RESEARCH ARTICLE

Joint Federated Learning Using Deep Segmentation and the Gaussian Mixture Model for Breast Cancer Tumors

NGUYEN TAN Y[ID][1]**, PHAM DUC LAM**[ID][2]**, VO PHUC TINH**[ID][1]**, DUY-DONG LE**[ID][3]**, NGUYEN HOANG NAM**[ID][4]**, AND TRAN ANH KHOA**[ID][4]

[1]Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 729000, Vietnam
[2]Faculty of Engineering and Technology, Nguyen Tat Thanh University, Ho Chi Minh City 70000, Vietnam
[3]University of Economics Ho Chi Minh City, Ho Chi Minh City 70000, Vietnam
[4]Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 729000, Vietnam

Corresponding author: Tran Anh Khoa (trananhkhoa@tdtu.edu.vn)

ABSTRACT Medical image segmentation is crucial for deep learning (DL) applications in clinical settings. Ensuring accurate segmentation is challenging due to diverse image sources and significant data sharing and privacy concerns in centralized learning setups. To address these challenges, we introduce a novel federated learning (FL) framework tailored for breast cancer. First, we use random regions of interest (ROIs) and bilinear interpolation to determine pixel color intensity based on neighboring pixels, addressing data inconsistencies from heterogeneous distribution parameters and increasing dataset size. We then employ the UNet model with a deep convolutional backbone (Visual Geometry Group [VGG]) to train the augmented data, enhancing recognition during training and testing. Second, we apply the Gaussian Mixture Model (GMM) to improve segmentation quality. This approach effectively manages distinct data distributions across hospitals and highlights images with a higher likelihood of tumor presence. Compared to other segmentation algorithms, GMM enhances the salience of valuable images, improving tumor detection. Finally, extensive experiments in two scenarios, federated averaging (FedAvg) and federated batch normalization (FedBN), demonstrate that our method outperforms several state-of-the-art segmentation methods on five public breast cancer datasets. These findings validate the effectiveness of our proposed framework, promising significant benefits for the community and society.

INDEX TERMS Federated learning, meta-global, Gaussian mixture model, segmentation, breast tumor.

I. INTRODUCTION

Medical image segmentation is a fundamental task in automated medical image analysis, underpinning essential applications including diagnosis, prognosis, treatment planning, image reconstruction, and evaluating patient treatment responses. This is achieved using a variety of methods ranging from basic to advanced, such as digital signal processing and machine learning [1], [2], [3]. Many diverse

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Zaman Chowdhury.

types of cancer exist; however, breast cancer is one of the most dangerous diseases affecting women, and there is still no natural cure [4]. Many studies have evaluated breast cancer as a complex disease characterized by the uncontrolled growth of cells in the breast [5]. Breast cancer presents in many forms depending on the cells that become cancerous and can spread beyond the breast through the blood and lymphatic vessels, affecting other body parts. To determine whether symptoms indicate malignant or benign breast cancer, doctors conduct a thorough physical examination and diagnostic tests using various methods to understand the underlying cause of the

tumor. The following imaging tests to diagnose breast cancer are standard as mammography and ultrasound in [6] and [7].

1) **Mammography**: Mammography is a widely used method for visualizing the internal structures of the breast. If a doctor suspects a lump or an area is suspicious, they may order a mammogram. Further tests may be recommended based on the findings.

2) **Ultrasound**: A breast ultrasound employs sound waves to generate detailed images of the tissues deep within the breast, aiding doctors in distinguishing between solid masses, tumors, and benign cysts. With these images as a guide, doctors can tailor appropriate treatment regimens to each patient.

During digital mammography screening, each breast is typically imaged from two views: the medial oblique and the cranial-caudal view [6]. Ultrasound imaging is often used to determine if an abnormality detected during a mammogram or physical exam is a fluid-filled cyst or a solid tumor. However, breast ultrasound is not usually used for cancer screening because it can miss early signs of the disease [7]. The results from these diagnostic methods allow for a thorough assessment of the breast tissue, supporting early detection and accurate diagnosis of abnormalities. However, current datasets are curated for specific use cases, locations, and devices, often focusing only on manual tumor image segmentation.

In recent decades, with the advent of artificial intelligence, basic machine learning (ML) [8], [9], DL [10] models have been proven effective in detecting breast cancer and supporting early diagnosis, thereby improving patient survival. Furthermore, DL requires less human intervention for feature extraction compared to classical ML techniques. However, concerns about data segregation, sharing, and inconsistency have led to a loss of flexibility and increased privacy concerns in medical treatment facilities.

