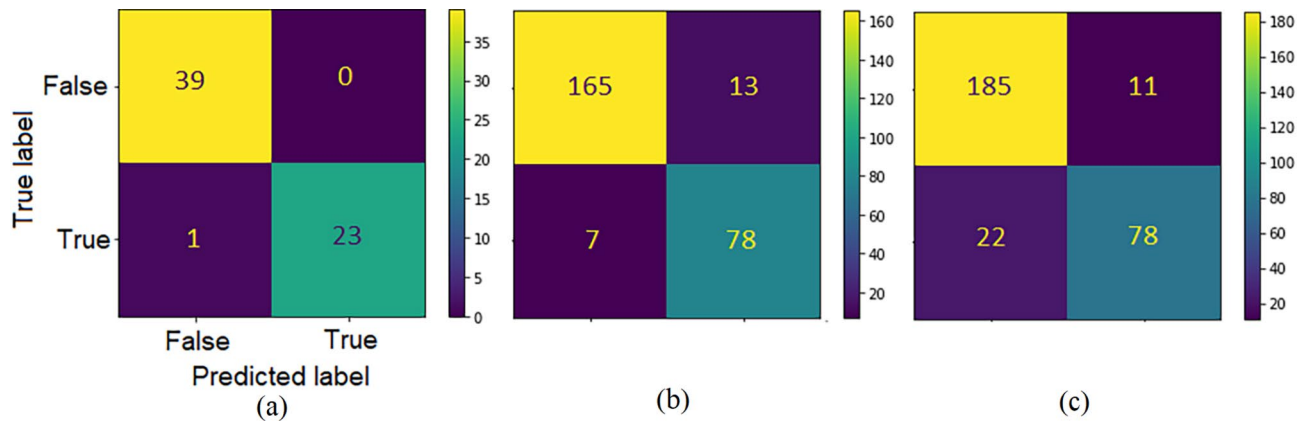
For analyzing the performance of the proposed framework, a train-validation-test split of 80%-10%-10% respectively is utilized for all the datasets in this work. To deal with the the class-imbalance in the datasets, binary focal cross entropy loss function is used during the training process for performance evaluation of classification task<sup>55</sup>. Further, it enhances the performance by focusing more on hard training samples<sup>55</sup>. For assessment of the proposed model, different standard classification metrics are used, such as, Accuracy (Ac), Precision (Pr), sensitivity/recall (Se), specificity (Sp), area under receiver operating characteristics (ROC) curve (AUROC). Further, we examine the area under precision recall curve (AUPRC) for analyzing the performance with respect to false positives and class imbalance. Description of these metrics is provided in<sup>56,57</sup>. Further, the confusion matrices are also presented for analyzing the class-wise performance of all datasets used in this work.

**Performance analysis**

Table 1 and Fig. 3 show the performance metrics and confusion matrices of the proposed framework respectively for breast lesion classification in all the three datasets used in this work. It can be observed from the table that for the public dataset *UDIAT*, the overall classification performance is highest, with an accuracy of 0.98, indicating that the proposed method is able to classify most of the breast lesions correctly. Further, it can be observed that the specificity, AUROC and AUPRC are all 1, which demonstrates that all benign lesions are predicted correctly. Further, the confusion matrix in Fig. 3 (a) for this dataset shows that there are no false positives and only one false negative, which demonstrates the outstanding classification performance of the proposed framework for this dataset. For the *BUSI* dataset, the overall classification accuracy is 0.93 and the areas under both curves are above 0.95, which indicates that the capability of the proposed framework in classifying the lesions with high performance for this dataset. However, it can be observed from the confusion matrix Fig. 3 (a) that the overall false positives are higher than the false negatives, however, the balance between the sensitivity and precision is still better, as indicated by *AUPRC* value of 0.98. Finally, for the *IHR* data, the performance is slightly low,

Dataset	Ac	Pr	Se	Sp	F1	AUROC	AUPRC
UDIAT	0.98	1.00	0.96	1.00	0.98	1.00	1.00
BUSI	0.93	0.86	0.92	0.93	0.89	0.97	0.98
IHR	0.89	0.88	0.78	0.94	0.83	0.94	0.89

**Table 1.** Performance metrics of the proposed framework.



**Fig. 3.** Confusion matrices for (a) *UDIAT*, (b) *BUSI*, and (c) *IHR* datasets. Here, False denotes Benign class, and True denotes malignant class.

however, it can be observed from the Table 1 and confusion matrix in Fig. 3 that the number of falsely detected malignant cases are lower, indicating a fair specificity, although the sensitivity is quite low. This may attribute to the presence of non pre-processed images in this dataset. However, the overall fair performance indicates the noise robustness capability of the proposed framework.

