

Article

Federated Learning-Based Detection of Invasive Carcinoma of No Special Type with Histopathological Images

Bless Lord Y. Agbley ¹, Jianping Li ^{1,*}, Md Altab Hossin ², Grace Ugochi Nneji ³, Jehoiada Jackson ³, Happy Nkanta Monday ¹ and Edidiong Christopher James ³

¹ School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; agbleybless@outlook.com (B.L.Y.A.); mh.nkanta@std.uestc.edu.cn (H.N.M.)

² School of Innovation and Entrepreneurship, Chengdu University, Chengdu 610106, China; altabbd@163.com

³ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; ugochinneji@std.uestc.edu.cn (G.U.N.); kofijackson@uestc.edu.cn (J.J.); edianajames@yahoo.com (E.C.J.)

* Correspondence: jpli2222@uestc.edu.cn

Abstract: Invasive carcinoma of no special type (IC-NST) is known to be one of the most prevalent kinds of breast cancer, hence the growing research interest in studying automated systems that can detect the presence of breast tumors and appropriately classify them into subtypes. Machine learning (ML) and, more specifically, deep learning (DL) techniques have been used to approach this problem. However, such techniques usually require massive amounts of data to obtain competitive results. This requirement makes their application in specific areas such as health problematic as privacy concerns regarding the release of patients' data publicly result in a limited number of publicly available datasets for the research community. This paper proposes an approach that leverages federated learning (FL) to securely train mathematical models over multiple clients with local IC-NST images partitioned from the breast histopathology image (BHI) dataset to obtain a global model. First, we used residual neural networks for automatic feature extraction. Then, we proposed a second network consisting of Gabor kernels to extract another set of features from the IC-NST dataset. After that, we performed a late fusion of the two sets of features and passed the output through a custom classifier. Experiments were conducted for the federated learning (FL) and centralized learning (CL) scenarios, and the results were compared. Competitive results were obtained, indicating the positive prospects of adopting FL for IC-NST detection. Additionally, fusing the Gabor features with the residual neural network features resulted in the best performance in terms of accuracy, F1 score, and area under the receiver operation curve (AUC-ROC). The models show good generalization by performing well on another domain dataset, the breast cancer histopathological (BreakHis) image dataset. Our method also outperformed other methods from the literature.

Keywords: breast cancer; deep learning; federated learning; invasive carcinoma of no special type; whole slide images; histopathological image analysis



Citation: Agbley, B.L.Y.; Li, J.; Hossin, M.A.; Nneji, G.U.; Jackson, J.; Monday, H.N.; James, E.C. Federated Learning-Based Detection of Invasive Carcinoma of No Special Type with Histopathological Images. *Diagnostics* **2022**, *12*, 1669. <https://doi.org/10.3390/diagnostics12071669>

Academic Editor: Dechang Chen

Received: 26 May 2022

Accepted: 5 July 2022

Published: 9 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancers are among many diseases the research community is working hard to detect with the aid of automated systems [1–4]. While tumors can be benign or malignant, the former is not considered cancerous as their cells are more regular, develop more slowly, and are not invasive of tissues around them. Malignant tumors, on the other hand, are cancerous tumors. The term “breast cancer” is used to refer to malignant tumors originating from the lobules or ducts of the breast or, in rare cases, the stromal tissue (the connective fatty and fibrous breast tissues), which are then further propagated to other parts of the body by the lymphatic system [5]. They are caused by genetic abnormalities where about 5% to 10% of the cases are due to inheritance from parents. About 90% of the hereditary cases are a result of mutations in the breast cancer genes BRCA1 and BRCA2 [6–9]. Breast cancer

makes up 30% of all new cancer cases diagnosed in women. The American Cancer Society estimates 287,850 new invasive cancer cases in women in 2022, and 43,250 individuals will die of the disease. Additionally, an estimate of 2710 men is expected to be diagnosed with invasive breast cancer in 2022. Out of this number, about 530 are likely to die from it [10].

Invasive carcinoma of no special type (IC-NST), previously referred to as invasive ductal carcinoma (IDC) or breast cancer not otherwise specified (NOS) [11], is known to be a very prevalent kind of breast cancer, making up about 80% of all breast cancers according to the American Cancer Society.

The procedure for diagnosing breast cancer involves a combination of different tests, including physically examining the breast, mammograms, and biopsy. Other examinations with ultrasound and breast magnetic resonance imaging (MRI) may also be considered [12]. The study carried out by Collins et al. indicated that core needle biopsy provides high rates for diagnosing IC-NST in over 90% of the cases, and it is a primary technique for evaluating histopathological features among pathologists [13]. The process, however, involves manual feature engineering by expert pathologists who must carefully observe and examine the glass slide of biopsy specimens. Capturing these specimens into images makes them available for use in computer-aided detection/diagnosis [14]. Approaches such as machine learning (ML) and, more specifically, deep learning (DL), have been explored in recent years due to their successes in aiding in the prognosis and diagnosis of other medical conditions [15–22]. These techniques provide mathematical models for automating the detection process. However, to effectively take good advantage of the ability of DL models to eliminate manual feature engineering, the methods tend to require large amounts of data. The regulations governing the release of patient data are, however, stringent, resulting in a small amount of publicly available data.

To address this privacy concern and encourage collaborative learning, we employ federated learning in this study. This paper contributes the following to help in the automated detection of IC-NST:

- We proposed a multimodal network by introducing two input modalities obtained separately by extracting histopathology image features with GaborNet and pre-trained ResNet.
- We formulated the problem of automatically diagnosing IC-NST as a federated learning challenge to leverage its privacy preservation capability.
- We conducted experiments to evaluate our approach and compared the results against some techniques in the literature in terms of accuracy, F1-score, and balance accuracy metrics.
- We also assessed how well our models generalized by evaluating them on another domain dataset from a different repository.
- Visualizations of the layers of both ResNet and GarborNet were provided to give more insight into the models and enhance their explainability.

We organized the subsequent sections in the following manner. Section 2 presents literature related to IC-NST and motivation for the study. In Section 3, we present our methodology. We implemented our approach, presented our experimental results, and compared them with other methods proposed in the literature in Section 4. Finally, we concluded in Section 5 and presented some future research directions.

2. Related Works

IC-NST detection has been a very important topic for researchers applying ML and DL in addressing breast cancer issues. Romano and Hernandez [23] performed a study on IC-NST in which they trained an enhanced convolutional neural network (CNN) model and analyzed the results of their model with a patch-based IC-NST classification problem. Their study achieved an F-score and balanced accuracy of 85.28% and 85.41%, respectively. Brancati et al. [24] explored some techniques to automatically analyze hematoxylin and eosin (HE)-stained histopathological images of breast cancer and lymphoma. They also proposed their approach for two use cases: the detection of IC-NST in breast histopathological images

and the classification of lymphoma subtypes. They reported having improvements of 5.06% in the F-score and 1.09% in the accuracy measure over previous methods. Liu et al. [25] proposed a new index of the energy to Shannon entropy ratio (ESER), which they used for classifying the tissues. They utilized principal component analysis (PCA) together with an ML technique to autonomously classify THz signals from IC-NST samples and had AUCs greater than 0.89, 92.85% precision, 89.66% sensitivity, and 96.67% specificity. Chapala et al. [26] proposed the use of the ResNet50 framework to automatically detect IC-NST from breast histopathology images (BHI). Their model achieved an accuracy of 91%. Celik et al. [27] conducted a study involving the automatic detection of IC-NST using a deep transfer learning technique. They used some pre-trained models of DL, including ResNet50 and DenseNet161, on BHI for their study. They obtained an F-score of 92.38% and a balanced accuracy value of 91.57% on the DenseNet161 model. They also obtained an F-score of 94.11% and a balanced accuracy value of 90.96% on the ResNet-50 model. Chand et al. [28] performed a more inclusive study on the various methods used over the recent years for IC-NST detection.

Transfer learning (TL) is a popular technique used to improve the performance of models for which there are inadequate training samples. They are also very convenient for reducing the training time by basing on the pre-trained weight of earlier layers in the network to fine-tune new models. Popular TL architectures are trained with the ImageNet dataset. For a medical use case such as the task of IC-NST detection, ImageNet-based transfer learning methods become out of the domain. Hence, Laith et al. [29] proposed a novel TL technique for medical imaging tasks with limited labeled datasets. They take advantage of the unlabeled medical data available to train models that learn domain-specific features, and the learned weights are used to fine-tune models for which there are limited datasets. The authors also proposed a new deep convolutional neural network (DCNN) to exploit recent advances in the area. To evaluate their proposed TL and DCNN models, they performed separate experiments involving automated skin cancer and breast carcinoma classification. Among the many promising results achieved by their method, their model obtained an increment from an 86.0% F1-score when trained without any form of TL to 96.25% with TL, as well as an even higher F1-score of 99.25% with the double-TL technique on a two-class (normal and diabetic foot ulcer) skin classification task. In a similar paper, the authors [30] proposed a hybrid DCNN to classify HE-stained breast biopsy images into four targets (invasive carcinoma, in situ carcinoma, benign tumor, and normal tissue). They empirically showed that domain-based TL does well at optimizing the performance of the models. Augmenting the training instances reduced overfitting by providing more instance variations for the model to explore for generalization.