Many recent studies have proposed collaborative learning to overcome these limitations in DL, explicitly focusing on FL [11], [12], which enables the combined training of a global model across medical treatment facilities without sharing sensitive data. This approach enhances model efficiency and generalizability by synthesizing knowledge from datasets managed at different imaging centers. The ability of FL to improve privacy and security is crucial, especially in sensitive fields, such as medicine, ensuring that personal information remains confidential. Additionally, FL offers advantages, such as improved overall efficiency, decentralization, and the potential for edge computing, making it easily deployable in hospitals without complex and expensive server systems.

In summary, early detection and timely diagnosis of breast cancer play a crucial role in improving prognosis and treatment outcomes [11], [12]. Regular checkups and timely medical care are vital for everyone, with a particular emphasis on women over 40 years old. Proactive measures, such as decentralization and FL, can significantly contribute to advancing medical image diagnosis and segmentation, leading to better healthcare outcomes in managing breast cancer.

This research introduces a computer vision technique called image segmentation, which classifies each pixel in an image as either tumor or background. Our study pioneers the use of FL for image segmentation in breast cancer. The first highlight is the use of rich datasets, both discrete and mixed, to ensure accuracy and reliability, and to facilitate sharing findings within the research community. Second, experiments and evaluations on non-independent and identically distributed (Non-IID) datasets, combined with techniques like random regions of interest (ROIs) and bilinear interpolation, demonstrate that the UNet3+ model is suitable for the breast cancer imaging landscape in FL. Third, this study employs clustering algorithms to improve segmentation image quality. Finally, comparison results show that our method achieves higher dice coefficient (DC) results than the baseline [13] and FedAvg (using the UNet3+ model) [14] methods across several different data types.

The following section introduces the outstanding contributions of this research with detailed explanations:

1) We propose a novel FL framework for breast cancer image segmentation, aiming to improve tumor detection (Section IV-A). The objective is to segment breast images from x-ray and ultrasound data using distributed nodes, each equipped with a unique local dataset. We evaluate globally trained models against individually trained models and a centrally aggregated model across diverse data types. The evaluation results demonstrate that the proposed framework is highly reliable and ready to support hospitals in diagnosing breast cancer.

2) We propose an effective solution for addressing data inconsistency caused by nonuniform distribution parameters within a dataset, thereby increasing its size. Section IV-B provides a detailed explanation. The approach involves using random regions of interest (ROIs) and bilinear interpolation techniques to determine the color intensity of a fixed pixel based on the values of four diagonally arranged neighboring pixels. We employ the UNet model in Section IV-C with a deep convolutional backbone (Visual Geometry Group [VGG]) to train the augmented data, enhancing recognition during both the training and testing phases.

3) We apply the Gaussian Mixture Model (GMM) in Section IV-D to improve image segmentation quality by leveraging both generative and discriminative learning. First, it addresses the issue of separate data distribution across hospitals. Second, it handles multiple levels of data distribution, allowing the GMM to highlight more valuable images whose high prominence corresponds to the likelihood of tumor occurrence. This approach yields consistently high-quality segmentation results, providing reliable information regarding tumor locations, as detailed in Algorithm 1.
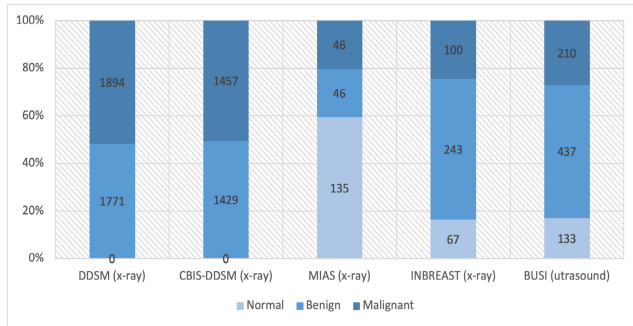
FIGURE 1. Summary of the number of tumor types by disease family of different breast cancer data.

4) We introduce two evaluation scenarios, federated averaging (FedAvg) and federated batch normalization (FedBN), to address challenges in training models within federated environments, as presented in Section IV-E. To assess the reliability and effectiveness of this study, we employ five performance indicators and three loss indicators, as outlined in Section IV-F and Algorithm 2. The results consistently demonstrate significantly higher performance compared to the baseline-centralized method [13] and FedAvg + UNet3+ [14] across various assessments. These outcomes substantiate the effectiveness of the proposed research direction, promising substantial benefits for the community and society.

In summary, this study investigates the advantages of FL in medical image segmentation tasks, especially for diverse and heterogeneously distributed datasets from various imaging centers. The identified technical objectives, while beneficial, raise questions regarding performance, which we aim to address:

- **RQ1:** Do GMM and UNet accurately segment tumors?
- **RQ2:** Does the UNet3+ model improve performance with limited datasets?

The rest of the paper is structured as follows. Section II provides an overview of related work in this domain. Next, Section III outlines the datasets and their associated challenges. The proposed methodology is detailed in Section IV. Sections V and VI cover the implementation details, experimental results, and efficiency and effects of the meta-global model. Finally, Section VII summarizes the critical findings and concludes the paper.