We analyze the variability of different parameters and classification performance metrics across the three datasets in Supplementary Table 1. We estimated the statistical parameters such as the standard deviation (*SD*) and the inter-quartile range (*IQR*) of the total number of images, number of benign and malignant images, and the classification performance metrics across the datasets. Substantial variations were observed in the number of images with a *SD* of 804.16 and *IQR* of 951.5. Similar variations can be noted for the number of benign and malignant images across the datasets. Further, the classification performance variability demonstrates higher *SD* (0.08 and 0.06) and *IQR* (0.09 and 0.08) for *Se* and *F1* respectively, while lower values (*SD*: 0.02, *IQR*: 0.03) for *AUROC*.

### Effectiveness of 2D-VMD

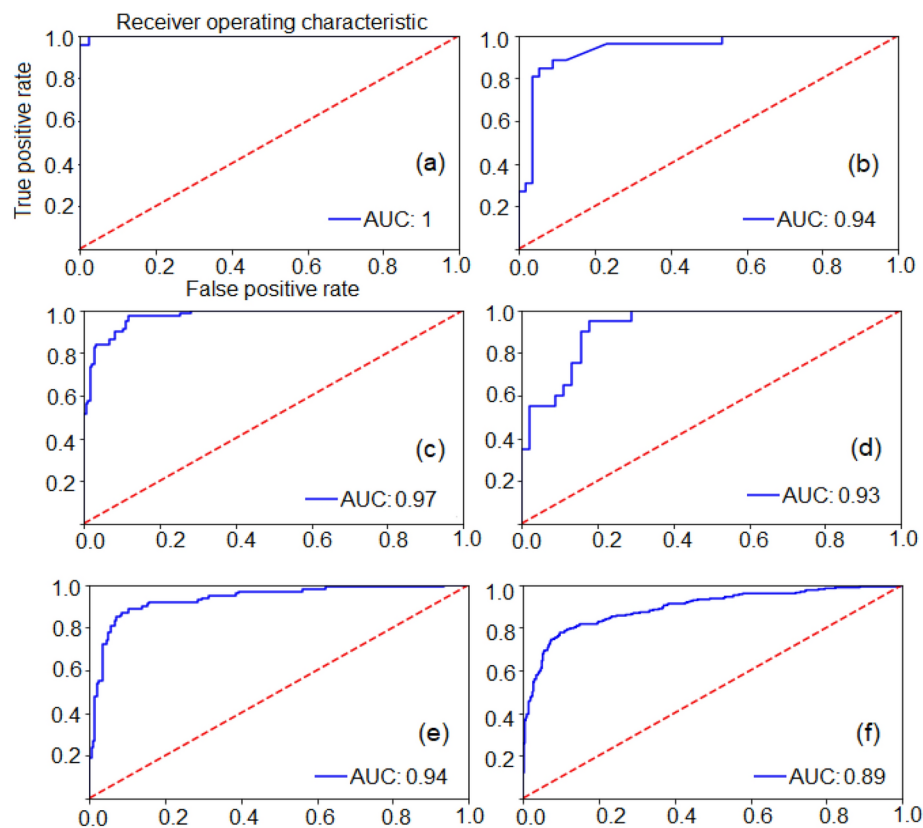
In this sub-section, we analyze the performance of the proposed framework without using the 2D-VMD prior to the deep learning network training. Figure 4 shows the *AUROC* curves obtained with and without 2D-VMD for all datasets used in this work. We also show the *AUPRC* curves in Supplementary Fig. 3. It can be observed that there is a considerable decrease in the areas under these curves when the proposed framework is built without 2D-VMD. Since, the proposed network trains from the raw B-mode images which makes it slightly difficult for the model to extract more effective high level features, the dip in the performance is expected. These results further demonstrate that 2D-VMD can extract more representative discriminatory feature modes which can guide the deep learning network for obtaining enhanced feature maps, resulting in improved classification performance.

