Check for updates

Deep learning classification of deep ultraviolet fluorescence images toward intra-operative margin assessment in breast cancer

Tyrell To[1], Tongtong Lu[2], Julie M. Jorns[3], Mollie Patton[3], Taly Gilat Schmidt[2], Tina Yen[4], Bing Yu[2] and Dong Hye Ye[5]*

[1]Department of Electrical and Computer Engineering, Marquette University, Opus College of Engineering, Milwaukee, WI, United States, [2]Joint Department of Biomedical Engineering, Marquette University and Medical College of Wisconsin, Milwaukee, WI, United States, [3]Department of Pathology, Medical College of Wisconsin, Milwaukee, WI, United States, [4]Department of Surgery, Medical College of Wisconsin, Milwaukee, WI, United States, [5]Department of Computer Science, Georgia State University, Atlanta, GA, United States

Background: Breast-conserving surgery is aimed at removing all cancerous cells while minimizing the loss of healthy tissue. To ensure a balance between complete resection of cancer and preservation of healthy tissue, it is necessary to assess themargins of the removed specimen during the operation. Deep ultraviolet (DUV) fluorescence scanning microscopy provides rapid whole-surface imaging (WSI) of resected tissues with significant contrast between malignant and normal/benign tissue. Intra-operative margin assessment with DUV images would benefit from an automated breast cancer classification method.

Methods: Deep learning has shown promising results in breast cancer classification, but the limited DUV image dataset presents the challenge of overfitting to train a robust network. To overcome this challenge, the DUV-WSI images are split into small patches, and features are extracted using a pre-trained convolutional neural network—afterward, a gradient-boosting tree trains on these features for patch-level classification. An ensemble learning approach merges patch-level classification results and regional importance to determine the margin status. An explainable artificial intelligence method calculates the regional importance values.

Results: The proposed method's ability to determine the DUV WSI was high with 95% accuracy. The 100% sensitivity shows that the method can detect malignant cases efficiently. The method could also accurately localize areas that contain malignant or normal/benign tissue.

Conclusion: The proposed method outperforms the standard deep learning classification methods on the DUV breast surgical samples. The results suggest that it can be used to improve classification performance and identify cancerous regions more effectively.
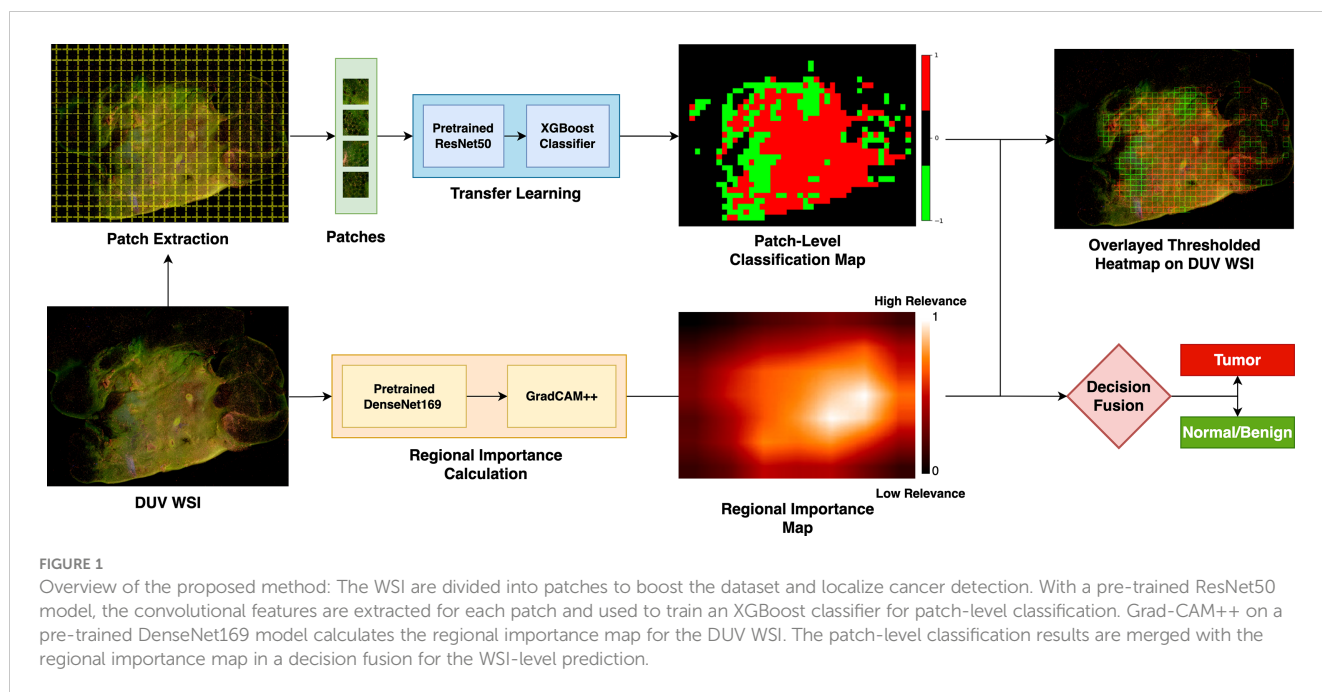
1 Introduction

In the United States, approximately 1 in 8 women will develop breast cancer in their lifetime (1). More than half of women who have surgery for breast cancer undergo breast-conserving surgery (BCS) or a lumpectomy (2, 3). These treatments are designed to remove all tumor tissue while keeping as much normal/benign tissue as possible. Patients may develop cancer recurrence if the cancer cells are not entirely detected and reside on the edges of the specimen tissue (positive margin). With positive margins, patients may experience additional surgery to remove the remaining cancer cells (4).

The evaluation of the intra-operative margin is currently performed by radiographic examination and confirmed by post-operative histological analysis of hematoxylin and eosin (H&E) images. As a preferable alternative, having a microscopic imaging system for margin assessment during surgery is beneficial. Real-time whole-surface imaging (WSI) of resected tissues during BCS can be provided by deep ultraviolet fluorescence scanning microscopy (DUV-FSM) (4, 5). DUV-FSM uses surface ultraviolet excitation without destructive techniques and unnecessary optical sectioning techniques to create high-quality images of specimens. These DUV images help to detect cancer cells at the edge of the surgical specimen as they show excellent color and texture contrast between cancer and normal/benign breast tissue. In previous intra-operative margin assessment research (5), texture analysis (TA) methods were used to detect tumors in ex vivo breast surgical tissues with the DUV dataset, demonstrating excellent sensitivity and specificity. However, limitations exist due to potential bias, errors during dataset construction, and reduced diagnostic accuracy for less common tissue types. Deep learning models offer an alternative solution for automated breast cancer classification of DUV images, as they can automatically learn complex patterns and hierarchical features directly from the data. This approach enables improved generalization and adaptability across diverse tissue types, overcoming the limitations of TA methods and potentially leading to more accurate and robust classification results.