II. RELATED WORKS

Medical image classification is extremely important today because it helps medical professionals and doctors provide additional image-based diagnostic support using algorithms in image processing and computer vision [2], [3]. Recent advances in DL models have achieved notable progress in semantic image segmentation [15], [16], [17]. The application of DL models to breast tumor segmentation has attracted significant attention in the research community [18]. Notable studies have extensively used ultrasound imaging

and digital mammography, with some incorporating magnetic resonance imaging to segment breast tumors [19], [20]. A study has demonstrated that UNet and its variants have emerged as popular architectures for this task [21].

Despite these advances, breast tumor segmentation remains an open challenge due to complexities such as shape, boundaries, curvature, intensity, and texture. Current DL models focus on pixel-level predictions, often considering interactions between pixels and semantic relationships in the data. The attention mechanism in DL networks is crucial for recognizing distinguishing features [22]. Soft attention is preferred because it is easy to train and optimize [23].

Several studies based on federated learning (FL) have demonstrated its effectiveness in the medical field. For example, one study [24] predicted mortality and length of hospital stay using distributed electronic health records from many different hospitals. Indeed, data sharing between hospitals brings great benefits to hospitals, doctors, and patients, demonstrating the great potential of FL. Another study [25] introduced computer vision and FL solutions to increase data labeling efficiency, showing superior results compared to FedAvg, especially with the Covid-19 image dataset, which is very encouraging in the clinical medical field.

In the field of breast cancer, there have been many studies on image segmentation based on DL models. However, there is little applied research in the FL setting, where hospitals participate in the FL process by sharing imaging features rather than data. The research in [26] focuses on image classification, introducing a new memory-aware curriculum learning method for the FL environment but using only the mammography dataset for classification. The authors of the large-scale computational pathology study [27] aimed to demonstrate the feasibility and results of FL on breast cancer datasets to show its superiority in ensuring privacy using multi-instance learning.

Another study called FedMix [28] proposes that each client updates its federated model by integrating and effectively using all available labeled data, from strong pixel-level labels to the weakest image-level class labels. Based on these local models, the authors proposed an adaptive weight assignment process for local clients, where each client learns to aggregate weights during the global model update process.

After evaluating the studies and the related DL and FL models used in visual segmentation of breast cancer images, we found it necessary to develop new solutions in breast cancer image segmentation using FL. This study presents the idea of increasing focus on important regions in breast tumor segmentation, using UNet for tumor detection, and exploring a decentralized learning model, which poses significant distributed challenges. Additionally, the use of FL brings many benefits, as hospitals can share data, improve the quality of image diagnosis, and maintain patient information confidentiality. The authors believe that this study pioneers breast cancer research, bringing significant benefits to the community and society.

TABLE 1. Summary of breast cancer data types.

Details	DDSM	CBIS-DDSM	MIAS	INBREAST	BUSI
Number of Images	3665 mammography images	2886 mammography images	227 mammography images	410 mammography images	780 breast ultrasound images
Image Modalities	Digitized film mammography (FM) images and images from full-field digital mammography (FFDM).	Digital mammography (DM) and film mammography (FM).	Digitized film mammography (FM) images and images from full-field digital mammography (FFDM).	Digital Imaging and Communications in Medicine (DICOM) format.	Digital Imaging with JPEG format.
Image Categories	Benign (non-cancerous) or malignant (cancerous).	Benign (non-cancerous) or malignant (cancerous).	Masses and microcalcifications	Masses, microcalcifications, and architectural distortions.	Lesions, tumors, and other abnormalities.
Image Annotations	Regions of interest (ROIs) where abnormalities are present.	Location of abnormalities.	Information about the presence of abnormalities.	Information about the presence of abnormalities.	Information about the presence of abnormalities.
Image Size and Resolution	Images have varying sizes and resolutions 5395×3131 pixels.	Images have varying sizes and resolutions 5395×3131 pixels.	Images have 1024×1024 pixels, the location and severity.	Images have 3328×4084 pixels, shape, and severity.	Images have varying sizes, from 408×288 pixels to 2531×1675 pixels.
Purpose	Breast cancer detection and diagnosis.	Detection and classification.	Research and benchmarking.	Detection, diagnosis, and image segmentation.	Presence detection, diagnosis, and image analysis.
Availability	Public 2001	Public 2016	Public 1994	Public 2011	Public 2020

III. DATASETS

Section III provides a concise overview of five crucial datasets for this research. In image processing, computer vision, and particularly deep learning (DL), the selection of input data significantly impacts the research process and outcomes. This section introduces the standard and relevant raw data used in the study. Breast cancer is the most prevalent and deadliest cancer worldwide. In 2022, an estimated 287,500 new cases of breast cancer were diagnosed in women in the United States, including 51,400 new cases of noninvasive breast cancer.