### Comparison with state-of-the-art methods

The proposed framework was analyzed with respect to some of the commonly existing methods for breast lesion classification in this sub-section. We present the comparative evaluation results for the two public datasets followed by that for the in clinic data. Table 2 shows the lesion classification accuracies and other metrics reported by some of the existing works for the two public datasets used in this work. It can be observed that for the *UDIAT* dataset, the proposed framework outperformed the existing methods in terms of classifying the lesions, with a considerable increase in all metrics. For the *BUSI* data, the proposed framework performed comparable classification with respect to a few existing methods and outperformed rest of the methods. Since most of the methods reported accuracy, we further estimated the improvement in accuracy of the proposed method with respect to these methods. For the existing dataset *UDIAT*, the proposed method achieved improvements in the range of 4 – 32% relative to these existing state-of-the-art methods, with a mean improvement (standard deviation) of 14.37%(8.42%). While for *BUSI*, the proposed method was comparable to two existing state-of-the-art methods based on hybrid of Alexnet, MobilenetV2, Resnet50<sup>30</sup> and spatial attention with Resnet18<sup>40</sup>, and achieved improvements in the range of 2 – 13% relative to the remaining existing state-of-the-art methods, with a mean improvement (standard deviation) of 5.75%(4.52%).

For the in-clinic data, we implemented some of the commonly used deep learning models in<sup>27–29,36,39</sup> among others to assess the proposed framework, and the corresponding accuracies obtained are presented in Fig. 5. These models are selected based on their wide usage and ease of implementation. It is clearly observable from the figure that the proposed framework outperforms these existing models for the *IHR* data. We further estimate the total number of model parameters and the floating point operations (FLOPS) required by each model for analyzing the corresponding computational complexities. Figure 6 presents these values for all implemented models and the proposed model. It is evident from the figure that the proposed model achieves the classification with a considerable reduction in the parameters and FLOPS with respect to existing methods. We further

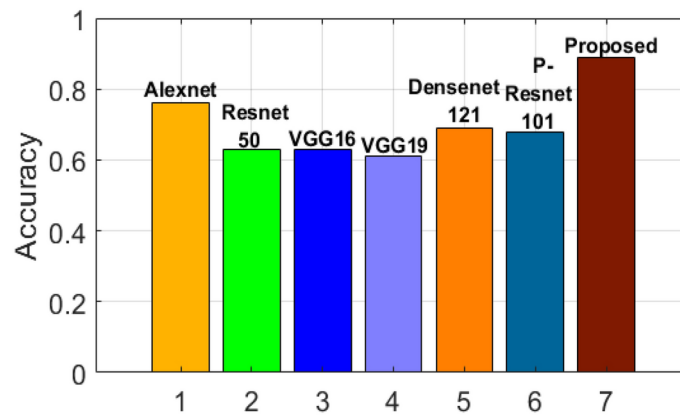




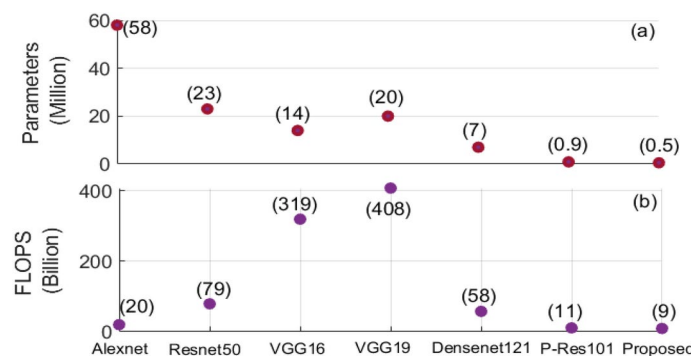
**Fig. 4.** AUROC curves for: (a) and (b) *UDIAT* dataset with and without using 2D-VMD respectively, (c) and (d) *BUSI* dataset with and without using 2D-VMD respectively, (e) and (f) *IHR* dataset with and without using 2D-VMD respectively.