In recent years, the medical domain has explored deep-learning methods for disease detection. Image classification algorithms can aid in detecting and localizing disease regions of interest (ROI). Breast cancer classification and margin assessment are major research areas that have shown promising results (6, 7). For instance (8), classified breast cancer by combining a convolutional graph network with a convolutional neural network (CNN). In their research, a modified 8-layer CNN and a two-layer graph convolutional network were fused to obtain higher classification accuracy (9). classified breast cancer with a patch-level CNN. Their study extracted spatial information of individual patches with integrated CNN and filter algorithms (10). segmented tumor and non-tumor areas with a patch-level CNN for tumor localization, which allowed deeper tissue analysis (11). adopted a patch-level CNN to classify breast cancer subtypes (12). used a patch screening method with a CNN and clustering methods. The use of patches helped differentiate the tumor types from one another (13). used a transfer learning approach to determine breast cancer grades. This study helped determine the severity of specific tumor grades with high accuracy (14). investigated the best features with transfer learning and determined the best features for predicting specific tumor grades. To deal with data imbalance (15), incorporated a transfer learning method in their study to improve classification results by using a VGG-19 model pre-trained on ImageNet weights.

The researchers mentioned above trained and tested the network on a publicly available H&E image dataset. At the moment, there is difficulty in obtaining a large training set of DUV images due to its relative novelty and practical application on a small group of subjects. Well-trained models may generalize poorly on a limited DUV dataset. To prevent the problem of overfitting, the DUV WSI images are divided into small, non-overlapping patches. In this way, the network will be able to learn better generalizations. This patch methodology boosts the training data at the patch level and helps localize cancer regions in DUV WSI for margin evaluation. A transfer learning approach is utilized to extract convolutional features from DUV patches with a pre-trainedconvolutional neural network and input them into decision-tree-based classifiers for DUV patch-level classification. As an ensemble approach, the patch-level classification results are merged with regional importance map calculations using Gradient-weighted Class Activation Mapping++ (Grad-CAM++) (16) into a single binary decision for the WSI level. The

FIGURE 1

Overview of the proposed method: The WSI are divided into patches to boost the dataset and localize cancer detection. With a pre-trained ResNet50 model, the convolutional features are extracted for each patch and used to train an XGBoost classifier for patch-level classification. Grad-CAM++ on a pre-trained DenseNet169 model calculates the regional importance map for the DUV WSI. The patch-level classification results are merged with the regional importance map in a decision fusion for the WSI-level prediction.

performance accuracy of the proposed approach is tested by classifying 60 images at the WSI level.

2 Method

Figure 1 describes the proposed breast cancer classification method for DUV image margin assessment. First, DUV WSIs are divided into patches to enhance the dataset and localize cancerous regions. A pre-trained ResNet50 model (17) extracts convolutional features for each patch, which are then used to train an XGBoost classifier (18) for patch-level classification. The regional importance map for the DUV WSI is calculated using Grad-CAM++ (16) on a pre-trained DenseNet169 model (19). Finally, the patch-level classification results merge with the regional importance map in a decision fusion for the WSI-level prediction.

2.1 Patch extraction

A DUV WSI is divided into multiple DUV patches. The x_i is denoted as a DUV WSI for a sample i. A 2D grid system is designed to be based on each sample's field of view (FOV) Ω_i with non-overlapping patches Ω_i^j,

$$\Omega_i = \cup_{j=1}^{N} \Omega_i^j, \text{and} \quad \Omega_i^k \cap \Omega_i^l = \varnothing \quad \text{for} \quad \forall\, k, l, \quad (1)$$

where N represents the total amount of patches in x_i, k and l represent any patch indices.

Since each DUV WSI has different dimensions, the images are resized to the closest dimensions divisible into non-overlapping patches Ω_i^j with a constraint size of 400×400 pixels.

The minimal alterations to the dimensions should not affect the quality of the morphological characteristics, such as cell density and

infiltration, because the DUV WSI images are very large. Afterward, a 2D grid system is implemented based on each sample's FOV Ω_i with non-overlapping cells Ω_i^j.

The patch images are extracted from each non-overlapping patch Ω_i^j of a DUV WSI. To determine the validity of a patch, it is converted to grayscale, and its pixels are analyzed. If most of the pixels in the patch Ω_i^j are at least 80% foreground, then it is counted as a valid patch p_i^j. A pixel is a foreground if its grayscale value is greater than 5. It is important to remove the dark background to help with the localization and detection of breast cancer. This process is shown in Algorithm 1 *ExtractPatches*, and Algorithm 2 *GetPercentage*.

```
x_i ← DUV WSI sample i
Ω_i ← 2D grid system on each sample's FOV
Ω_i^j ← Non-overlapping patch with 400 × 400 pixels
p_i^j ← valid patch image
M ← number of DUV WSI samples
for i < M do
    Load in sample x_i
    for Ω_i^j in Ω_i do
        if GetPercentage(Ω_i^j) ≥ 0.2 then
            Save Ω_i^j as valid patch p_i^j for Ω_i
        end if
    end for
end for
```

ALGORITHM 1

ExtractPatches is used to extract foreground patches for each DUV WSI

```
BackgroundThreshold ← 5
ImageSize ← 400 × 400 pixels
```

```
PixelCounter ← 0
for x < 400 do
  for y < 400 do
    if Image[x][y] ≥ BackgroundThreshold then
      PixelCounter ← PixelCounter → 1
    end if
  end for
end for
  return PixelCounter / ImageSize
```

ALGORITHM 2
GetPercentage is used to determine the foreground percentage of a patch

2.2 Patch classification with transfer learning

Convolutional neural networks are widely used for image classification problems and have the potential to diagnose diseases. In architectures like ResNet, the 'vanishing gradient' problem exists. The architecture of a standard ResNet50 is visualized in Figure 2. This happens as the increased number of layers can cause information loss of certain/specific information, preventing the network from training well. Information loss is a critical problem to overcome, especially with training with the limited DUV dataset. Even though the gradient of ResNets can transfer directly to the identity function from the previous layers to the forward layers, the summation of the identity function and output H can nullify the information flow in the network. A robust model architecture is needed to prevent overfitting for a small BCS dataset.