- The University of South Florida's Department of Radiology developed the Digital Database for Screening Mammography (DDSM), which includes 2620 scanned mammograms with verified patient information. This dataset supports the development and evaluation of algorithms for breast cancer detection and diagnosis [29].
- The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is derived from the DDSM dataset and improved by a trained mammographer. It includes only abnormal images for computer-aided classification and detection tasks [30].
- The Mammographic Image Analysis Society (MIAS) Digital Mammogram Database, developed with the University of Aberdeen, includes digital mammograms with annotations by expert radiologists. It is widely used in breast cancer research [31].
- The INBREAST database, created by the University of Santiago de Compostela, contains 410 mammography images from 2012, categorized into various classifications and presented in DICOM format [32].
- The Breast Ultrasound Image (BUSI) dataset consists of ultrasound images from 600 women aged 25 to 75, providing a comprehensive resource for breast cancer

detection and diagnosis research [33]. Table 1 summarizes the datasets and data calibration in this study.

Researchers enhance the robustness and generalizability of breast cancer models by integrating datasets from multiple hospitals. Using medical datasets requires strict adherence to ethical considerations for patient privacy. An analysis of five datasets revealed imbalances due to heterogeneous distribution (Fig. 1). Addressing these imbalances involves preprocessing and employing segmentation algorithms. In the data processing phase, DDSM and CBIS-DDSM data show a 50% abnormality rate. Similar patterns are seen in MIAS and INBREAST datasets, with 80% normal and 20% abnormal data. Training data across hospitals are weighted for balanced representation. In BUSI ultrasound data, 70% are labeled normal, and 30% show abnormalities.

Handling data imbalance in FL focuses on effective learning rather than simple data ratios. Sample balancing is not applied due to limited data and varying tumor detection targets. Distribution disparities are addressed through training with a DL model and controlled with GMM in postprocessing. Completely trusting the global model is impractical, requiring independent data control at each hospital. Retraining with fine-tuning, a form of transfer learning, uses pretrained model weights on hospital data. After examining common breast cancer datasets, the focus shifts to processing for high-quality datasets and accurate evaluations, enabling practical applications in Section IV.

IV. METHODOLOGY
A. PROPOSED FRAMEWORK OVERVIEW

This section delineates the proposed approach within the FL framework for breast cancer segmentation. Section IV-A expounds on the comprehensive framework structure. Section IV-B details the data processing solution encapsulated
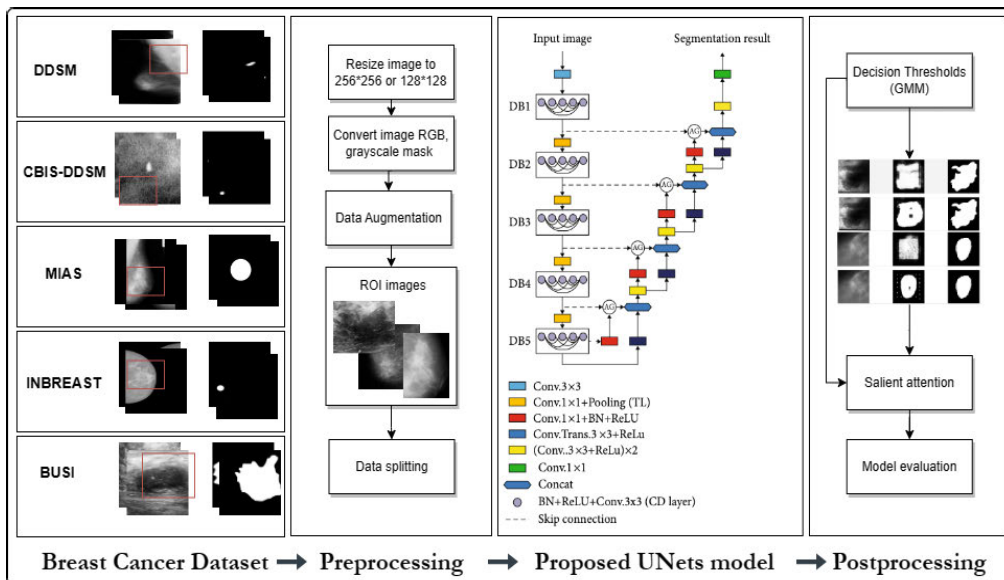
FIGURE 2. Proposed training strategy: Utilize breast cancer datasets (DDSM, CBIS-DDSM, MIAS, INBREAST, BUSI), preprocess by converting DICOM to 2D images, resizing, augmenting data, extracting ROI, and splitting data; train UNet with hybrid segmentation loss; and postprocess using GMM decision thresholds, focusing on salient regions and evaluating performance.