Despite the promising results of all the methods proposed in the literature involving DL and TL techniques, privacy concerns regarding health data add more limitations to the availability of publicly accessible data. Institutions are usually obliged to run models on their localized datasets to keep their patients' data private. Motivated by the successful use of FL by Google to achieve a high performing model for predicting words on their Gboard [31,32], some recent works in computerized medical image analysis [19] have adopted FL to help protect patients' privacy. This work also uses FL to bridge the IC-NST data availability gap and to encourage collaboration between different institutions without sharing their patients' data.

3. Proposed Approach

Data scarcity affects the generalization of deep learning models as there are not enough instance variations to represent the distribution under study adequately. This is a big challenge; hence, institutions resort to collaborating by donating their data to a common repository [33] for centralized learning (CL). While this helps in addressing the challenge of availability of big data to improve model generalization, it also introduces the challenge of data privacy [34]. This is a significant concern in health, adding to the already existing data

limiting factors such as the lack of experts to annotate and digitize traditionally acquired data.

We proposed federated learning-based IC-NST detection in which the clients utilize pre-trained residual models and the Gabor network to extract features from their private IC-NST datasets. The architectural diagram of the proposed model is represented in Figure 1. It consists of the global FL model on the sever (A), the client model architectures residing on each client node (B), and the private datasets for each client, also residing on the client node. More details on our approach are provided in the following sub-sections.

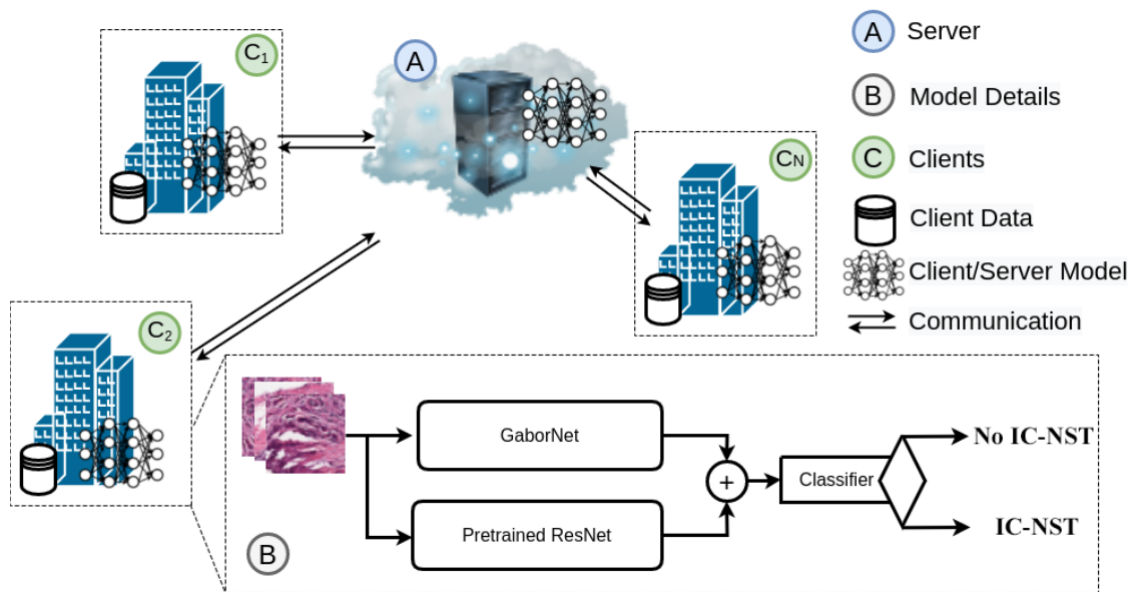


Figure 1. Architecture of the proposed approach with GaborNet and ResNet models on a federated setting.

3.1. Federated Learning (FL)

FL eliminates the need for one common data bank for access by the participating institutions. Their data remain with them while they contribute training parameters to help improve each other’s local models. We present our proposed FL approach in Algorithm 1. Using federated averaging [31], the parameters of each local client model are averaged on a central server. These averaged parameters are then used to update the participating clients for a number of rounds. For each round of server update, each of the clients trains its local model for several epochs. Given N number of clients, each client has its own local IC-NST patch dataset, X_i and, labels, Y_i , where i specifies the client. The set of all the client data are $\{X_1, \dots, X_N\}$ with labels, $\{Y_1, \dots, Y_N\}$. A global model M is agreed upon by all the clients, and a clone m_i is made by each of the clients and initialized with the parameters of M . The entire local model list is thus $\{m_1, \dots, m_N\}$ parameterized with $w_i = W$, where w_i is the local parameters while W represents the global parameters.

The server aggregation phase is carried out after each round of local training. The local parameters $\{w_1, \dots, w_N\}$ are transmitted to the server for aggregation using the federated averaging (FedAvg) algorithm to compute the average of all the local parameters, which becomes the new global parameter, W , as shown in Equation (1) below:

$$\underset{w}{\text{minimize}} F(w) := \frac{1}{N} \sum_{i=1}^N F_i(w), \tag{1}$$

where $F_i(w)$ represents the risk function for client i . The risk function is further presented in Equation (2) below:

$$F_i(w) = \mathbb{E}_{(x_i, y_i) \sim \mathcal{D}_i} [\mathcal{L}(m_w(x_i), y_i)], \quad (2)$$

where \mathcal{D}_i represents the distribution of each client data, \mathcal{L} represents the loss function, m_w represents the local model, and $m_w(x_i)$ represents the model prediction for a data instance x_i . Approximating the risk with an empirical function, we have Equation (3):

$$F_i(w) \approx \frac{1}{n_i} \sum_{i=1}^{n_i} [\mathcal{L}(m_w(x_i), y_i)] \quad (3)$$

We used the cross entropy loss function for the optimization of our models. For balanced classes, the loss is formulated as follows:

$$\mathcal{L}(m_w(x_i), y_i) = -\log \left(\frac{\exp(x[y])}{\sum_j \exp(x[j])} \right) \quad (4)$$

$$\mathcal{L}(m_w(x_i), y_i) = -x[y_i] + \log \left(\sum_j \exp(x[j]) \right) \quad (5)$$

Algorithm 1: Proposed FL-Based Approach.

Input: N clients; R : communication rounds; E : local epochs; global parameters: W ; local parameters: $\{w_i, \dots, w_N\}$; η : learning rate; b : size of batch; X : images; Y : labels; Client data: $\{X_1, \dots, X_N\}$; client labels: $\{Y_1, \dots, Y_N\}$; Local models: $\{m_i, \dots, m_N\}$; Optimizers: $\{optim_1(\cdot), \dots, optim_N(\cdot)\}$

Output: Global model: M_W

```

1 Initialize model parameters:  $\{w_i, \dots, w_N\} \leftarrow W$ 
2 for  $r = \{1, \dots, R\}$  do
3   for  $i = \{1, \dots, N\}$  do
4     for  $e = \{1, \dots, E\}$  do
5       Get training patches for client  $i$ ,  $X_{i_{train}}$ 
6       Apply augmentations
7        $Examples_{train} \leftarrow augData(X_{i_{train}})$ 
8        $map_1 \leftarrow ResNet(Examples_{train})$ 
9        $map_2 \leftarrow GaborNet(Examples_{train})$ 
10       $map_{fused} \leftarrow fusion(map_1, map_2)$ 
11       $pred = classifier(map_{fused}, Y_{train_i})$ 
12      Compute prediction loss
13       $\mathcal{L}_{pred}(map_{fused}, Y_{train_i})$ 
14       $= weight[Y_{train_i}] \left( -x[Y_{train_i}] + \log \left( \sum_j \exp(x[j]) \right) \right)$ 
15      Update  $w_i \leftarrow optim_i(\mathcal{L}_{pred})$ 
16    end
17    Get validation patches for client  $i$ ,  $X_{i_{val}}$ 
18    Validate and compute metrics
19  end
20  Server Aggregation
21   $W \leftarrow FedAvg(\{w_i, \dots, w_N\})$ 
22  Testing with global model
23  Get test patches  $X_{test}$ 
24  Deploy weights to client models
25   $\{w_i, \dots, w_N\} \leftarrow W$ 
26 end
27 return  $M_W$ 

```

3.2. FL Model Architecture

Taking advantage of TL for fine-tuning models trained on smaller datasets [35–38], this study utilizes residual network models pre-trained with the ImageNet dataset for automatic feature extraction from the histopathology patches to produce one feature modality. A second feature modality is obtained using a Gabor network.