Dataset	Method (Ref.)	Ac	Pr	Se	Sp	F1	AUROC
UDIAT	VGG19 <sup>28</sup>	0.94	-	0.91	0.94	-	-
	Pre-trained Resnet101 <sup>39</sup>	0.85	-	-	-	0.83	0.91
	Modified VGG16 with attention mechanism <sup>41</sup>	0.93	0.92	<b>0.96</b>	0.90	0.94	-
	Hierarchical dense feature aggregation network <sup>58</sup>	0.90	0.91	0.79	0.96	0.84	0.87
	Dual-branch learning network <sup>60</sup>	0.82	0.70	-	0.84	0.72	0.80
	Cross-task guided network <sup>61</sup>	0.86	0.75	0.87	0.85	0.80	0.86
	Hybrid learning <sup>62</sup>	0.85	-	0.80	0.90	0.80	0.87
	Vision Transformer <sup>65</sup>	0.74	-	-	-	-	0.80
	Proposed	<b>0.98</b>	<b>1.00</b>	<b>0.96</b>	<b>1.00</b>	<b>0.98</b>	<b>1.00</b>
BUSI	Hybrid of Alexnet, MobilenetV2, Resnet50 <sup>30</sup>	<b>0.93</b>	-	-	-	-	-
	Spatial attention with Resnet18 <sup>40</sup>	<b>0.94</b>	-	-	-	-	-
	Pre-trained Resnet101 <sup>39</sup>	0.85	-	-	-	0.83	0.91
	Hierarchical dense feature aggregation network <sup>58</sup>	0.91	0.84	0.84	<b>0.94</b>	0.84	0.92
	Dual global attention neural network <sup>59</sup>	-	0.83	0.85	-	-	-
	Dual-branch learning network <sup>60</sup>	0.84	0.73	-	0.87	0.76	0.83
	Cross-task guided network <sup>61</sup>	0.90	0.85	0.84	0.93	0.85	0.89
	Hybrid learning <sup>62</sup>	0.87	-	0.83	0.92	0.83	0.91
	CNN-transformer <sup>63</sup>	0.86	0.86	0.86	0.85	0.86	-
	MaskFormer <sup>64</sup>	0.89	0.94	0.86	0.93	<b>0.89</b>	0.78
	Vision Transformer <sup>65</sup>	0.82	-	-	-	-	0.91
	Proposed	<b>0.93</b>	<b>0.86</b>	<b>0.92</b>	0.93	<b>0.89</b>	<b>0.97</b>

**Table 2.** Comparative analysis of state-of-the-art methods with respect to the proposed method for breast lesion classification. Significant value are in bold.



**Fig. 5.** Comparison results of the proposed framework with some common methods for breast lesion classification for *IHR* data. Note: P denotes pre-trained.



**Fig. 6.** Computational complexities of the proposed framework and some common methods for breast lesion classification for *IHR* data: (a) Total number of model parameters, (b) FLOPS. Note: P denotes pre-trained.

estimated that the proposed method achieved a reduction of 44 – 99% in parameters and 18 – 97% in FLOPS as compared to existing methods. These results demonstrate the superiority of the proposed framework with respect to the existing deep learning networks in terms of performance as well as computational complexity.

### Interpretability of the proposed network

To further analyze the proposed network, we extract the feature maps of the intermediate and final layers for two example cases of benign and malignant lesions. Supplementary Fig. 4 demonstrates these maps for some intermediate layers and the final layer. It can be observed that the average pooling focuses on the general features of the depth-wise separable convolution output while max pooling focuses on the localised features as indicated by slightly higher contrast and activations (darker areas). Mixed-pooling output has a contrast between the average and max pooling outputs. The fourth depth-wise separable convolution output has lower activations (lighter areas) for the lesion area and some background pixels. It shows the differences at pixel-level for lesion and the surrounding areas. Further, the weighted sum mixed pooling output demonstrates a higher contrast and activations for lesion area and some surroundings. It can also be observed that the preceding max pooling output is weighted more as indicated by higher activations with respect to the average pooling output. The activations for benign lesion are observed to be lower as compared to malignant lesion. The squeeze-and-excite output slightly refines the activations which are flattened to a vector. Finally it can be observed that the amplitudes of the samples of the flattened vector are small for benign and higher for malignant cases. This output is converted to probability values of “ $(1.22 \times 10^{-33})$ ” for benign and “1” for malignant, by the sigmoid activation. This demonstrates the explicability of the proposed network layers to classify the lesions.

### Discussion

The proposed framework is rigorously evaluated by using two different public datasets and a private dataset without any prior processing. The corresponding performance metrics highlight its effectiveness across all three datasets. While performance on the *IHR* dataset is marginally lower compared to the other two datasets, it remains robust. This indicates that our framework maintains a reliable performance, even on the data without any prior pre-processing. It can be observed that the framework is able to achieve a high specificity with a minimal trade-off in the sensitivity, as indicated by the *AUPRC* values. This shows promising results of using

deep learning with variational modes that can potentially guide radiologists in identifying pathologies with more accuracy. We also demonstrated the confusion matrices for all datasets for a thorough analysis and performance visualization of the proposed framework. The indication of low false positives for *UDIAT* and *IHR* data from these matrices demonstrates the capability of the framework toward reducing benign biopsies. Further, since the benign lesions are high in number as compared to malignant lesions, the trade off between the false positives and false negatives is justifiable. The high variability in sensitivity and *F1* score across the datasets attributes to the corresponding high variability in the number of benign and malignant lesions as well as total images across the datasets. Alternatively, since the number of images were highly dispersed across the datasets along with some inconsistencies such as noisy images in *IHR* versus the public datasets, there were performance fluctuations across the datasets. However, the *AUROC* values were consistent across the datasets.