in Algorithm 1—a vital core element of this research. Section IV-C introduces the selected model architecture and its practical application. The ensuing postprocessing step is discussed in Section IV-D and succinctly detailed in Algorithm 2 to explain the FL framework implementation. This section elucidates the pivotal choice of GMM for determining the segmentation threshold, ensuring precision. The detailed steps are outlined in Fig. 2. After identifying and explaining the critical algorithms, the application of scenarios in real-world situations is discussed. Section IV-E introduces two primary scenarios, FedAvg and FedBN, tailored for different clients, each representing a hospital with distinct datasets. Section IV-F delves into the selected methods for evaluating the research results. These evaluation methods shape the research and facilitate practical applications in research centers and hospitals with diverse breast cancer data. Sections V and VI detail the simulation implementation and the experimental results.

B. PREPROCESSING

1) REGION OF INTEREST (ROI) EXTRACTION

In medical images for breast cancer analysis, as detailed in Section III, we identified the imperative need to enhance image datasets. Typically, these datasets are relatively small, and annotations tend to be sparse in comparison to natural images [34]. Hence, this section introduces data augmentation methods for the datasets outlined in Table 1. While recent data augmentation methods, such as Cutout [35], Mixup [36], and their derivatives [37] and combinations [38], have predominantly employed a fixed set of image transformations, the proposed approach addresses

the unique challenges posed by the necessity to augment images with multiple annotations targeting various shapes, including simultaneously segmenting different object regions and their contours.

The proposed solution is applicable to all the included datasets (DDSM, CBIS-DDSM, MIAS, INBREAST, and BUSI) given their relatively large scans, with an average height ranging from 2531 to 5295 pixels and an average width from 1675 to 3131 pixels. Images were extracted from full-sized scans and downsized to 256×256 or 128×128, and the ROIs were extracted using masks.

Each preprocessed image was contextually extracted based on a bilinear interpolation visual algorithm. The data were expanded by randomly positioning ROIs in the image to augment the dataset. This process involves bilinear interpolation, allowing the conversion of full-size images to downsized images with pixel matching. This method ensures the accurate calculation of the intensity for each pixel, unlike other interpolation techniques, such as nearest neighbor interpolation [39] and binary interpolation. In this study, we use the values of the nearest four pixels located diagonally from a given pixel to determine the appropriate color intensity values for that pixel, as illustrated in Fig. 3.

Suppose we want to find the value of an unknown function f at the coordinate point (x, y) and that we know the value of f at four coordinate points $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, and $Q_{22} = (x_2, y_2)$.

First, we performed linear interpolation in the x direction.

$$R1 = Q11(1 - w_x) + Q21w_x,$$
$$R2 = Q12(1 - w_x) + Q22w_x,$$
$$P = (1 - w_y)R1 + R2w_y,$$

$$= Q11(1-w_x)(1-w_y) + Q21w_x(1-w_y)$$
$$+ Q12(1-w_x)w_y + Q22w_xw_y \qquad (1)$$

where $w_x = \frac{x-x_1}{x_2-x_1}$ and $w_y = \frac{y-y_1}{y_2-y_1}$. x and y are pixel points that are reduced in size. We achieved the same results when performing the interpolation first in the y direction and then in the x direction.

Next, the DDSM images contain variable-sized borders and white patches to hide the patient's personal information. The MIAS dataset has a background data area at the border and does not contain mammograms. The BUSI dataset changes at different rates, mistaking data for a tumor and ignoring essential information. The DDSM image in the MIAS dataset was cropped by 5% on each side, with a bilinear size reduction and segmentation mask template combined into 10-pixel to 50-pixel source removal data to remove borders and redundancy. The INBREAST dataset contains four distinct classes: mass, calcifications, asymmetries, and distortions, with no background tissue descriptors, small ROI are dilated by 10% edges. Moreover, BUSI easily focuses on the tumor area, so the abnormal points are not at the edge based on the mask layer provided. Determining suitable augmentation solutions for datasets is also a challenge and is addressed in this study.

2) PERSPECTIVE TRANSFORMATION

In the realm of medical imaging, especially in breast tumor analysis, the examination of abnormal tumor data is crucial for distinguishing between benign and malignant tumors. Achieving homogeneity across all datasets is vital; thus, a minimum fixed padding is applied around the ROI during the preprocessing stage. Each ROI is extracted and dilated 30-pixel from the four corners of the tumor's bounding box, with the ROI randomly positioned within these bounds. Extracting each ROI in an area larger than its size ensures that tumors significantly larger than the background are not overlooked. The size of this dataset is based on the abnormal volume obtained through zooming in and randomly cropping images according to the tumor characteristics. In the data processing step illustrated in Fig. 2, the images containing overlay text are preserved or predominantly transformed to have a black background. This process is done to retain tumor formation while considering the complexity of the semantic segmentation model, which faces challenges due to diverse prediction distributions.