ResNet: ResNet18 and ResNet50 are two of the deep learning networks proposed by He et al. [39], in which they showed that residual learning by referencing the layer's input enables the construction of deeper networks without increasing the error. Similar to the original ResNet18 and 50 networks, our implementation consists of a convolutional layer with 7×7 filters, an output channel of 64, a stride of 2, and padding of 3×3 . Following the convolution layer is a 3×3 max-pooling with a stride of 2. Four sequential layers consisting of blocks of convolution, batch normalization, and ReLU layers are then added. The major difference between the ResNet18 and 50 networks is the number of blocks used to compose each layer. While ResNet18 consists of 2 blocks per layer, ResNet50 contains 3, 4, 6, and 3 blocks, respectively, for each layer. Average pooling is added after the last layer of both networks. We modified the linear layer in the fully connected block to output a feature size of 448.

GaborNet: Gabor filters are linear band pass filters that analyze frequencies in localized regions for directions defined by the kernel [40,41]. Given a kernel size (p, q) , phase offset, ψ , spatial aspect ratio, γ , and standard deviation, σ , a complex Gabor kernel is obtained using the function $g(\cdot)$ in Equation (6). The real and imaginary components are Equations (7) and (8), respectively.

$$g(\cdot) = \exp\left(-\frac{p'^2 + \gamma^2 q'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{p'}{\lambda} + \psi\right)\right) \quad (6)$$

$$g(\cdot) = \exp\left(-\frac{p'^2 + \gamma^2 q'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{p'}{\lambda} + \psi\right) \quad (7)$$

$$g(\cdot) = \exp\left(-\frac{p'^2 + \gamma^2 q'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{p'}{\lambda} + \psi\right) \quad (8)$$

Here, $p' = p \cos \theta + q \sin \theta$ and $q' = -p \sin \theta + q \cos \theta$. A total of 96 kernels were defined by varying the function parameters to produce different frequencies at different directions. We used the kernels along with 2D convolution to extract features from the images. Another convolutional layer with a 3×3 kernel size and an output size of 384 are added. We also added a linear layer consisting of 64 output size.

Classifier: The features obtained from the ResNet (map_1) and the features extracted with the GaborNet (map_2) are fused [19,42] and fed into a classifier for prediction. The classifier consists of a linear layer with a 256 output size, a ReLU activation layer, a batch normalization layer, a dropout of 0.5, and a final linear layer with an output of 2 for the two classes.

3.3. Centralized Learning (CL)

Since the major practice over these past years is to have the training data centralized in one device or server, we considered CL our benchmark technique for comparison with the proposed FL approach. The CL technique involves running the same deep learning networks described in Section 3.2 for the FL approach in a non-distributed fashion. Hence, the datasets are kept at one location, and the various model variations are evaluated on them. Those models specific to the CL are named CL+ResNet18, CL+ResNet18+Gabor, CL+ResNet50 and CL+ResNet50+Gabor. CL in the naming refers to the centralized learning mode, while ResNet18, ResNet50, and GaborNet refer to the feature extraction techniques. The "+" symbol implies the combination of different techniques.

3.4. Datasets

Deep learning relies heavily on datasets to automatically extract features that uniquely characterize the various target classes. We used two histopathological datasets from different repositories for our study. These datasets are described in the following subsections.

3.4.1. Breast Histopathology Image (BHI) Dataset

We utilized the Breast Histopathology Image (BHI) dataset [41,43,44] for training. The original dataset contains whole slide images (WSI) taken at $40\times$ magnifying factor. In total, 277,524 patches of size 50×50 were obtained from the whole slide images, of which 198,738 patches are for IC-NST-negative and 78,786 are for IC-NST-positive cases, as shown in Figure 2. Samples of the two classes in the BHI dataset are shown in Figure 3a.

Considering a subject from the BHI dataset, for instance, the corresponding patches are aligned to enable us to visualize the entire tissue without a mask of the malignant tumor region as shown in Figure 4a. The region indicated by the deep red mask is the cancerous region, shown in Figure 4b. The tissue slice with the binary positive and negative target patches are also shown in Figure 4c. The pre-processing included augmenting the training and validation patch instances by randomly flipping the images horizontally and vertically and then applying a z-score normalization.

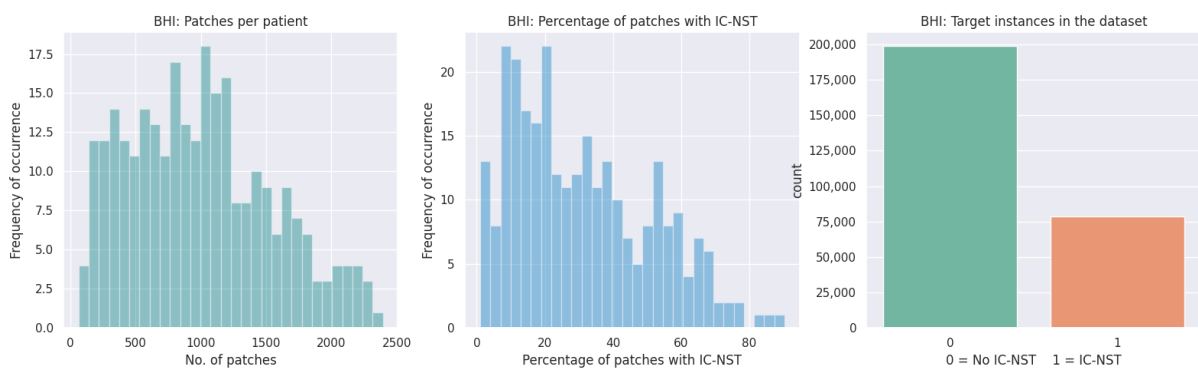


Figure 2. Patches available in the BHI dataset.

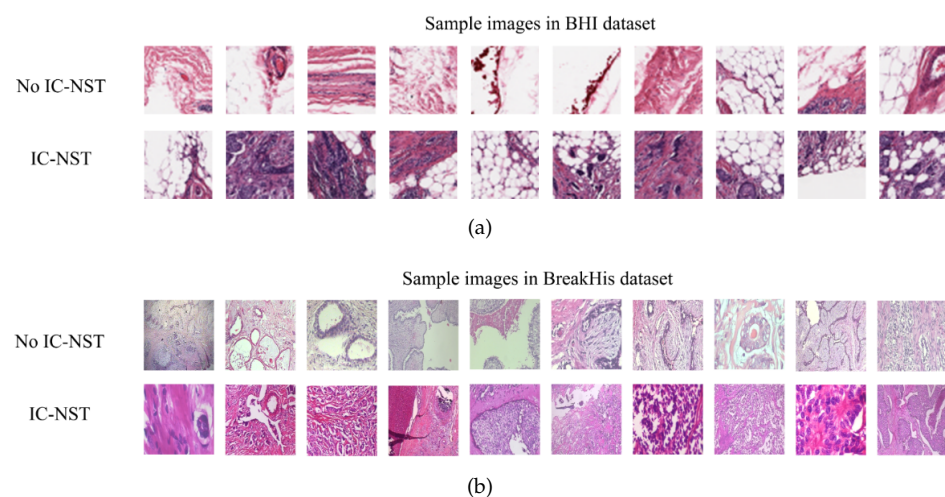


Figure 3. Histopathology image samples in the (a) BHI and (b) BreakHis datasets.

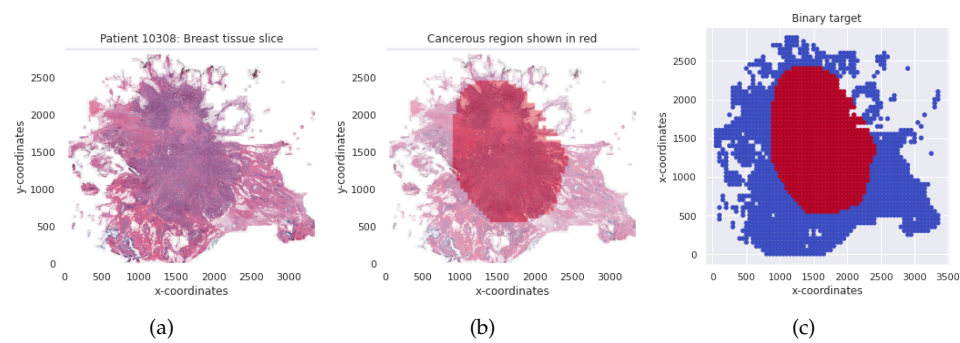


Figure 4. (a) Breast tissue slice for an IC-NST subject. (b) The red colored mask shows the cancerous region. (c) Binary target per tissue slice for the IC-NST subject.