This study presents another exhaustive analysis of the proposed deep learning network with and without using 2D-VMD in Fig. 4 and Supplementary Fig. 3. A higher steepness is observed in the ascent towards the left upper corner in Fig. 4 for all datasets in the presence of 2D-VMD as compared to that in its absence, which affirms an excellent trade-off between sensitivity and specificity with 2D-VMD. This is essential in clinical practice, where missing a diagnosis is much more unacceptable than false positive. Supplementary Fig. 3 shows the *AURPC*, which is a more informative parameter in situations of class imbalance compared to the *AUROC* curve, and thereby it could further assert the model's precision in identifying true positive cases. It can be observed that it also showed a high precision across a range of real levels with 2D-VMD as compared to that without it, which is another important factor in a clinical context where a false positive may lead to unnecessary invasive or expensive procedures. Thereby, the effectiveness of 2D-VMD in improving the performance of deep learning model is clearly evident from these results. Finally, we compared our proposed framework with some of the common DL networks reported in recent studies. Firstly, we selected the corresponding existing studies utilizing the two public datasets *UDIAT* and *BUSI*, and presented the classification accuracies reported by these works in Table 2. It is assertive that the proposed framework performs significantly better on *UDIAT* data, and marginally closer to that for the *BUSI* data. Further, we implemented some of the common DL models for *IHR* after experimentally selecting them based on their adaptations in literary works. The corresponding results in Fig. 2 and 6 exhibit the capability of our framework in surpassing the existing models in terms of classification performance and computational complexity.

The integration of 2D-VMD with DL for breast lesion classification can be a unique contribution to the field of ultrasound image classification, unlike most of the recent related studies. For example, the ensemble convolutional neural network (CNN) employed in the study of W. K. Moon et al. in<sup>27</sup>, has demonstrated the effectiveness of image fusion in classification tasks, however, the image decomposition techniques still remain unexplored. Furthermore, the weakly supervised approach in the model developed by J. Kim et al. in<sup>26</sup>, leverages class activation mapping to alleviate the need for detailed annotations, however, it failed to address the potential of signal processing methods to further refine the discriminatory features. The multi-scale kernel approach in the model developed by X. Qi et al. in<sup>34</sup>, highlighted the potential benefits of architectural innovation in complementary neural networks, however, our study demonstrates on the other hand that the decomposition-based feature enhancement can simplify the architecture of the DL models. Further, in comparison with the multi-view learning strategy introduced in<sup>42</sup>, our study underscores the substantial impact of pre-processing strategies on model performance. The utilization of 2D-VMD has led to a perceptible increase in both *AUROC* and *AUPRC* values, which suggests that the approach utilized in the proposed study, successfully captures nuanced features in ultrasound images unlike the traditional DL models. The methodical divergence illustrates the multifaceted approaches for the collective goal of advancing diagnostic accuracy in medical imaging, where each study might contribute to a gathered understanding and improvement of ML application in clinical settings.

The proposed DL model is less complex with respect to the existing traditional models in terms of the total number of layers, parameters, and operations. An expected outcome is that it can facilitate a quicker training process, and moreover, enhance the model's feasibility for real-time applications. A potential clinical impact of this outcome is that it prioritizes timely diagnosis of malignant lesions. Further, our dual-stage approach, which starts with discriminative low-level feature extraction using 2D-VMD, followed by a high-level feature extraction and classification, demonstrates the ability of the proposed framework to process breast ultrasound images efficiently and with lower computational demands. This can deliver high performance without extensive computational and time complexity, which are typically associated with DL models. Another expected outcome of our approach is that it can detect the benign lesions more precisely with a potential clinical impact of minimizing the unnecessary biopsies for benign lesions. Finally, the activation maps demonstrate the interpretability of the proposed network unlike the existing networks which lack in explainability, towards classification outcomes. In the future direction, we plan to incorporate a larger private dataset for validating our proposed framework. In addition to that, we plan to test the proposed framework in real time to test its efficacy in clinical settings. Further, the proposed method can be adapted to other imaging modalities, such as, 2D magnetic resonance imaging (MRI), computed tomography (CT), and mammography. For these adaptations, slight parameter modifications may need to be done for the VMD and the deep learning network depending upon the initial experimental assessment.