For this issue, a transfer learning approach determines whether each patch is malignant or normal/benign for all N DUV patches $p_i^j, j = \{1, ..., N\}$. In this approach, the features are extracted from the final layer of a pre-trained ResNet50 network (17) on the ImageNet dataset and fed into an XGBoost classifier (18). The XGboost classifier predicts the binary output $y_i^j \in \{-1, +1\}$ as malignant or normal/benign for each patch p_i^j. This determines the tumor ROI in the DUV WSI based on y_i^j with the relative patch locations (Refer to Figure 1).

2.3 Regional importance calculation with Grad-CAM++

Grad-CAM++ (20) is an explainable artificial intelligence approach that generates visual explanations behind a model's decisions. When applied to deep neural networks, it can visualize gradients with pixel-wise weighted feature maps. This technique explains the output layer decisions while considering the spatial information and high-level semantics from the previous convolutional layers. Since the details in the patches and the DUV images are very complex, it is crucial to implement Grad-CAM++ on a network that can retain as much information as it can. A proposed network called DenseNet (Densely Connected Convolutional Network) (19) is a modified version of ResNet where each layer is connected directly with every other layer.
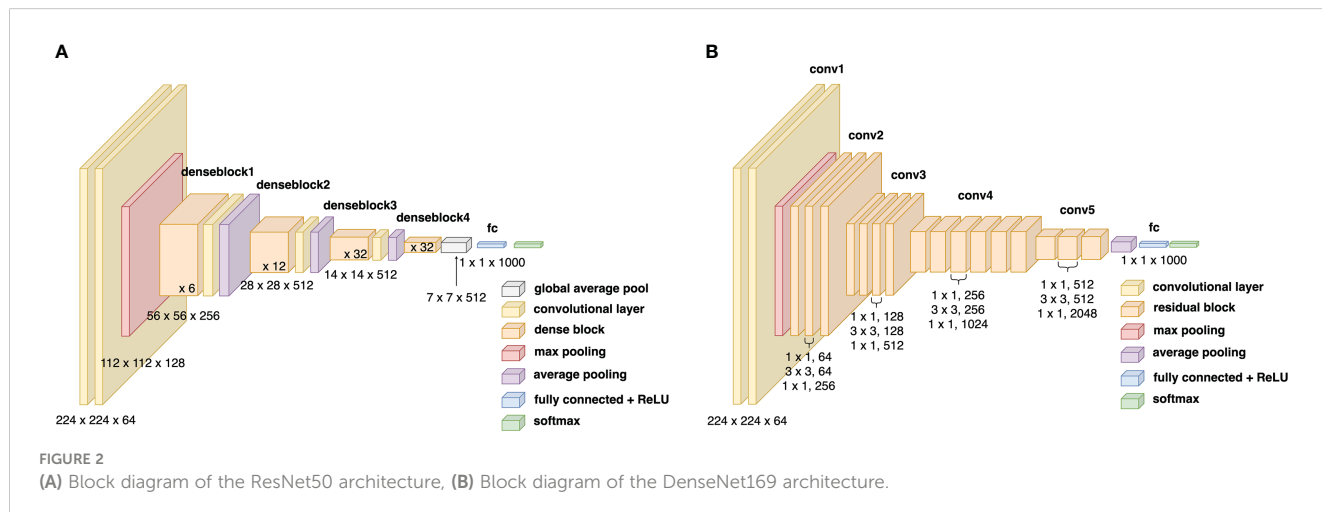
As shown in this equation, the *lth* layer receives the feature-maps of all preceding layers, $x_0, ..., x_{(l-1)}$, as input,

$$x_l = H_l([x_0, ..., x_{(l-1)}]), \qquad (2)$$

where $[x_0, ..., x_{(l-1)}]$ refers to the concatenation of the feature maps produced in layers 0 to $l - 1$.

The novelty of DenseNet is that it removes repeated feature maps after the concatenation of the feature maps, allowing for fewer parameters. By concatenating these feature maps from different feature maps, it increases the variation in the inputs of the following layers. The bottleneck and compression layers of the DenseNet architecture are effective against overfitting due to fewer parameters needed. The bottleneck layers help reduce the number of inputs from previous k-output feature maps, which improves computational efficiency. For the compression layers, they improve model compactness by reducing the feature maps that a dense block generates. The feature maps will overlap at the last output layer, highlighting the most relevant features, which is helpful for Grad-CAM++.

The following equations describe Grad-CAM++'s process of extracting weight maps from feature maps. The final classification score is defined as Y^c for class c as a linear combination of global average pooled convolutional feature maps A^k in the last layer over the FOV Ω_i:

FIGURE 2
(A) Block diagram of the ResNet50 architecture, **(B)** Block diagram of the DenseNet169 architecture.

$$Y^c = \sum_k w_k^c \cdot \sum_{l \in \Omega_i} A_l^k \qquad (3)$$

The gradient of the classification score Y^c for class c before the softmax layer with respect to the final convolutional layer feature map activation A^k is defined as $\frac{\partial Y^c}{\partial A^k}$. The weights w_k^c for a particular feature map A^k are defined by its global average pooled gradients:

$$w_k^c = \frac{1}{Z} \sum_{l \in \Omega_i} \frac{\partial Y^c}{\partial A_l^k}, \qquad (4)$$

where Z represents the activation map's number of pixels.

With Grad-CAM implementations, visualizations are limited if there are multiple instances of a class in the input image x_i, as different spatial footprints of classes can cause different feature maps. The feature maps with small footprints will not be seen in the final saliency map.

To fix this issue, a weighted average of the pixel-wise gradients can be taken instead, where Equation 3 is restructured as:

$$w_k^c = \sum_{l \in \Omega_i} \alpha_l^{kc} \cdot ReLU\left(\frac{\partial Y^c}{\partial A_l^k}\right), \qquad (5)$$

where $ReLU$ is the rectified linear unit activation function and α_l^{kc} correspond to the pixel-wise gradients for class c and convolutional feature map A^k.