We identified 279 unique patients in the BHI dataset; hence, we obtained an identically independent distribution (IID) by allocating 93 patients randomly to each client. All the patches of the patients assigned to the clients were used to train their private models.

3.4.2. Breast Cancer Histopathological (BreakHis) Image Dataset

Samples obtained through surgical (open) biopsy (SOB) were prepared into histopathological images to create the Breast Cancer Histopathological Image dataset by staining with hematoxylin and eosin (HE) [45–47]. The microscopic breast tumor images were taken at four different magnifying factors: $40\times$, $100\times$, $200\times$, and $400\times$. A total of 9109 images were present, of which 2480 belonged to the benign class, while 5429 belonged to the malignant class. The images were 3-channel RGB images of size 700×460 . Based on the appearance of the cells, the images can be categorized into the benign (adenosis, fibroadenoma, phyllodes, and tubular adenoma) class and the four invasive cancer subtypes (ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma). Samples in this dataset are graphically presented in Figure 3b.

In order to handle the difference in the sizes between the BHI and BreakHis datasets, we cropped 50×50 patches from the center of all the images considered for evaluation from the BreakHis dataset. Cropping from the center retains information about the context without losing the target class.

3.4.3. Handling Imbalance Data Label for Loss Computation

There is an imbalance in the number of instances in the target classes. Imbalanced classes tend to learn the class with larger training instances more effectively than the class with the smaller instances, thereby affecting the overall model performance. To overcome this issue, we computed the weights, given each class, j in $\{1, \dots, J\}$, for each client as follows:

$$\mu_{i,j} = \frac{|X_i|}{J \times |Y_{i,j}|} \quad (9)$$

where $|X_i|$ is the number of instances for client i , and $|Y_{i,j}|$ is the total number of instances in class j for client i . The loss function in Equation (5) is modified to factor in the weight of each class as follows:

$$\mathcal{L}(m_w(x_i), y_i) = \mu_{i,j} \left(-x[y_i] + \log \left(\sum_j \exp(x[j]) \right) \right) \quad (10)$$

4. Experiments

4.1. Experiment

Different sets of experiments were performed on both the CL and FL setups. Specifically, we ran experiments using only ResNet18 and ResNet50 for feature extractions and

also combined the GaborNet for each of the cases. We performed training, validation, and testing using the BHI dataset. Exploring the dataset, 279 unique patients were identified, and hence we allocated 93 patients to each of the clients. This ensured that the patches owned by each of the clients belong to a patient assigned to them. Thus, every patch located on each client node was private to that client. The patches at each client node were then split into train, dev and test sets of 70%, 15% and 15% respectively. The train and dev sets, shown in Table 1, were kept private to each client. The test sets at each client node are combined to produce one set for evaluating the global model. While each client validated their respective models on their dev-set, the combined test-set was used in evaluating the global model. The data augmentation techniques used for both the training and development sets at the client level included random horizontal and vertical flips and z-score normalization. During training, local parameters were sent to the server for averaging after every 2 epochs for a total of 30 rounds. Due to the imbalanced nature of the dataset, we computed the class weights for each client according to Equation (9) using the weighted cross-entropy loss function as our criterion. The stochastic gradient descent (SGD) with varying learning rates was used. We scheduled the learning rate using Equation (11):

$$lr_{next} = lr_1 \times 0.085^{e+1}, \quad (11)$$

where the initial learning rate was $lr_1 = 1.0$, and $e \in \mathbb{N}^+$. A batch size of 512 was used for all the experiments.

We provide a visualization of the features extracted by the Global FL+ResNet50+Gabor model in Figure 5. In Figure 5a, we show a sample of an original malignant image and its corresponding gray image. We also show the magnitude and Gabor phase obtained after applying a filter with a wavelength of 10 at an orientation of 90 degrees to the image. In Figure 5b, we show a grid of different filters (filter bank) used to extract different orientations of frequency-based features from the sample image. The extracted features are shown in Figure 5c. Next, Figure 5d,e show the features extracted by the first and second layers of the model. Finally, we visualized the model using Grad-CAM to provide insight on the model's localization for its predictions [48]. We used PyTorch and an NVIDIA GeForce RTX 2080 Ti accelerator with a memory of 11 Gb for the experiments.

Table 1. Private train and development splits for clients on the BHI dataset (patch-wise).

	Client 1	Client 2	Client 3
Train set	63,988	65,963	65,079
Dev set	11,966	13,252	12,561

4.2. Metrics, Results, and Discussion

For each experiment, we calculated the true positive predictions (TP), the false positive predictions (FP), the true negative predictions (TN), and the false negative predictions (FN). These were further used to compute commonly used metrics for medical data classification tasks [24,27,49,50] and the performance shown in Tables 2 and 3. Columns C_1 , C_2 , and C_3 in the tables refer to client 1, client 2, and client 3, respectively. G refers to the global model (obtained by the FL) or the centralized model.

Specificity is a measure of the proportion of negatives the model classified correctly. Specificity is computed with

$$Specificity = \frac{TN}{TN + FP}. \quad (12)$$

From the result, the best specificity performance was attained by the federated model using ResNet18 and GaborNet features. We noticed that the results obtained by the FL models outperformed the CL models with respect to the specificity metrics. Additionally, fusing the features from the GaborNet and the ResNets tended to further improve the specificity.

Recall gives the proportion of positives correctly classified by the model and is obtained with the equation

$$Recall = \frac{TP}{TP + FN} \times 100. \quad (13)$$

From Table 2, the best recall performance of 82.15% was attained when we performed centralized learning with both ResNet18 and GaborNet features. Likewise, the subsequent best recall of 79.90% was obtained by training the CL model with both ResNet50 and GaborNet feature extractors.

Precision reflects the accuracy of the positive predictions. It is obtained by dividing the TP by the total positives, as shown in Equation (14).

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

The best precision, 80.05%, was obtained by using both ResNet18 features and the GaborNet features on FL. Using ResNet50 and GaborNet features on FL resulted in the next best result of 79.60%.