## Conclusion

In this study, a novel variational mode directed deep learning framework is proposed for ultrasound-based breast lesion classification. An exhaustive evaluation using two public datasets and an in-clinic dataset demonstrates the proficiency of the proposed framework in classification of breast lesions. The utilization of 2D-VMD with a low complex deep learning model demonstrates a considerable improvement in the overall classification performance by enhancing the discriminatory features obtained in the variational modes. The comparative evaluation of the proposed framework with respect to some of the reported networks for public datasets indicate

the supremacy of the proposed framework. Further implementations of the recently used networks on the in-clinic data demonstrate the superiority of the proposed framework. Finally, the low complexity of the DL model with only a few convolutional layers demonstrates its suitability for real-time breast lesion classification.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The requested data may include figures that have associated raw data. Because the study was conducted on human volunteers, the release of patient data may be restricted by Mayo policy and needs special request. The request can be sent to: Karen A. Hartman, MSN, CHRC | Administrator - Research Compliance | Integrity and Compliance Office | Assistant Professor of Health Care Administration, Mayo Clinic College of Medicine & Science | 507-538-5238 | Administrative Assistant: 507-266-6286 | hartman.karen@mayo.edu Mayo Clinic | 200 First Street SW | Rochester, MN 55905 | mayoclinic.org. We do not have publicly available accession codes, unique identifiers, or web links.

Received: 1 November 2024; Accepted: 16 April 2025

Published online: 24 April 2025

## References

1. Siegel, R. L. et al. Cancer statistics 2024. *CA A Cancer J. Clin.* **74**(1), 12–49 (2024).
2. Sood, R. et al. Ultrasound for breast cancer detection globally: A systematic review and meta-analysis. *J. Glob. Oncol.* **5**(5), 1–17 (2019).
3. Yuan, W. H. et al. Supplemental breast cancer-screening ultrasonography in women with dense breasts: A systematic review and meta-analysis. *Br. J. Cancer* **123**(4), 673–688 (2020).
4. Rebolj, M. et al. Addition of ultrasound to mammography in the case of dense breast tissue: Systematic review and meta-analysis. *Br. J. Cancer* **118**(12), 1559–1570 (2018).
5. Lee, J. et al. Evaluation of the effect of computer aided diagnosis system on breast ultrasound for inexperienced radiologists in describing and determining breast lesions. *Med. Ultrason.* **21**(3), 239–245 (2019).
6. Nicosia, L. et al. Evaluation of computer-aided diagnosis in breast ultrasonography: Improvement in diagnostic performance of inexperienced radiologists. *Clin. Imaging* **82**, 150–155 (2022).
7. Ramadan, S. Z. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: A review. *J. Healthc. Eng.* **2020** (2020).
8. Shah, D. et al. Enhancing the quality and authenticity of synthetic mammogram images for improved breast cancer detection. *IEEE Access* **12**, 12189–12198 (2024).
9. Shah, D. et al. Reliable breast cancer diagnosis with deep learning: DCGAN-driven mammogram synthesis and validity assessment. *Appl. Comput. Intell. Soft Comput.* **2024**, 1122109 (2024).
10. Shah, D. et al. Optimizing breast cancer detection with an ensemble deep learning approach. *Int. J. Intell. Syst.* **2024**, 5564649 (2024).