The gradient weights α_l^{kc} can be derived for a particular class c and activation map A^k. Combining Equation 2 and Equation 3:

$$Y^c = \sum_k \left\{ \sum_{l \in \Omega_i} \alpha_l^{kc} \cdot ReLU\left(\frac{\partial Y^c}{\partial A_l^k}\right) \right\} \left[\sum_{l \in \Omega_i} A_l^k \right] \qquad (6)$$

Without a threshold need, the $ReLU$ can be removed from the derivation since there is no loss of generality. Hence:

$$\frac{\partial^2 Y^c}{\partial (A_l^k)^2} = 2 \cdot \alpha_l^{kc} \cdot \frac{\partial^2 Y^c}{\partial (A_l^k)^2} + \sum_{l \in \Omega_i} A_l^k \left\{ \alpha_l^{kc} \cdot \frac{\partial^3 Y^c}{\partial (A_l^k)^3} \right\} \qquad (7)$$

To isolate α_l^{kc}, Equation 6 is rearranged:

$$\alpha_l^{kc} = \frac{\frac{\partial^2 Y^c}{(\partial A_l^k)^2}}{2 \frac{\partial^2 Y^c}{\partial (A_l^k)^2} + \sum_{l \in \Omega_i} A_l^k \frac{\partial^3 Y^c}{(\partial A_l^k)^3}} \qquad (8)$$

The Grad-CAM++ weights are calculated by substituting Equation 7 in Equation 4:

$$w_k^c = \sum_{l \in \Omega_i} \left[\frac{\frac{\partial^2 Y^c}{(\partial A_l^k)^2}}{2 \frac{\partial^2 Y^c}{(\partial A_l^k)^2} + \sum_{l \in \Omega_i} A_l^k \frac{\partial^3 Y^c}{(\partial A_l^k)^3}} \right] \cdot relu\left(\frac{\partial Y^c}{\partial A_l^k}\right) \qquad (9)$$

The regional importance map R^i for a DUV WSI x_i is computed using a linear combination of forward activation maps.

$$R^i = ReLU\left(\sum_k w_k^c \cdot A_l^k \right), \qquad (10)$$

This highlights the most significant features in the final classification with a positive correlation with pixel intensity and classification score with the application of the $ReLU$ function to a linear combination of activation maps. In this study, the GradCAM

++ implementation is applied on a pre-trained DenseNet169 model with ImageNet weights and extracts the feature map as the regional importance map at the Norm5 layer. The architecture of the DenseNet169 model can be visualized in Figure 2.

2.4 WSI classification with decision fusion

Given the patch-level classification labels $y_i^j \in \{+1, -1\}$ for all patches $j = \{1, ..., M_i\}$ and regional importance map R^i, a decision fusion method is applied to determine the WSI-level classification label $y_i \in \{+1, -1\}$.

First, the regional importance r_i^j is computed for each patch p_i^j by taking the average value of R^i over a patch's FOV Ω_i^j.

$$r_i^j = \frac{1}{|\Omega_i^j|} \sum_{l \in \Omega_i^j} R_l^i, \qquad (11)$$

where $|\Omega_i^j|$ is the number of pixels for each patch (400×400 pixels).

The weight is defined as w_i^j for each patch p_i^j based on the thresholded regional importance value r_i^j.

$$w_i^j = \begin{cases} 0 & \text{if } r_i^j < 0.25 \\ r_i^j & \text{otherwise} \end{cases} \qquad (12)$$

This weighting scheme ignores the patches with low importance, either malignant or normal/benign, in the fused decision for the DUV WSI.

Next, the patch-level classification label $y_i^j \in \{+1, -1\}$ and weight w_i^j are multiplied for each patch as u_i^j.

$$u_i^j = w_i^j \cdot y_i^j, \qquad (13)$$

Now, the total amount of malignant patches H_i is calculated,

$$H_i = \sum_{j=1}^{} \begin{cases} 1 & \text{if } u_i^j > 0, \\ 0 & \text{otherwise} \end{cases} \qquad (14)$$

Finally, the WSI-level classification label $y_i \in \{-1, +1\}$ is determined by comparing H_i to a certain percentage q of the total foreground patches M_i.

$$y_i = \{ +1 \text{ if } H_i > q \cdot M_i, 0 \le q \le 1 - 1 \text{ otherwise} \qquad (15)$$

where it maps the positive (malignant) and negative (benign) value to $+1$ and -1, respectively.

3 Results

Automated DUV WSI classification is used to determine if a sample is malignant or normal/benign tissue. The proposed method was assessed with 5-fold cross-validation. The DUV WSI data was split while managing balanced class labels and specific patient samples to stay only in training or testing data splits. There was no fine-tuning of the pre-trained ResNet50 with ImageNet weights for the Transfer Learning part. The XGBoost classifier's

hyperparameters were default settings and did not involve any tinkering. There was no hyperparameter tuning for the pre-trained DenseNet169 model with ImageNet Weights for the Regional Importance Calculation. The proposed approach was compared with a standard ResNet50 model and Patch Classification with Majority Voting. The ResNet50 (17) trained for 100 epochs on the limited DUV WSI data with a batch size of 4, a learning rate of 0.006, and a dropout of 40%. Hyperparameter tuning has been done for the ResNet50 model. The Patch Classification with Majority Voting is derived from the architecture of the proposed method's patch classification and transfer learning portion. This approach is augmented with a patch majority voting scheme for binary classification between malignant and normal/benign for the WSIs.

3.1 Dataset

The breast cancer dataset consists of DUV images from 60 samples (24 normal/benign and 36 malignant). This DUV dataset was collected from the Medical College of Wisconsin (MCW) tissue bank (4) with a custom DUV-FSM system. The DUV-FSM used a deep ultraviolet (DUV) excitation at 285 nm and a low magnification objective (4X), which achieved a small spatial resolution from 2 to 3 μ m. To enhance fluorescence contrast, breast tissues are stained with propidium iodide and eosin Y. This technique produces images of the microscopic resolution, sharpness, and contrast from fresh tissue stained with multiple fluorescence dyes.

The 60 DUV images were divided into 34468 patches with a size of 400×400 pixels at 4X magnification (9444 malignant and 25024 normal/benign samples). Moreover, when the classifiers were trained, horizontal/vertical flips and 90-degree rotations were used to boost the data. The pathologists annotated and delineated tumors from the corresponding H&E images for ground-truth labels. The DUV images are registered and compared with H&E images manually. Table 1 shows the dataset's distribution of malignant and normal/benign images in extracted patches, and DUV WSIs.

3.2 Visual inspection

The breast cancer location and detection results on several malignant and normal/benign DUV images for qualitative evaluation are in the last column of Figure 3. The Grad-CAM++ influenced results are shown as the red (malignant) and green (benign) patches with high importance. The H&E images annotated by the pathologists are displayed for comparison with their DUV counterparts. In comparison, the DUV images have a higher color contrast than the H&E images when analyzing the malignant (pink/yellow) and normal/benign (light/dark green) tissues.