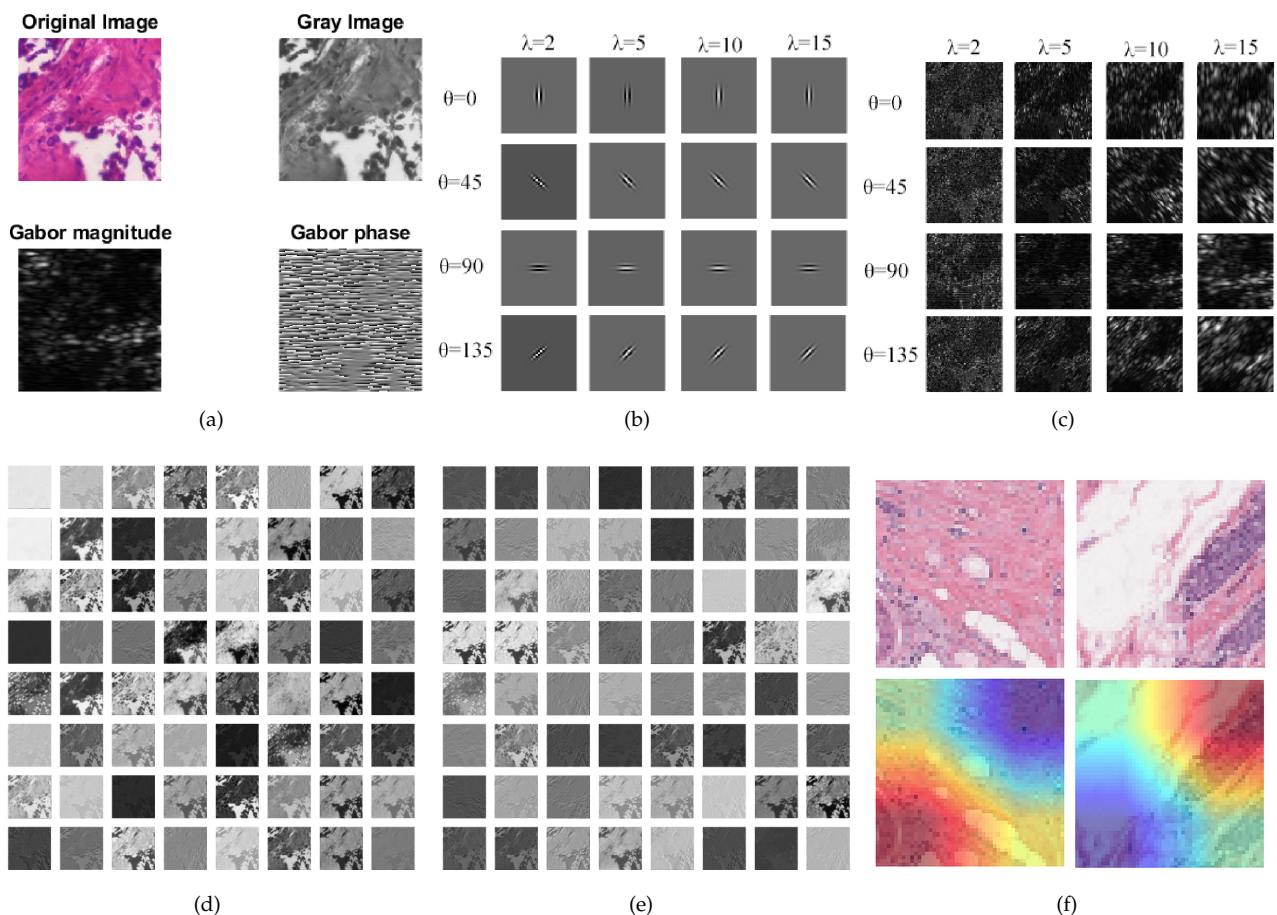


Figure 5. Visualizations: (a) Original histopathology image, gray-scale image and the Gabor magnitude and phase for wavelength $\lambda = 10$ and orientation $\theta = 90$. (b) Sample Gabor filter bank with different λ and θ values. (c) Features extracted with the filter bank. (d) Features learned by the first CNN layer of ResNet50. (e) Features learned by the second CNN layer of ResNet50. (f) Using Grad-CAM to visualize some images.

Table 2. Patch-wise result comparison on the test datasets with the CL and the FL global models.

Metric	Precision (%)				Recall (%)				Specificity (%)			
	C ₁	C ₂	C ₃	G	C ₁	C ₂	C ₃	G	C ₁	C ₂	C ₃	G
CL+ResNet18	-	-	-	78.46	-	-	-	77.37	-	-	-	90.40
FL+ResNet18	71.50	78.55	82.72	79.15	77.62	73.02	59.15	73.75	86.72	91.94	94.42	91.22
CL+ResNet18+Gabor	-	-	-	75.92	-	-	-	82.15	-	-	-	88.23
FL+ResNet18+Gabor	72.66	77.42	83.14	80.05	75.53	68.77	64.96	73.85	87.16	90.94	84.05	91.68
CL+ResNet50	-	-	-	79.18	-	-	-	75.13	-	-	-	91.07
FL+ResNet50	63.81	76.42	79.66	78.59	79.91	68.50	74.10	77.35	77.75	90.45	91.45	90.48
CL+ResNet50+Gabor	-	-	-	76.02	-	-	-	79.90	-	-	-	88.61
FL+ResNet50+Gabor	71.36	73.09	82.32	79.60	78.69	70.98	65.46	76.42	85.27	88.95	93.65	91.15

Table 3. Patch-wise comparison of our proposed models with studies from the literature on the BHI dataset.

Metric	Accuracy (%)				F1-Score (%)				BA Score (%)			
	C ₁	C ₂	C ₃	G	C ₁	C ₂	C ₃	G	C ₁	C ₂	C ₃	G
CL+ResNet18	-	-	-	86.35	-	-	-	86.32	-	-	-	83.89
FL+ResNet18	83.00	83.66	83.44	85.78	83.09	83.01	82.57	85.64	81.82	78.64	76.78	82.48
CL+ResNet18+Gabor	-	-	-	86.33	-	-	-	86.47	-	-	-	85.19
FL+ResNet18+Gabor	83.54	84.04	84.99	86.13	83.62	83.76	84.43	85.97	81.34	79.85	79.50	82.77
CL+ResNet50	-	-	-	86.12	-	-	-	86.01	-	-	-	83.10
FL+ResNet50	80.57	84.18	86.05	86.40	81.20	83.36	85.91	86.36	82.28	80.92	82.78	83.92
CL+ResNet50+Gabor	-	-	-	85.90	-	-	-	86.00	-	-	-	84.26
FL+ResNet50+Gabor	84.02	82.84	84.88	86.57	84.28	84.47	84.35	86.49	83.27	79.59	79.55	83.78

Accuracy measures the percentage of correctly classified cases, including both the positive and negative classes. The computation is shown in the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100. \tag{15}$$

From Table 3, the accuracy obtained by training ResNet18 in a centralized fashion with and without the Gabor features are at par with each other with a difference of only 0.02%. Combining GaborNet and ResNet18 feature extractors in a FL fashion produced an improvement of 0.35% over using only ResNet18 for FL. For ResNet50, the centralized model without the Gabor features performed better than its Gabor feature combination.

The F_1 score calculates the harmonic mean of the precision and sensitivity as formulated in Equation (16).

$$F_1 = 2 \times \frac{precision \times sensitivity}{precision + sensitivity} \tag{16}$$

The best F_1 score (86.49%) was obtained by the FL model trained on both ResNet50 and GaborNet feature combination. This result was closely followed by the CL model trained on features from both ResNet18 and GaborNet, with a difference of 0.02. The GaborNet improved the result in these two cases.

Balance accuracy is another important evaluation metric used to measure the effectiveness of binary classifiers, especially when the classes are imbalanced. It is the average of the recall and specificity, as shown in Equation (17).

$$\text{Balance Accuracy} = \frac{\text{Recall} + \text{Specificity}}{2} \quad (17)$$

CL with both ResNet18 and the GaborNet features achieved the highest balanced accuracy result, followed by CL with both ResNet50 and GaborNet features. The results indicate an improvement in the balance accuracy when the GaborNet features were fused with the ResNet features in both cases.

The receiver-operating curve (ROC) is a plot of *recall* against $1 - \text{specificity}$. Area under the curve (AUC) is the most widely used metric for measuring the overall performance of methods as it shows the chances that a randomly selected positive instance is more highly ranked than a randomly chosen negative one. Figure 6 shows the ROC curve and the corresponding area under the receiver-operating curve (AUC-ROC) results for each of the clients and the final global model for FL+ResNet50+GaborNet (the model with the best accuracy).

Confusion matrix helps visualize the number of actual and predicted instances of a dataset. Figure 7 shows the Confusion matrices of the FL+ResNet50+Gabor model on the BHI dataset for the global FL model, Figure 7a, and for the local models of each client, Figure 7b–d. From the figures, all the local models performed well, given the imbalanced BHI dataset having more negative patch instances compared to the number of positive patch instances. The generalization of the local models were reflected in the performance of the global model.

4.3. Evaluation on the BreakHis Dataset

To assess how well the trained models have generalized, we evaluated the performance of our approach on another domain data. The BreakHis dataset introduced in Section 3.4.2 was used to perform additional evaluations. All the 2480 images in both the subtypes and magnifying factor categories were combined for the benign class to form the negative (no IC-NST) instances. The malignant images were shuffled, and 3020 samples were randomly selected for the positive (IC-NST) instances. The results are reported in Table 4. From the table, the highest accuracy of 86.57% was obtained on the BHI dataset with our proposed FL+ResNet50+Gabor model. The corresponding accuracy with the same model for the BreakHis dataset is 86.32%, indicating only a 0.26% difference in the accuracy metric. However, the BreakHis dataset's accuracy outperformed that of the BHI dataset on the CL+ResNet50+Gabor model. All the BHI results for specificity outperformed those of the BreakHis dataset, indicating that more negative targets were correctly detected from the BHI dataset. Conversely, all the recall results of the BreakHis dataset outperformed those of the BHI dataset, indicating the correct detection of more positive targets in the BreakHis dataset. These can be better visualized with the confusion matrices in Figure 8.

Table 4. Patch-wise result comparison on the test datasets with the CL and the FL global models.