3) RESIZING

To enhance the specificity of the dataset, additional data augmentation tools are employed to improve performance, recognize diverse tumor morphologies, and prevent overfitting during deep neural network training. The UNet model discussed in Section IV-C uniformly enhances the image and mask. These enhancements include changes in brightness and contrast by 15%, shifts, rotations, image scaling by 15 degrees, a 50% image flip, using PadIfNeeded to expand the image and create new blank areas if necessary, cropping
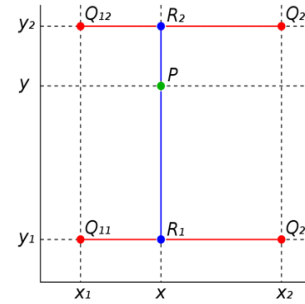

FIGURE 3. Four red dots represent data points, and the green dot signifies the point of interpolation.

random portions of the input, adding 20% Gaussian noise to the input image, and performing a random four-point 50% perspective transform of the information using the perspective operation. Algorithm 1 details the data processing steps.

The image intensity and mask values are normalized between 0 and 1. A $128 \times 128 \times 3$ random foreground was extracted for hospital contexts, and the corresponding $128 \times 128 \times 1$ label is a grayscale image conveying a value of 1 for the tumor area and 0 for the background. The output includes post-processed image combinations, providing investigators with information on intensity mapping and tumor size.

4) GEOMETRIC TRANSFORMATIONS

This comprehensive, performance-optimized approach to medical image transformation operations offers a concise yet powerful tool for various computer vision tasks, including segmentation. This process includes adjustments in brightness and contrast, shifts, rotations, flips, scaling, and random cropping to diversify the dataset as in Fig. 4. These transformations ensure accurate pixel intensity and enhance the dataset by introducing variability, thereby improving the robustness and generalizability of the deep learning models.

The preprocessing step used Python as the programming language, with OpenCV and PyTorch being the primary libraries. OpenCV was used for image processing tasks, such as contour detection, perspective transformation, resizing, and augmentations.

C. MODEL ARCHITECTURE SELECTION

This study advocates the use of the UNet architecture, a widely recognized framework for semantic segmentation tasks in medical images, characterized by simple connections. It explores two variants of the UNet architecture, UNet2+ and UNet3+, which feature simple connections nested within each other and dense concatenation to extract comprehensive information across scales. The preferred backbone for these variants is the state-of-the-art UNet3+ model, leveraging the VGG architecture [40], rather than older methods, such as PSPNet, DeepLabV2, DeepLabV3, DeepLabV3, and Attention UNet, [16], [41], [42], [43], [44].

Algorithm 1 Preprocessing

Input: a given data $X = \{x_1, x_2, \ldots, x_i\}$,
$Y = \{y_1, y_2, \ldots, y_i\}$, $offset = 30$
Output:
$X' = \left\{x'_1, x'_2, \ldots, x'_i\right\}$, $Y' = \left\{y'_1, y'_2, \ldots, y'_i\right\}$
Randomly initial X
for $x, y \in zip(X, Y)$ **do**
    /*region of interest*/
    $r \leftarrow$ finding contours for x follow y box $\leftarrow$ points
    after the rectangle $r + offset$
    /*perspective transformation*/
    $M \leftarrow$ the perspective transformation matrix $\leftarrow$ box
    $x, y \leftarrow$ directly warp the rotated rectangle to get the
    straightened rectangle $\leftarrow M$
    /*resizing*/
    $x', y' \leftarrow$ from Eq. 1
    /* geometric transformations*/
    $x', y' \leftarrow$ Compose[Change brightness and
    contrast;Shift, Rotate, Flip and Scale; PadIfNeeded;
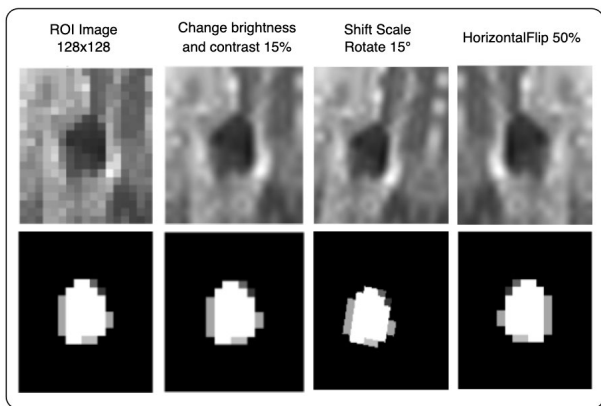   RandomCrop] $\leftarrow x', y'$.
end

FIGURE 4. Example of geometric transformations applied to segmentation images: ROI Image (128 × 128), Change brightness and contrast (15%), Shift, scale, rotate (15°), Horizontal flip (50%).