11. Shah, D. et al. Dual-view deep learning model for accurate breast cancer detection in mammograms. *Int. J. Intell. Syst.* **2025**, 7638868 (2025).
12. Fujioka, T. et al. The utility of deep learning in breast ultrasonic imaging: A review. *Diagnostics* **10**(12), 1055 (2020).
13. Berg, W. A. et al. Toward AI-supported US triage of women with palpable breast lumps in a low-resource setting. *Radiology* **307**(4), e223351 (2023).
14. Kumar, V. et al. Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PLoS One* **13**(5), e0195816 (2018).
15. Webb, J. M. et al. Comparing deep learning-based automatic segmentation of breast masses to expert interobserver variability in ultrasound imaging. *Comput. Biol. Med.* **139**, 104966 (2021).
16. Huang, Q. et al. Machine learning in ultrasound computer-aided diagnostic systems: A survey. *BioMed. Res. Int.* **2018** (2018).
17. Kumar, V. et al. Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning. *IEEE Access* **8**, 63482–63496 (2020).
18. Webb, J. M. et al. Automatic deep learning semantic segmentation of ultrasound thyroid cineclips using recurrent fully convolutional networks. *IEEE Access* **9**, 5119–5127 (2021).
19. Saini, M., Fatemi, M. & Alizad, A. Fast inter-frame motion correction in contrast-free ultrasound quantitative microvasculature imaging using deep learning. *Sci. Rep.* **14**(1), 26161 (2024).
20. Kim, S. Y. et al. Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses. *Sci. Rep.* **11**(1), 395 (2021).
21. Wang, Y. et al. The clinical application of artificial intelligence assisted contrast-enhanced ultrasound on BI-RADS category 4 breast lesions. *Acad. Radiol.* **30**, S104–S113 (2023).
22. Zhang, X. et al. Deep learning-based radiomics of B-Mode ultrasonography and shear-wave elastography: Improved performance in breast mass classification. *Front. Oncol.* **10**, 1621 (2020).
23. Han, S. et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys. Med. Biol.* **62**(19), 7714–7728 (2017).
24. Fujioka, T. et al. Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. *Jpn. J. Radiol.* **37**(6), 466–472 (2019).
25. Fujioka, T. et al. Classification of breast masses on ultrasound shear wave elastography using convolutional neural networks. *Ultrason. Imaging* **42**(4–5), 213–220 (2020).
26. Kim, J. et al. Weakly-supervised deep learning for ultrasound diagnosis of breast cancer. *Sci. Rep.* **11** (2021).
27. Moon, W. K. et al. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Progr. Biomed.* **190**, 105361 (2020).
28. Daoud, et al. Breast tumor classification in ultrasound images using combined deep and handcrafted features. *Sensors* **20**(23), 6838 (2020).
29. Zhang, N. et al. Application of deep learning to establish a diagnostic model of breast lesions using two-dimensional grayscale ultrasound imaging. *Clin. Imaging* **79**, 56–63 (2021).
30. Eroğlu, Y. et al. Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. *Comput. Biol. Med.* **133**, 104407 (2021).
31. Zhao, Z. et al. Application of deep learning to reduce the rate of malignancy among BI-RADS 4A breast lesions based on ultrasonography. *Ultrason. Med. Biol.* **48**, 2267–2275 (2022).
32. Gu, Y. et al. Ultrasound-based deep learning in the establishment of a breast lesion risk stratification system: A multicenter study. *Eur. Radiol.* **33**, 2954–2964 (2023).