As seen in these DUV WSI images, the proposed method was able to accurately localize malignant and normal/benign areas with the aid of Grad-CAM++ regional importance maps. The DUV images contain overlaid regions focusing on malignant (red

| | Annotated H&E Image | DUV WSI | Regional Importance | DUV WSI w/ Overlay |
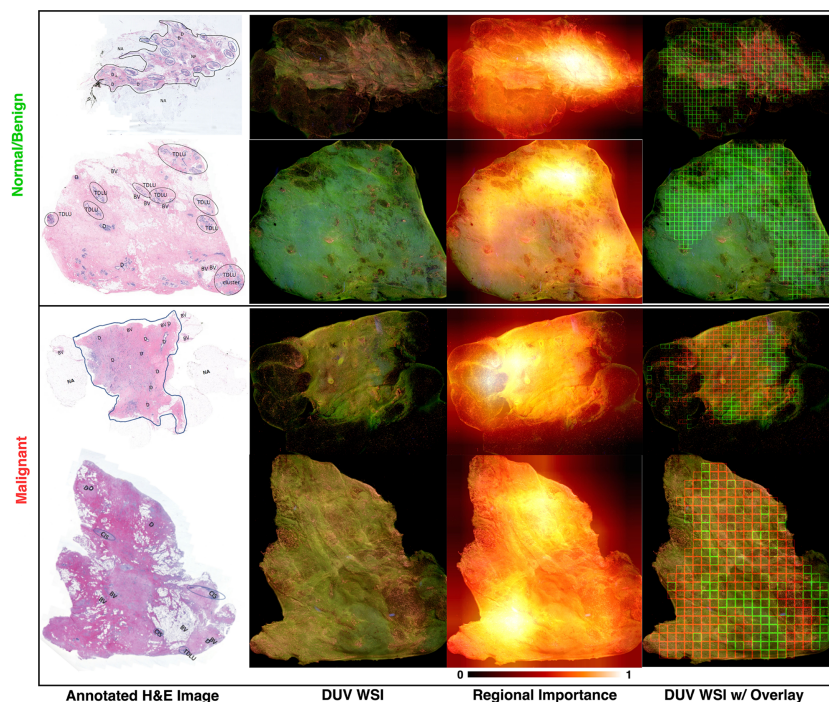
FIGURE 3

DUV WSI with corresponding H&E image and regional importance map for malignant and normal/benign samples: Regional importance areas contain high entropy, high content/semantics (3rd column). It focuses on the strong semantic areas to help in the classification. Note that the images are inputs into a pre-trained DenseNet169 with ImageNet weights, and Grad-CAM++ extracted the heatmaps where the relevancy of the tissue areas is determined from 0 to 1. Red and green bounding boxes in DUV WSIs (last column) represent malignant and normal/benign patches thresholded by regional importance, respectively. Large DUV WSIs are scaled down for visualization.

Confusion Matrix of the proposed method (Patch Classification with Regional Importance).
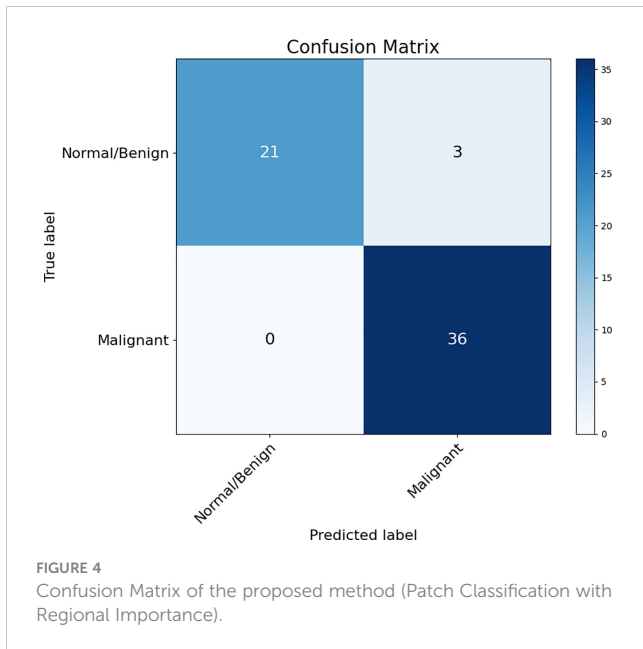
patching strategy can capture pathological traits like high cell density and infiltration.

3.3 Classification performance

The accuracy, sensitivity, specificity, and AUC score of malignant/benign binary classification on DUV-WSI images are measured for quantitative analysis. The classification performance of 5-fold cross-validation across 60 DUV-WSI images is shown in Table 2. The accuracy, sensitivity, and specificity were calculated with Equations 16-18, respectively:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \qquad (16)$$

$$Sensitivity = \frac{TP}{TP + FN}, \qquad (17)$$

$$Specificity = \frac{TN}{TN + FP}, \qquad (18)$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives. A confusion matrix summarizes the proposed method's performance with the predictions against known ground truth labels in Figure 4.

Compared to standard ResNet50, the proposed method significantly improves classification performance with a 13.3%