The study also includes the UNet+GM method to streamline the investigation and comparison. All networks undergo optimization using the loss function proposed in [45]. The selection of UNet3+ is crucial for the complex task of tumor segmentation, especially when dealing with substantial medical breast images and the available computational resources within the hospital. The UNet3+ model serves as an extension of the UNet architecture, building on UNet2+ by introducing full-scale bypass connections. Apart from the nested bypass connections, UNet3+ establishes connections between all encoder and decoder layers, creating a fully connected structure, ensuring that information from all levels of the encoder is accessible at each level of the decoder. The overarching goal of UNet3+ is to enhance the segmentation accuracy.

D. POSTPROCESSING

1) ESTIMATE GMM COMPONENTS

In Section IV-B, we addressed the challenges of data enrichment and model selection to delineate the segmentation region in any given image. However, as highlighted, choosing DL models to achieve highly accurate results presents its own set of challenges. Therefore, a solution is essential at the postprocessing stage. In this method, we employ the Gaussian Mixture Model (GMM). Given the impracticality of relying solely on the global model and the need for independent control of data distribution across each hospital or retraining with fine-tuning, a transfer learning approach becomes imperative. This approach emphasizes the use of pretrained models on hospital data to address the nonuniform data distribution.

The number of components in the GMM heavily depends on the output image, and no one-size-fits-all threshold exists for complete segmentation. Manual thresholding, for example, using a fixed threshold of 0.5/1.0 with the ground truth, may lead to suboptimal results, as illustrated in Fig. 5, where noise and deformities in the mammary gland affect the segmentation process. Clustering with multiple components using the GMM has proven effective in overcoming these challenges, yielding reasonable results within the range of tumor hyperintensities. The output optimization solution uses the GMM to acquire a comprehensive representation of class recognition knowledge in a discriminant feature space [46].

2) INDIVIDUAL COMPONENTS

Partly inspired by probing the data structure through intra-class clustering, the GMM [47], as implemented in this combined training scheme, offers several advantages. First, the GMM provides the dual benefits of generative and discriminative learning. Through online expectation-maximization (EM)-based productive optimization, the GMM adeptly fits the data distribution, even as the feature space expands. The feature space is trained end-to-end discriminatively, aligning with the guidance of the GMM classifier to maximize pixelwise prediction performance.

Second, the GMM's explicit modeling of the data distribution facilitates handling examples at multiple distribution levels. Saliency maps are directly input into post-production as additional information based on areas in the image with high saliency values, corresponding to an increased likelihood of tumor presence. Thus, saliency maps must exhibit suitable quality and provide reliable information about tumor locations. Conversely, poor-quality saliency maps can adversely influence model performance.

3) RECONSTRUCTION

For density estimation and clustering, we assume that the output data are generated from a mixture of multiple Gaussian distributions. Each Gaussian component in the mixture represents the density of a cluster k in the data. The GMM represents the probability density function as the total weight

of many Gaussian distributions, which is given according to the following Eq. 2:

$$p(v) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x \mid \mu_k + q_{\mu k}, \sigma_k^2 + q_{\sigma^2 k}\right) \quad (2)$$

In this equation, $q_{\mu k}$ and $q_{\sigma^2 k}$ are drawn from a uniform distribution for each component. Additionally, K is the number of Gaussian components in the mixture, and π_k is the weight (or mixing coefficient) of the k-th Gaussian component, satisfying Eq. 3:

$$\sum_{k=1}^{K} \pi_k = 1 \quad (3)$$

where $\mathcal{N}\left(x \mid \mu_k, \sigma_k^2\right)$ is a Gaussian distribution with mean μ and covariance matrix σ^2.

In the EM algorithm, GMM training involves estimating the parameters of Gaussian components. The **E (Expectation)** step calculates the probability of each data point belonging to each Gaussian component. The **M (Maximization)** step updates the parameters based on the probabilities obtained in the **E** step. The details for the postprocessing stage are explained in Algorithm 2.

Algorithm 2 Postprocessing: EM Algorithm for GMM

---

with each i **do Input**: a given data $X = \{x_1, x_2, \ldots, x_n\}$
$\qquad \pi = \{\pi_1, \pi_2, \ldots, \pi_K\}$,
Output: $\mu = \{\mu_1, \mu_2, \ldots, \mu_K\}$,
$\qquad \sigma^2 = \{\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_K\}$
Randomly initial π, μ, σ^2
for $t \in T$ **do**
$\quad$ //E step **for** $n \in N$ **do**
$\qquad$ **for** $k \in K$ **do**
$\qquad\qquad \gamma\left(z_{nk}\right) = \frac{\pi_k N\left(x_n \mid \mu_k, \sigma^2_k\right)}{\sum_{j=1}^{K} \pi_j N\left(x_n \mid \mu_j, \sigma^2_j\right)}$
$\qquad$ **end**
$\qquad y_n = \arg\max_k \gamma\left(z_{nk}\right)$
$\quad$ **end**
$\quad$ //M step **for** $k \in K$ **do**
$\qquad \mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$
$\qquad \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$
$\qquad \pi_k = \frac{1}{N}\sum_{n=1}^{N} \gamma\left(z_{nk}\right)$
$\quad$ **end**
$\quad \mu$ mixture component means use to state of environment for policy each i
$\quad \pi$ scale weight of mixture components
$\quad y$ prediction is conducted by group label for client each i
end