Datasets	BHI						BreakHis					
	Acc	Bac	F1	Pre	Rec	Spe	Acc	Bac	F1	Pre	Rec	Spe
CL+ResNet18	86.35	83.89	86.32	78.46	77.37	90.40	85.71	85.61	86.94	87.23	86.65	84.55
FL+ResNet18	85.78	82.48	85.64	79.15	73.75	91.22	84.14	84.15	85.34	86.63	84.11	84.19
CL+ResNet18+Gabor	86.33	85.19	86.47	75.92	82.15	88.23	85.47	85.42	87.00	87.40	85.93	85.00
FL+ResNet18+Gabor	86.13	82.77	85.97	80.05	73.85	91.68	86.02	86.01	87.12	88.14	86.13	85.89
CL+ResNet50	86.12	83.10	86.01	79.18	75.13	91.07	85.63	85.67	86.71	88.13	85.33	86.01
FL+ResNet50	86.40	83.92	86.36	78.59	77.35	90.48	84.91	84.90	86.08	87.22	84.97	84.84
CL+ResNet50+Gabor	85.90	84.26	86.00	76.02	79.90	88.61	86.31	86.36	87.32	89.00	85.86	86.85
FL+ResNet50+Gabor	86.57	83.78	86.49	79.60	76.42	91.15	86.32	86.31	87.40	88.44	86.39	86.52

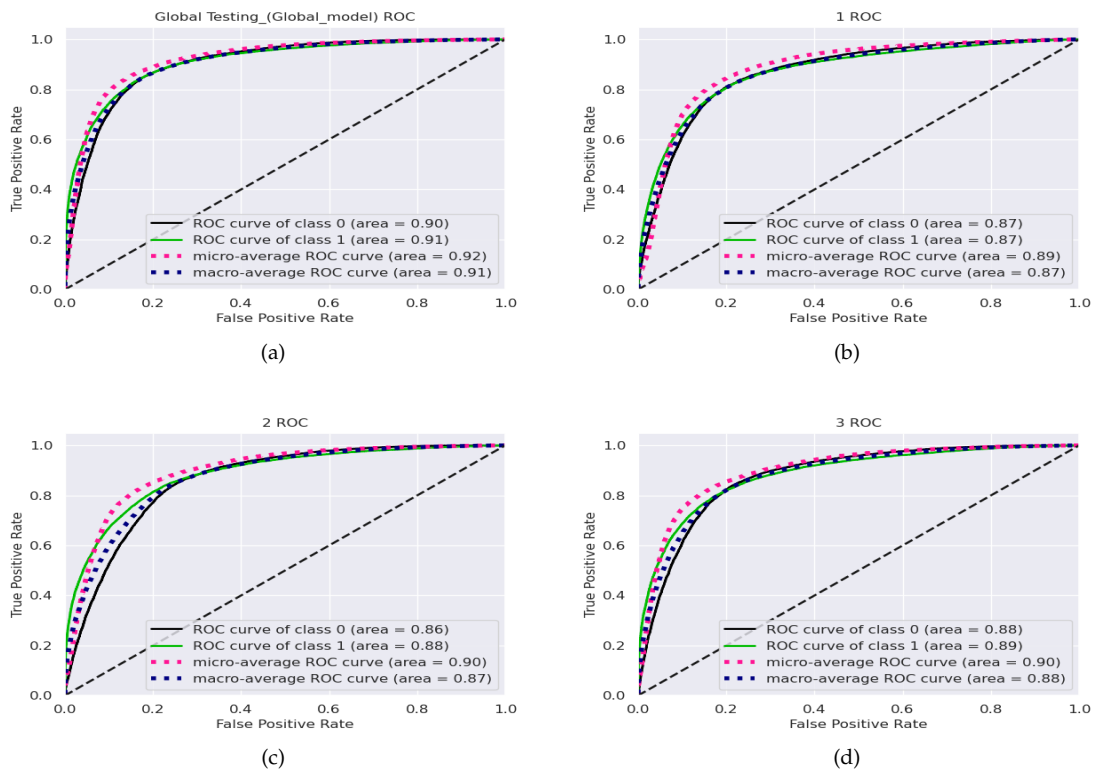


Figure 6. Receiver-operating curves for the (a) global model and the (b) client 1, (c) client 2, and (d) client 3 local models with the combined BHI test dataset and FL+ResNet50+GaborNet.

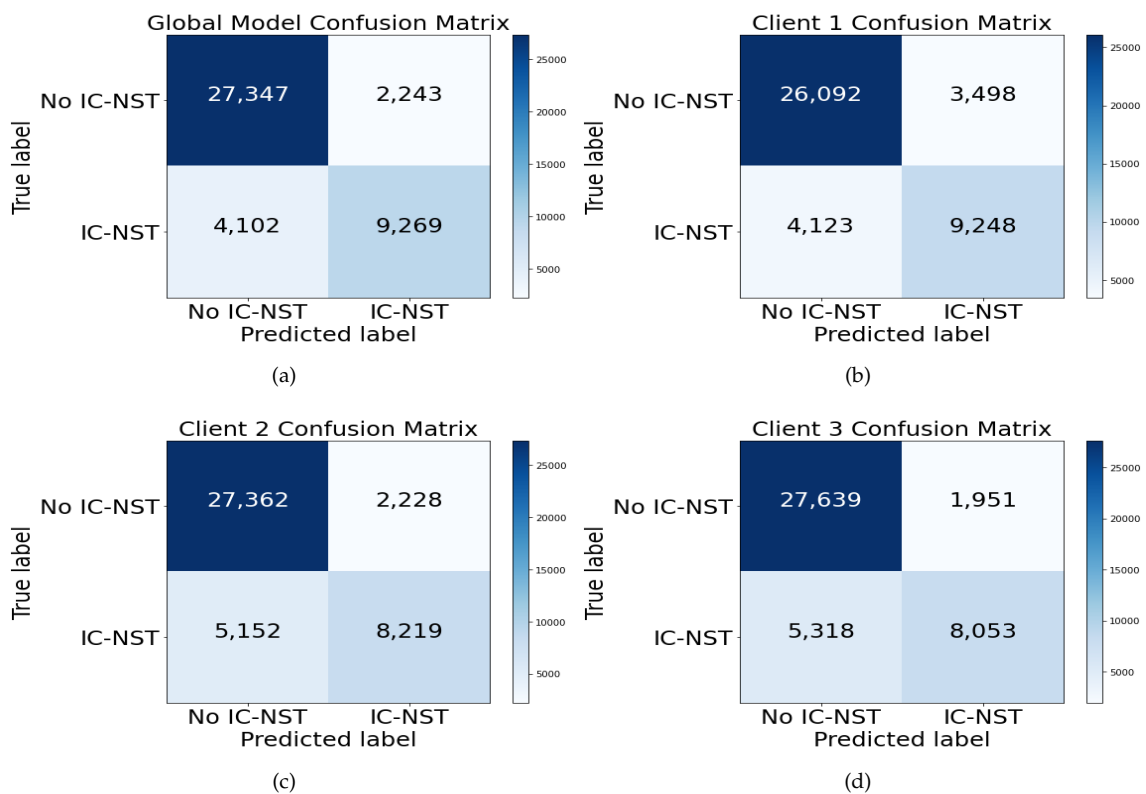


Figure 7. Confusion matrix for the (a) global model and the (b) client 1, (c) client 2, and (d) client 3 local models with the combined BHI test dataset and FL+ResNet50+GaborNet.

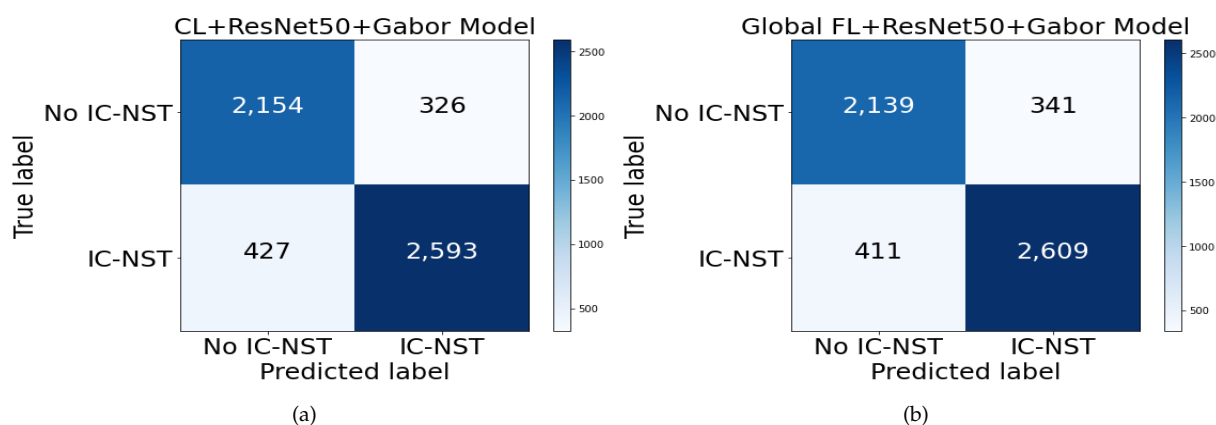


Figure 8. Confusion matrix for BreakHis dataset on (a) CL+ResNet50+Gabor and (b) FL+ResNet50+Gabor models.