squares) and normal/benign adipose (green) tissue. These DUV images show that the proposed fusion method with regional importance maps can localize ROI for a WSI-level decision with confident margin prediction accuracy. The results demonstrate that even with little training data, this deep learning approach with a
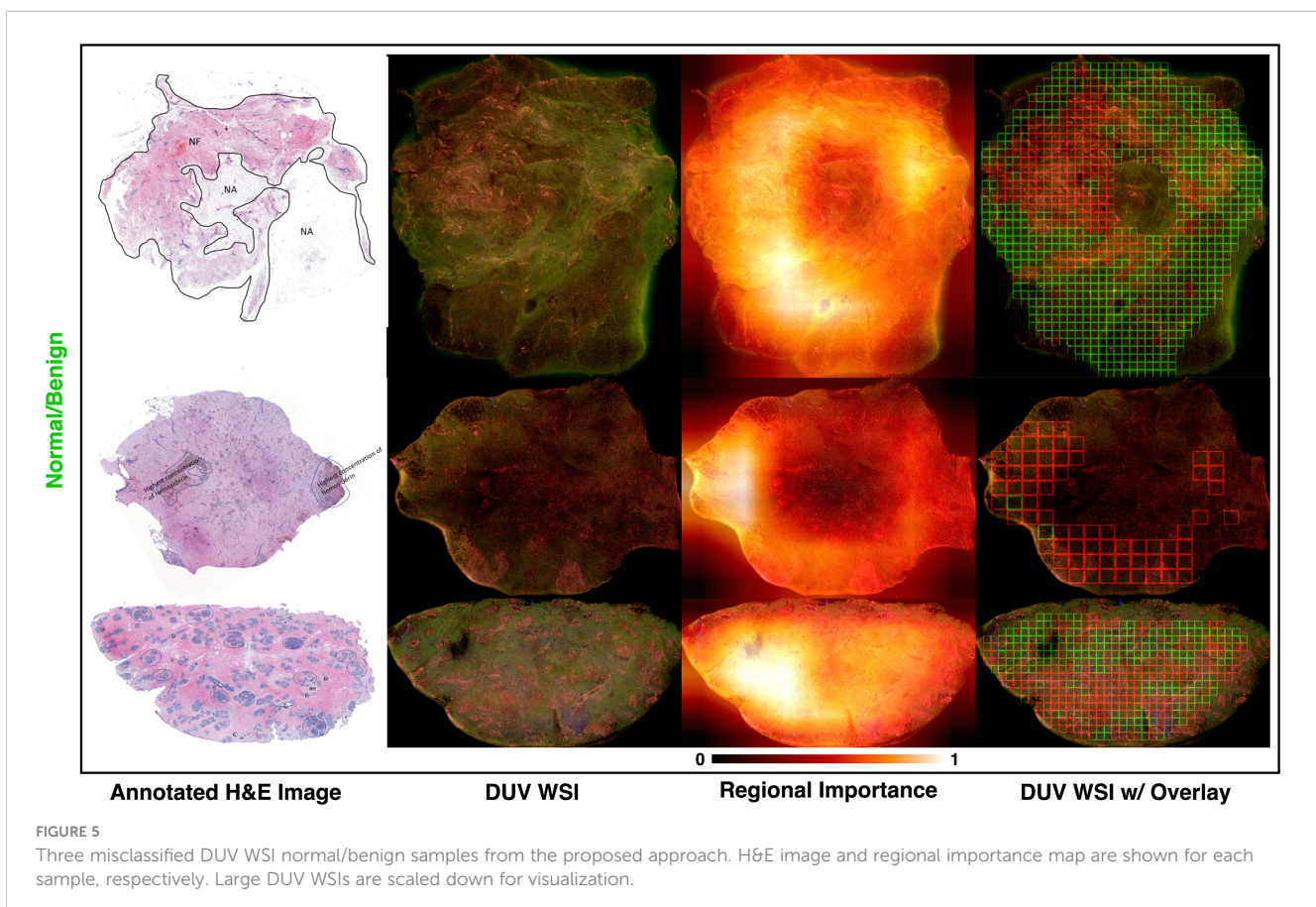
Three misclassified DUV WSI normal/benign samples from the proposed approach. H&E image and regional importance map are shown for each sample, respectively. Large DUV WSIs are scaled down for visualization.

TABLE 1 Dataset Information: Extracted Patches provides the number of patches that were extracted for normal/benign and malignant.

Extracted Patches		DUV WSIs	
	Number of Patches		Number of Samples
Normal/Benign	25,024	Normal/Benign	24
Malignant	9,444	Malignant	36
Total	34,468	Total	60

DUV WSIs provides the number of samples for normal/benign and malignant.

increase in accuracy. The proposed method is reliable, while the standard ResNet50 with DUV WSI has overfitting issues. The proposed method outperforms ResNet50 in sensitivity, specificity, and AUC scores by 8.3%, 20.8%, and 17.0%, respectively. Figure 6 illustrates the proposed method's ROC Curve, where the AUC is 17% higher than with the ResNet50 model. Compared with Patch Classification with Majority Voting, the proposed method outperforms by 1.7% and 5.6% for accuracy and sensitivity, respectively. Although the proposed method has a specificity that is 3.7% lower than another approach, it is still a good option for breast cancer classification. The proposed method achieved a perfect sensitivity rate of 100% compared to the other approach. Sensitivity is crucial in detecting malignant cases, a primary concern in breast cancer classification. The higher sensitivity rate of the proposed method makes it a valuable option despite having slightly lower specificity. These metrics demonstrate the advantage of the proposed method for intra-operative margin assessment, as it should reduce the likelihood of breast cancer margins being undetected during BCS.

3.4 Misclassified samples

The three normal/benign DUV WSI images that were misclassified as malignant are shown in Figure 5. The top sample contains a mixture of fat and fibrotic breast tissue. The dense, irregular, bumpy fibrotic tissue likely contributed to the misclassification. The middle sample has a pool of bleeding that probably resulted in misclassification. The bottom sample was considered normal/benign breast tissue with high fibro glandular density, indicating dense breast tissue. However, it is acknowledged that there may be some uncertainty in the sample classification due to other possible confounding factors. Thus, it is recommended to conduct additional analysis and confirmation to validate the classifications.

4 Discussion

This part discusses the results of the proposed breast cancer classification method for DUV image margin assessment, which combines patch-level classification using a transfer learning approach with regional importance maps generated by the Grad-CAM++ algorithm. The discussion involves the effectiveness of the method, its limitations, and potential future research directions to improve intra-operative margin assessment during breast-conserving surgery.

4.1 Methodology and results

This study used a pre-trained ResNet50 model with ImageNet weights to extract convolutional features from DUV WSI patches, which were then used to train an XGBoost classifier for patch-level classification. The regional importance maps were generated using the Grad-CAM++ algorithm on a pre-trained DenseNet169 model with ImageNet weights to localize crucial areas in the WSI image for margin assessment. The overall WSI label was determined by fusing the patch-level classification results with the regional importance map, which enhanced classification accuracy at the WSI level. The proposed method achieved an accuracy of 95% in determining DUV WSI and displayed 100% sensitivity in detecting malignant cases. Additionally, the approach could accurately localize malignant and normal/benign tissue areas, outperforming standard deep-learning classification methods on DUV breast surgical samples.

4.2 Comparison with existing methods

A previous study utilized a transfer learning-based approach with ResNet50 to classify H&E-stained histological breast cancer images from a different dataset (21). The method fine-tuned the

TABLE 2 5-fold Cross-Validation Classification Performance on 60 DUV WSI images: The proposed method (3) (Patch Classification with Regional Importance) significantly increases the classification accuracy, sensitivity, and specificity compared with the standard ResNet50 approach (1), and Patch Classification with Majority Voting (2).