33. Huang, Y. et al. Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. *Biomed. Eng. Online* **18**(1), 8 (2019).
34. Qi, X. et al. Automated diagnosis of breast ultrasonography images using deep neural networks. *Med. Image Anal.* **52**, 185–198 (2019).
35. Ragab, M. et al. Ensemble deep-learning-enabled clinical decision support system for breast cancer diagnosis and classification on ultrasound images. *Biology* **11**, 439 (2022).
36. Shia, W. C. & Chen, D. R. Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine. *Comput. Med. Imaging Graph.* **87**, 101829 (2021).
37. Kaplan, E. et al. Automated BI-RADS classification of lesions using pyramid triple deep feature generator technique on breast ultrasound images. *Med. Eng. Phys.* **108**, 103895 (2022).
38. Jabeen, K. et al. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors* **22**, 807 (2022).
39. Wilding, R. et al. Deep learning applied to breast imaging classification and segmentation with human expert intervention. *J. Ultrasound* **25**, 659–666 (2022).
40. Lu, S. Y. et al. SAFNet: A deep spatial attention network with classifier fusion for breast cancer detection. *Comput. Biol. Med.* **148**, 105812 (2022).
41. Kalafi, E. Y. et al. Classification of breast cancer lesions in ultrasound images by using attention layer and loss ensemble in deep convolutional neural networks. *Diagnostics* **11**, 1859 (2021).
42. Shin, S. Y. et al. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE Trans. Med. Imaging* **38**(3), 762–774 (2019).
43. Kim, H. E. et al. Transfer learning for medical image classification: A literature review. *BMC Med. Imaging* **22**(1), 69 (2022).
44. Hijab, A. et al. Breast cancer classification in ultrasound images using transfer learning. *2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME)*, 1–4 (2019).
45. Konstantin, D. & Zosso, D. Two-dimensional variational mode decomposition. *Energy Minim. Methods Comput. Vis. Pattern Recognit.* **8932**, 197–208 (2015).
46. Saini, M., Satija, U. & Upadhyay, M. D. DSCNN-CAU: Deeplearning- based mental activity classification for IoT implementation toward portable BCI. *IEEE Internet Things J.* **10**(10), 8944–8957 (2023).
47. Saini, M. et al. One-dimensional convolutional neural network architecture for classification of mental tasks from electroencephalogram. *Biomed. Signal Process. Control* **74**, 103494 (2022).
48. Bai, L., Zhao, Y. & Huang, X. A CNN accelerator on FPGA using depthwise separable convolution. *IEEE Trans. Circuits Syst. II Express Briefs* **65**(10), 1415–1419 (2018).
49. Saini, M. et al. BiCurNet: Premovement EEG-Based Neural Decoder for Biceps Curl Trajectory Estimation. *IEEE Trans. Instrum. Meas.* **73**, 1–11 (2024).
50. Saini, M., Afrin, H., Sotoudehnia, S., Fatemi, M. & Alizad, A. DMAeEDNet: Dense multiplicative attention enhanced encoder decoder network for ultrasound-based automated breast lesion segmentation. *IEEE Access* **12**, 60541–60555 (2024).
51. Bai, X. et al. Edge-guided adaptive multi-feature learning for breast ultrasound image segmentation. *Biomed. Signal Process. Control* **91**, 106076 (2024).
52. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132–7141 (2018).
53. Yap, M. H. et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Informat.* **22**(4), 1218–1226 (2018).
54. Al-Dhabyani, W., Gomaa, M., Khaled, H. & Fahmy, A. Dataset of breast ultrasound images. *Data Brief* **28** (2020).
55. Lin, T.-Y. et al. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 318–327 (2020).
56. Saini, M. et al. Discriminatory features based on wavelet energy for effective analysis of electroencephalogram during mental tasks. *Circuits Syst. Signal Process.* **41**(10), 5827–5855 (2022).
57. Hancock, J., Khoshgoftaar, T. M. & Johnson, J. M. Informative evaluation metrics for highly imbalanced big data classification. *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas*, 1419–1426 (2022).
58. Di, X. et al. Saliency map-guided hierarchical dense feature aggregation framework for breast lesion classification using ultrasound image. *Comput. Methods Progr. Biomed.* **215**, 106612 (2022).
59. Meng, H. et al. DGA-Net: a dual global attention neural network for breast lesion detection in ultrasound images. *Ultrasound Med. Biol.* **49**(1), 31–44 (2023).
60. Zhu, C. et al. DBL-Net: A dual-branch learning network with information from spatial and frequency domains for tumor segmentation and classification in breast ultrasound image. *Biomed. Signal Process. Control* **93**, 106221 (2024).
61. Yang, K. et al. CTG-Net: Cross-task guided network for breast ultrasound diagnosis. *PLoS One* **17**(8), e0271106 (2022).
62. Ru, J., Zhu, Z. & Shi, J. Spatial and geometric learning for classification of breast tumors from multi-center ultrasound images: A hybrid learning approach. *BMC Med. Imaging* **24**(1), 133 (2024).
63. Tagnamas, J. et al. Multi-task approach based on combined CNN-transformer for efficient segmentation and classification of breast tumors in ultrasound images. *Visual Comput. Ind. Biomed. Art* **7**(1), 2 (2024).
64. Anand, A., Jung, S. & Lee, S. Breast lesion detection for ultrasound images using MaskFormer. *Sensors* **24**(21), 6890 (2024).
65. Gheflati, B. & Rivaz, H. Vision transformers for classification of breast ultrasound images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE, 2022)*.

## Acknowledgements

Research reported in this publication was supported by National Institute of health grants R01CA239548, R01CA148994, R01CA195527. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The NIH did not have any additional role in the study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors would like to thank Ms. Cindy Andrist, Ms. Patricia O'Neil and Ms. Lesley Riess for their valuable help in patient recruitment.

## Author contributions

A.A. contributed to conceptualization, methodology, investigation, visualization, resources, funding acquisition, supervision, and project administration, reviewing and editing the manuscript; M.F. contributed to conceptualization, methodology, investigation, visualization, resources, funding acquisition, supervision, and project administration, reviewing and editing the manuscript; M.S. contributed to visualization, formal analysis, software, wrote the original draft of manuscript, reviewing and editing the manuscript; S.H. wrote parts of the original draft of manuscript, visualization, figure preparation, data analysis, reviewing and editing the manuscript, B.M. contributed to data curation, data analysis, validation and reviewing editing the manuscript. All authors reviewed the manuscript.



## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-99009-5>.

**Correspondence** and requests for materials should be addressed to A.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025