4.4. Comparison with Works in Literature

The BHI dataset is the most popular public IC-NST dataset; thus, we used it to compare with the results of other models in the literature. Table 5 shows the comparison in terms of accuracy, F_1 score, and the balance accuracy measure. From the table, our FL+ResNet50+GaborNet model outperformed the existing approaches in terms of accuracy and F_1 scores. The model proposed by Reza et al. [51] performed best in terms of balanced accuracy, with a value of 85.48%, surpassing our proposed ResNet18 + GaborNet with only a small margin of 0.29%. All our proposed models combining the ResNets with the GaborNet performed higher on the F1-score than the previous methods in the table.

Table 5. Patch-wise comparison of our proposed models with studies from the literature on the BHI dataset.

Paper	Model	Published Year	Accuracy (%)	F1 (%)	BA (%)
Cruz-Roa et al. [41]	CNN	2014		71.80	84.23
Janowczyk et al. [44]	AlexNet + Resize	2016		76.48	84.68
Reza et al. [51]	SMOTE	2018		85.78	85.48
Romano et al. [23]	CNN	2019		85.41	85.28
Kumar et al. [52]	CNN	2021	83.00		
Proposed (CL)	ResNet18+GaborNet	2022	86.33	86.47	85.19
Proposed (FL)	ResNet18+GaborNet	2022	86.13	85.97	82.77
Proposed (CL)	ResNet50+GaborNet	2022	85.90	86.00	84.26
Proposed (FL)	ResNet50+GaborNet	2022	86.57	86.49	79.55

5. Conclusions and Future Works

Invasive carcinoma of no special type is a deadly cancer that can be identified and classified by the microscopic similarity between cancer cells to normal tissue. Hence, a number of researchers have proposed methods using histopathological images to detect these kinds of cancers. Deep learning is one common technique for building such models; however, they require huge amounts of data, which is not always available in the medical field due to privacy concerns of patients.

In this research, we proposed an approach that leverages federated learning with pre-trained residual neural networks to securely train with local client data without sharing their training examples. We also introduced the Gabor network to extract features from the datasets and fused the feature map with the features extracted by the residual networks. A

custom classifier network was used to obtain the predictions. We used popular evaluation metrics in the health domain to assess the performance of our approach on the BHI and BreakHis datasets. Analyzing the recall, specificity, and confusion matrices for both the BHI and BreakHis dataset indicated good generalization and the ability of the model to detect data imbalance and correctly classify the instances. Promising results were obtained without losing model quality, indicating the feasibility of federated learning for several possible applications, such as collaborative training in sensitive and privacy preservation scenarios, and interobserver variability assessment, among many others.

In future work, we intend to train each client on datasets obtained from different repositories, thereby introducing the limitations that come with such non-identically and independently distributed data sources. We will analyze it and propose approaches to limit the effects on the generalizability of the global model. We will also explore domain-based transfer learning techniques and recent advances for improving model performance on federated learning.

Author Contributions: B.L.Y.A. and H.N.M.: Conceptualization, Methodology, Resources, Writing—original draft, Formal analysis, Investigation, Software, Validation, Visualization, Writing—review, and editing. J.L.: Funding acquisition, Project administration, Supervision. M.A.H., G.U.N., J.J. and E.C.J.: Data curation, Software, Validation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China, the National High Technology Research and Development Program of China, the project of Science and Technology Department of Sichuan Province (Grant No. 2021YFG0322), the project of Science and Technology Department of Chongqing Municipality, the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202114401), Chongqing Qinchengxing Technology Co., Ltd., Chengdu Haitian Digital Technology Co., Ltd., Chengdu Chengdian Network Technology Co., Ltd., Chengdu Civil-military Integration Project Management Co., Ltd., and Sichuan Yin Ten Gu Technology Co., Ltd.

Institutional Review Board Statement: Ethical review and approval were waived for this study because this study only makes use of publicly available data.

Informed Consent Statement: Not applicable.

Data Availability Statement: For our study, a publicly accessible dataset of breast histopathology images was used. A total of 277,524 patches were obtained from whole slide images (WSI) of subjects. 198,738 patches belong to the IC-NST negative class, while 78,786 patches belong to the IC-NST positive class [43,44]. A second dataset named Breast Cancer Histopathological (BreakHis) was used to perform ablation studies to validate the proposed models. The dataset consists of 2480 benign and 5429 malignant breast tumor tissue samples of different magnifying factors obtained from 82 patients. We used all the benign samples and 3020 malignant samples. The dataset is publicly available on the Laboratório Visão Robótica e Imagem website [45–47].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IC-NST	Invasive carcinoma of no special type
IDC	Invasive ductal carcinoma
NOS	Breast cancer not otherwise specified
BRCA1/2	Breast cancer gene 1/2
CAD	Computer-aided diagnosis/detection
ML	Machine learning
DL	Deep learning
TL	Transfer learning
CL	Centralized learning
FL	Federated learning

C_1, C_2, C_3	Client 1, 2 and 3
CNN	Convolutional neural network
ResNet	Residual network
GaborNet	Gabor network
ReLU	Rectified linear unit
SGD	Stochastic gradient descent
LR	Learning rate
BHI	Breast histopathology image
BreakHis	Breast cancer histopathological
SOB	Surgical (open) biopsy
HE	Hematoxylin and eosin
Acc	Accuracy
Bac	Balance accuracy
F1	F1-score
Pre	Precision
Rec	Recall
Spe	Specificity
TP	True positives
TN	True negatives
FP	False positives
FN	False negatives
ROC	Receiver-operating curve
AUC-ROC	Area under the receiver-operating curve

References

1. Cruz-Roa, A.; Gilmore, H.; Basavanthally, A.; Feldman, M.; Ganesan, S.; Shih, N.N.; Tomaszewski, J.; González, F.A.; Madabhushi, A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci. Rep.* **2017**, *7*, 46450. [CrossRef] [PubMed]
2. Klimov, S.; Miligy, I.M.; Gertych, A.; Jiang, Y.; Toss, M.S.; Rida, P.; Ellis, I.O.; Green, A.; Krishnamurti, U.; Rakha, E.A.; et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (DCIS) recurrence risk. *Breast Cancer Res.* **2019**, *21*, 83. [CrossRef] [PubMed]
3. Golatkar, A.; Anand, D.; Sethi, A. Classification of breast cancer histology using deep learning. In Proceedings of the International Conference Image Analysis and Recognition, Póvoa de Varzim, Portugal, 27–29 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 837–844.
4. Haq, A.U.; Li, J.P.; Ali, Z.; Khan, I.; Khan, A.; Uddin, M.I.; Agbley, B.L.Y.; Khan, R.U. Stacking approach for accurate Invasive Ductal Carcinoma classification. *Comput. Electr. Eng.* **2022**, *100*, 107937. [CrossRef]
5. Chugh, G.; Kumar, S.; Singh, N. Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis. *Cogn. Comput.* **2021**, *13*, 1451–1470. [CrossRef]
6. Gage, M.; Wattendorf, D.; Henry, L. Translational advances regarding hereditary breast cancer syndromes. *J. Surg. Oncol.* **2012**, *105*, 444–451. [CrossRef]
7. Diaz-Zabala, H.J.; Ortiz, A.P.; Garland, L.; Jones, K.; Perez, C.M.; Mora, E.; Arroyo, N.; Oleksyk, T.K.; Echenique, M.; Matta, J.L.; et al. A recurrent BRCA2 mutation explains the majority of hereditary breast and ovarian cancer syndrome cases in Puerto Rico. *Cancers* **2018**, *10*, 419. [CrossRef]
8. Madariaga, A.; Lheureux, S.; Oza, A.M. Tailoring ovarian cancer treatment: implications of BRCA1/2 mutations. *Cancers* **2019**, *11*, 416. [CrossRef]
9. Warner, E. Screening BRCA1 and BRCA2 mutation carriers for breast cancer. *Cancers* **2018**, *10*, 477. [CrossRef]
10. Breast Cancer Research Highlights. Available online: <https://www.cancer.org/research/acs-research-highlights/breast-cancer-research-highlights.html> (accessed on 15 May 2022).
11. Sinn, H.P.; Kreipe, H. A brief overview of the WHO classification of breast tumors. *Breast Care* **2013**, *8*, 149–154. [CrossRef]
12. Invasive Ductal Carcinoma (IDC). Available online: <https://www.breastcancer.org/types/invasive-ductal-carcinoma> (accessed on 15 May 2022)
13. Collins, L.C.; Connolly, J.L.; Page, D.L.; Goulart, R.A.; Pisano, E.D.; Fajardo, L.L.; Berg, W.A.; Caudry, D.J.; McNeil, B.J.; Schnitt, S.J. Diagnostic agreement in the evaluation of image-guided breast core needle biopsies: Results from a randomized clinical trial. *Am. J. Surg. Pathol.* **2004**, *28*, 126–131. [CrossRef]
14. Kanavati, F.; Tsuneki, M. Breast invasive ductal carcinoma classification on whole slide images with weakly-supervised and transfer learning. *Cancers* **2021**, *13*, 5368. [CrossRef] [PubMed]