	(1)	(2)	(3)
Accuracy	81.7%	93.3%	95.0%
Sensitivity	91.7%	94.4%	100%
Specificity	66.7%	91.2%	87.5%

FIGURE 6
ROC graph of the proposed method (Patch Classification with Regional Importance), ResNet50 approach, and Patch Classification with Majority Voting.
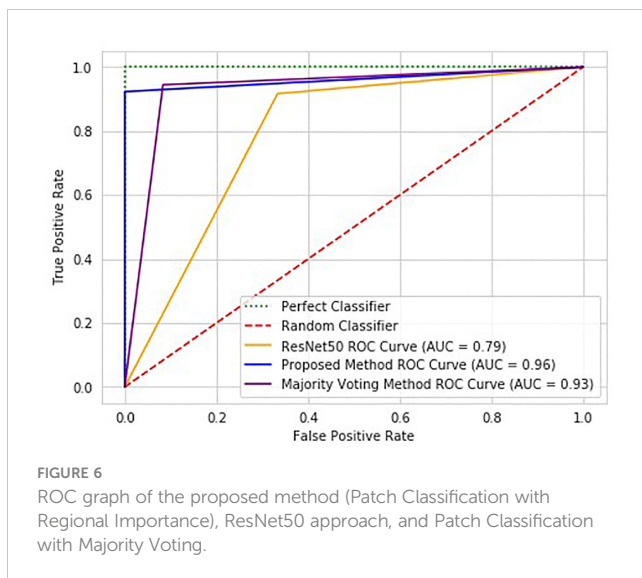
pre-trained ResNet50 network with ImageNet weights to adapt it to the domain-specific breast histology image classification task using the BACH 2018 dataset, which contains 400 breast histology microscopy images with 100 samples for normal, benign, *in situ* carcinoma, and invasive carcinoma. For the experiment, 320 samples were assigned for training/validation, and 80 samples were assigned for testing with balanced classes. To address the issue of limited data and large image sizes in the BACH dataset, patches of size 512 x 512 pixels were extracted with a 50% overlap between patches, resulting in a final amount of 14,000 patches. The training/validation set was augmented by applying data augmentation through varying degrees of rotation and flipping the extracted patches. The remaining 80 samples were used as a test set to evaluate the accuracy of classification methods. The approach achieved a whole-image classification accuracy of 97.50%, and a majority voting process was employed to classify breast histology images into four tissue sub-types: normal, benign, *in situ* carcinoma, and invasive carcinoma, based on patch-wise class labels estimated using the network. This ResNet50-based method demonstrated the powerful classification capacity of transfer learning approaches for automatically classifying breast cancer histology images, especially when dealing with limited training data.

The proposed ensemble learning-based approach showcases its robustness, even with a limited dataset, offering advantages comparable to the ResNet50-based method. In contrast, the proposed method focuses on margin assessment during breast cancer surgery using the DUV WSI dataset. The method uses a pre-trained ResNet50 model with ImageNet weights to extract convolutional features from DUV WSI patches. These extracted features are then used to train and test an XGBoost classifier for patch-level classification, providing a more robust and accurate representation of the input data. Combining the pre-trained ResNet50 model and the XGBoost classifier effectively utilizes the limited dataset available while minimizing overfitting, a critical aspect when dealing with a small dataset like the one used in the study.

The proposed ensemble learning method offers several advantages compared to the ResNet50-based approach. One key advantage is incorporating regional importance maps generated using the Grad-CAM++ algorithm on a pre-trained DenseNet169 model with ImageNet weights. These regional importance maps help focus on areas with the highest level of detail and semantic information within the DUV WSI images. By capturing the spatial distribution of the most discriminative features, the algorithm highlights the most critical regions in the image responsible for the classification decision. This emphasis on semantically significant areas substantially contributes to breast cancer margin assessment, enabling the method to hone in on the most relevant image regions that could contain malignant or benign tissue.

The proposed method aims to facilitate more accurate and efficient decisions during breast cancer surgery by concentrating on the critical aspects of margin assessment. This additional step offers a more streamlined decision-making process tailored for margin assessment during breast cancer surgery. In addition to patch-level classification, the proposed approach assigns a malignant or normal/benign label to each WSI. By focusing on this binary classification, the proposed method simplifies the task for surgeons, making it easier to determine whether the margins are clear or require further resection. On the other hand, the ResNet50-based method classifies images into four tissue types: normal, benign, *in situ* carcinoma, and invasive carcinoma. While this approach provides valuable information regarding the tissue types present, it may not directly apply to margin assessment in the same way as the proposed method, as it requires additional interpretation to decide the margin status.

Although a direct comparison between the two approaches may not be fair due to their fundamental differences, the proposed method highlights the potential advantages of deep learning techniques in breast cancer margin assessment using DUV images. By utilizing transfer learning, a robust gradient-boosting algorithm, and regional importance maps, the proposed approach demonstrates the effectiveness of deep learning in accurately identifying cancerous regions in DUV WSI images. This method paves the way for further research and development in this area, potentially contributing to improved surgical outcomes and reduced breast cancer re-excision rates after BCS.

4.3 Limitations and challenges

The dataset used in this study was relatively small, consisting of 60 images out of 66 available images. This limitation may restrict the generalizability of this study's findings, as the model's performance on more extensive or diverse datasets still needs to be investigated. Training a robust network without overfitting on such a small dataset can also be challenging, requiring careful selection of training parameters and model architecture.

Another challenge encountered in this study is the presence of mixed tissue types within many tissue samples. The label assigned to each patch may only represent a portion of the area within the patch, potentially complicating the classification process. Determining the

optimal patch size remains an open question; better classification results could be achieved by identifying and using an optimal patch size that maximizes the model's discriminatory power.

This dataset mostly contained samples that were either predominantly benign/normal or malignant, with six samples exhibiting a complex mixture of benign/normal and malignant tissues being disregarded for the sake of model training. While this approach trains the model on clean and pure samples, it may limit the generalizability of the proposed method to samples with a greater complexity of benign/normal and malignant tissues. Consequently, further investigation of the model's performance on more complex and diverse samples is essential to validate its efficacy and robustness in real-world scenarios.