15. Preuss, K.; Thach, N.; Liang, X.; Baine, M.; Chen, J.; Zhang, C.; Du, H.; Yu, H.; Lin, C.; Hollingsworth, M.A.; et al. Using Quantitative Imaging for Personalized Medicine in Pancreatic Cancer: A Review of Radiomics and Deep Learning Applications. *Cancers* **2022**, *14*, 1654. [[CrossRef](#)] [[PubMed](#)]
16. Oh, H.S.; Lee, B.J.; Lee, Y.S.; Jang, O.J.; Nakagami, Y.; Inada, T.; Kato, T.A.; Kanba, S.; Chong, M.Y.; Lin, S.K.; et al. Machine Learning Algorithm-Based Prediction Model for the Augmented Use of Clozapine with Electroconvulsive Therapy in Patients with Schizophrenia. *J. Pers. Med.* **2022**, *12*, 969. [[CrossRef](#)] [[PubMed](#)]
17. Awassa, L.; Jdey, I.; Dhahri, H.; Hcini, G.; Mahmood, A.; Othman, E.; Haneef, M. Study of Different Deep Learning Methods for Coronavirus (COVID-19) Pandemic: Taxonomy, Survey and Insights. *Sensors* **2022**, *22*, 1890. [[CrossRef](#)]
18. Arabahmadi, M.; Farahbakhsh, R.; Rezazadeh, J. Deep Learning for Smart Healthcare—A Survey on Brain Tumor Detection from Medical Imaging. *Sensors* **2022**, *22*, 1960. [[CrossRef](#)]
19. Agbley, B.L.Y.; Li, J.; Haq, A.U.; Bankas, E.K.; Ahmad, S.; Agyemang, I.O.; Kulevome, D.; Ndiaye, W.D.; Cobbinah, B.; Latipova, S. Multimodal Melanoma Detection with Federated Learning. In Proceedings of the 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 17–19 December 2021; pp. 238–244. [[CrossRef](#)]
20. Ahmed, S.M.; Mstafa, R.J. A Comprehensive Survey on Bone Segmentation Techniques in Knee Osteoarthritis Research: From Conventional Methods to Deep Learning. *Diagnostics* **2022**, *12*, 611. [[CrossRef](#)]
21. Kang, M.; Oh, J.H. Editorial of Special Issue “Deep Learning and Machine Learning in Bioinformatics”. *Int. J. Mol. Sci.* **2022**, *23*, 6610. [[CrossRef](#)]
22. Lee, J.; Chung, S.W. Deep Learning for Orthopedic Disease Based on Medical Image Analysis: Present and Future. *Appl. Sci.* **2022**, *12*, 681. [[CrossRef](#)]
23. Romano, A.M.; Hernandez, A.A. Enhanced deep learning approach for predicting invasive ductal carcinoma from histopathology images. In Proceedings of the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 25–28 May 2019; pp. 142–148.
24. Brancati, N.; De Pietro, G.; Frucci, M.; Riccio, D. A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. *IEEE Access* **2019**, *7*, 44709–44720. [[CrossRef](#)]
25. Liu, W.; Zhang, R.; Ling, Y.; Tang, H.; She, R.; Wei, G.; Gong, X.; Lu, Y. Automatic recognition of breast invasive ductal carcinoma based on terahertz spectroscopy with wavelet packet transform and machine learning. *Biomed. Opt. Express* **2020**, *11*, 971–981. [[CrossRef](#)]
26. Chapala, H.; Sujatha, B. ResNet: Detection of Invasive Ductal Carcinoma in Breast Histopathology Images Using Deep Learning. In Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020; pp. 60–67.
27. Celik, Y.; Talo, M.; Yildirim, O.; Karabatak, M.; Acharya, U.R. Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognit. Lett.* **2020**, *133*, 232–239. [[CrossRef](#)]
28. Yadavendra; Chand, S. A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method. *Mach. Vis. Appl.* **2020**, *31*, 46. [[CrossRef](#)]
29. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. *Cancers* **2021**, *13*, 1590. [[CrossRef](#)] [[PubMed](#)]
30. Alzubaidi, L.; Al-Shamma, O.; Fadhel, M.A.; Farhan, L.; Zhang, J.; Duan, Y. Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model. *Electronics* **2020**, *9*, 445. [[CrossRef](#)]
31. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20 th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
32. Hard, A.; Rao, K.; Mathews, R.; Ramaswamy, S.; Beaufays, F.; Augenstein, S.; Eichner, H.; Kiddon, C.; Ramage, D. Federated learning for mobile keyboard prediction. *arXiv* **2018**, arXiv:1811.03604.
33. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The future of digital health with federated learning. *NPJ Digit. Med.* **2020**, *3*, 119. [[CrossRef](#)]
34. Liu, B.; Ding, M.; Shaham, S.; Rahayu, W.; Farokhi, F.; Lin, Z. When machine learning meets privacy: A survey and outlook. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [[CrossRef](#)]
35. Cilia, N.D.; D’Alessandro, T.; De Stefano, C.; Fontanella, F.; Molinara, M. From Online Handwriting to Synthetic Images for Alzheimer’s Disease Detection Using a Deep Transfer Learning Approach. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 4243–4254. [[CrossRef](#)]
36. Panayides, A.S.; Amini, A.; Filipovic, N.D.; Sharma, A.; Tsafaris, S.A.; Young, A.; Foran, D.; Do, N.; Golemati, S.; Kurc, T.; et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1837–1857. [[CrossRef](#)]
37. Strodthoff, N.; Wagner, P.; Schaeffter, T.; Samek, W. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1519–1528. [[CrossRef](#)]

38. Bhowal, P.; Sen, S.; Yoon, J.H.; Geem, Z.W.; Sarkar, R. Choquet Integral and Coalition Game-Based Ensemble of Deep Learning Models for COVID-19 Screening From Chest X-Ray Images. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 4328–4339. [[CrossRef](#)] [[PubMed](#)]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
40. Agyemang, I.O.; Zhang, X.; Adjei-Mensah, I.; Agbley, B.L.Y.; Fiasam, L.D.; Mawuli, B.C.; Sey, C. Accelerating Classification on Resource-Constrained Edge Nodes Towards Automated Structural Health Monitoring. In Proceedings of the 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 17–19 December 2021; pp. 157–160. [[CrossRef](#)]
41. Cruz-Roa, A.; Basavanthally, A.; González, F.; Gilmore, H.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; Madabhushi, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. in Proceedings of the Medical Imaging 2014: Digital Pathology, San Diego, CA, USA, 15–20 February 2014; Volume 9041, p. 904103.
42. Fan, X.; Yao, Q.; Cai, Y.; Miao, F.; Sun, F.; Li, Y. Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1744–1753. [[CrossRef](#)] [[PubMed](#)]
43. Breast Histopathology Images. <https://www.kaggle.com/paultimothymooney/breast-histopathology-images> (accessed on 4 February, 2022)
44. Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Inform.* **2016**, *7*, 29. [[CrossRef](#)] [[PubMed](#)]
45. Breast Cancer Histopathological Database (BreakHis). <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/> (accessed on 28 June, 2022)
46. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2560–2567.
47. Spanhol, F.A.; Oliveira, L.S.; Cavalin, P.R.; Petitjean, C.; Heutte, L. Deep features for breast cancer histopathological image classification. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 1868–1873.
48. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
49. Ul Haq, A.; Li, J.; Memon, M.H.; Khan, J.; Ud Din, S. A novel integrated diagnosis method for breast cancer detection. *J. Intell. Fuzzy Syst.* **2020**, *38*, 2383–2398. [[CrossRef](#)]
50. Agbley, B.L.Y.; Li, J.; Haq, A.; Cobbinah, B.; Kulevome, D.; Agbefu, P.A.; Eleeza, B. Wavelet-based cough signal decomposition for multimodal classification. In Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 18–20 December 2020; pp. 5–9.
51. Reza, M.S.; Ma, J. Imbalanced histopathological breast cancer image classification with convolutional neural network. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018, pp. 619–624.
52. Kumar, K.; Saeed, U.; Rai, A.; Islam, N.; Shaikh, G.M.; Qayoom, A. IDC Breast Cancer Detection Using Deep Learning Schemes. *Adv. Data Sci. Adapt. Anal.* **2020**, *12*, 2041002. [[CrossRef](#)]