4.4 Future research directions

Future research can validate findings using larger and diverse datasets, optimizing patch size for classification to enhance performance, and exploring alternative evaluation metrics for more comprehensive comparisons. Assessing the clinical impact of the proposed method can guide surgical decision-making, potentially reducing the need for re-excision. Integrating the proposed method into surgical navigation systems can improve precision and minimize damage to healthy tissues. Investigating the method's performance on other medical imaging modalities can enhance its versatility and applicability to different diagnosis techniques, contributing to better patient outcomes.

5 Conclusion

Accurate margin assessment during breast cancer surgery is crucial for reducing breast cancer re-excision rates following breast-conserving surgery (BCS). This study addressed the challenge of locating cancerous regions in DUV whole-slide images (WSI) by proposing an automated method leveraging deep learning techniques' power. This methodology combines patch-level classification using transfer learning with regional importance maps generated through Grad-CAM++. Focusing on highly significant regions within the images assigns a malignant or normal/benign label to each WSI with increased confidence. This approach enhances the robustness of breast cancer classification in DUV WSI images, which is essential for accurate intra-operative margin assessment. The proposed method was evaluated on a dataset of 60 authentic DUV WSI images, and an impressive classification accuracy of 95.0% was achieved. This result indicates the potential of the proposed method for real-time assessment of margin status during breast cancer surgery and potentially other types of cancer surgeries. However, it is essential to acknowledge that the study has limitations, such as the relatively small dataset size and potential generalizability challenges for more complex tissue samples. Future work could focus on expanding the dataset, optimizing patch size, and investigating alternative evaluation metrics to validate further and enhance the proposed method. The approach used in this study could impact clinical

decision-making significantly. It can provide real-time guidance to surgeons during operations, ensuring they make more informed decisions about resection margins. It can also help develop personalized treatment plans. Overall, this study demonstrates the effectiveness of deep learning techniques in breast cancer margin assessment using DUV WSI images. Furthermore, this method can be integrated into surgical navigation systems to visualize cancerous regions and their boundaries in real time, improving the precision of surgical interventions and minimizing damage to healthy tissues. The potential clinical applications and the possibility for further refinement of this method offer a promising direction for future research and development in medical imaging.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, TY, BY, and DY; data curation, TL, MP, JJ, and TS; funding acquisition, TY, BY, and DY; investigation, TT, and DY; methodology, TT, and DY; software, TT; writing—original draft, TT; writing—review and editing, BY, and DY. All authors have read and agreed to the published version of the manuscript.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA* (2019) 321 (3):288–300. doi: 10.1001/jama.2018.19323

2. Kummerow KL, Du L, Penson DF. Nationwide trends in mastectomy for early-stage breast cancer. *JAMA Surg* (2015) 150(1):9–16. doi: 10.1001/jamasurg.2014.2895

3. Kantor O, Pesce C, Kopkash K. Impact of the society of surgical oncology-american society for radiation oncology margin guidelines on breast-conserving surgery and mastectomy trends. *J Am Coll Surgeons* (2019) 229(1):104–14. doi: 10.1016/j.jamcollsurg.2019.02.051

4. Lu T, Jorns JM, Patton M, Fisher R, Emmrich A, Doehring T, et al. Rapid assessment of breast tumor margins using deep ultraviolet fluorescence scanning microscopy. *J Biomed Optics* (2020) 25(12):126501. doi: 10.1117/1.JBO.25.12.126501

5. Lu T, Jorns JM, Ye DH. Automated assessment of breast margins in deep ultraviolet fluorescence images using texture analysis. *Biomed Optics Express* (2022) 13(9):5015–34. doi: 10.1364/BOE.464547

6. Foomani FH, Anisuzzaman DM, Niezgoda J, Niezgoda J, Guns W, Gopalakrishnan S, et al. Synthesizing time-series wound prognosis factors from electronic medical records using generative adversarial networks. *J Biomed Inf* (2022) 125:103972. doi: 10.1016/j.jbi.2021.103972

7. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol* (2020) 65(20):20TR01. doi: 10.1088/1361-6560/ab843e

8. Zhang Y-D, Satapathy SC, Guttery DS, Górriz JM, Wang S-H. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Inf Process Manage* (2021) 58(2):102439. doi: 10.1016/j.ipm.2020.102439

9. Singh R, Ahmed T, Kumar A, Singh AK, Pandey AK, Singh SK. (2020). Imbalanced breast cancer classification using transfer learning, in: *IEEE/ACM transactions on computational biology and bioinformatics*, Vol. 18. pp. 83–93. doi: 10.1109/TCBB.2020.2980831

10. Gamble P, Jaroensri R, Wang H. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med* (2021) 1 (1):14. doi: 10.1038/s43856-021-00013-3

11. Jaber MI, Song B, Taylor C. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Res* (2020) 22(1):12. doi: 10.1186/s13058-020-1248-3

12. Li Y, Wu J, Wu Q. Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *IEEE Access* (2019) 7:21400–8. doi: 10.1109/ACCESS.2019.2898044

13. Farahmand S, Fernandez AI, Ahmed FS. Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer. *Modern Pathol* (2022) 35(1):44–51. doi: 10.1038/s41379-021-00911-w

14. Couture HD, Williams LA, Geradts J. Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype. *Breast Cancer* (2018) 4(1):30. doi: 10.1038/s41523-018-0079-1

15. Ye J, Luo Y, Zhu C, Liu F, Zhang Y. (2019). Breast cancer image classification on wsi with spatial correlations, in: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,Brighton, UK pp. 1219–23. doi: 10.1109/ICASSP.2019.868256 0

16. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* Lake Tahoe, NV, USA. (2017) p. 839–47. doi: 10.1109/WACV.2018.00097

17. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, (CVPR), Las Vegas, NV, USA pp. 770–8. doi: 10.1109/CVPR.2016.90

18. Shokouhmand A, Aranoff ND, Driggin E, Green P, Tavassolian N. Efficient detection of aortic stenosis using morphological characteristics of cardiomechanical signals and heart rate variability parameters. *Sci Rep* (2021) 11(1):1–14. doi: 10.1038/s41598-021-03441-2

19. Huang G, Liu Z, Kilian Q. Weinberger. densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA* (2017) p. 2261–9. doi: 10.1109/CVPR.2017.243

20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. (2017). Grad-cam: visual explanations from deep networks *via* gradient-based localization, in: *Proceedings of the IEEE international conference on computer vision*, (ICCV), Venice, Italy pp. 618–26. doi: 10.1109/ICCV.2017.74

21. Vesal S, Ravikumar N, Davari A. Classification of breast cancer histology images using transfer learning. *ArXiv* (2018). doi: 10.1007/978-3-319-93000-8_92