E. SCENARIO IMPLEMENTATIONS

This section introduces two main FL scenarios, FedAvg [13] and FedBN [48], to address challenges in training models on

decentralized data while preserving data privacy and security. A fundamental FL algorithm, FedAvg focuses on training a global model by aggregating local updates from different hospitals. However, it may face suboptimal performance due to sparse and inconsistent hospital data. A modified version of batch normalization for FL, FedBN addresses problems related to data distribution differences between hospitals by aggregating the mean and variance statistics. This approach enhances model convergence and ensures more consistent normalization during training. The FedAvg and FedBN algorithms are explained in detail in the following section, providing insight into their operation for the depicted scenario in Fig. 6.

- **Data Collection**: Data are collected from clients, each with access to their local datasets. The data remain decentralized and do not need to be shared with a central cloud.
- **Data Partitioning**: Local datasets are partitioned or sampled into smaller subsets (batches) per hospital. Each device uses local data to train the local model.
- **Initialization**: A global model is instantiated in the cloud.
- **Client Update**: Each hospital (participant) downloads the global model and trains the model locally using its dataset (DDSM, CBIS-DDSM, MIAS, INBREAST, or BUSI) and calculates model updates (UNet, as discussed in Section IV-C).
- **Model Aggregation**: Model updates from all hospitals are sent to the cloud server, where they are aggregated by calculating their average. In FedBN, the batch normalization values are kept separate for each hospital.
- **Global Model Updates**: The average model updates are applied to the global model.
- **Iterative Process**: The process of updating the client, synthesizing, and updating the global model is repeated for multiple rounds or epochs until the global model converges or reaches the desired performance level.

Sections V and VI present the simulation settings and results.

F. TRAINING STRATEGY

After completing the steps and constructing the scenario, we initiated the training process for the entire system. To assess the reliability and potential effectiveness of this study, we identified and emphasized the predicted tumors by examining the similarities between the predictions and annotated segments (ground truths). This evaluation is conducted using five performance evaluation indices and three loss indices: precision (specificity), recall (sensitivity), accuracy/range index (ACC), Dice coefficient (DC), area under the receiver operating characteristic curve, Jaccard loss (intersection over union [IoU]), focal loss, and multiscale structural similarity index measure (MS-SSIM) loss [49], [50]. Algorithm 3 provides a detailed explanation of the training strategy. All presented performance indicators are based on calculating the confusion matrix for the binary segmentation mask.
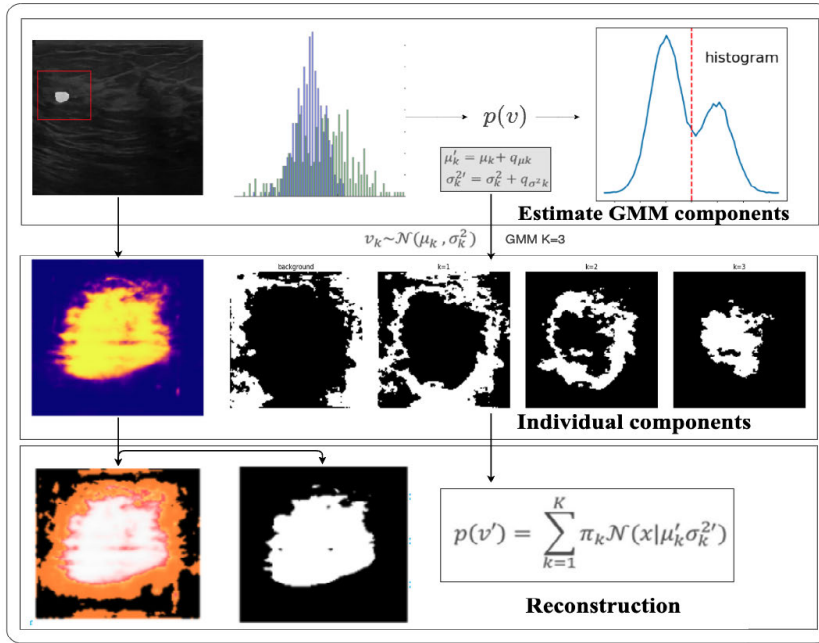
FIGURE 5. Diagram of postprocessing. This augmentation is exclusively conducted while validating the segmentation network.
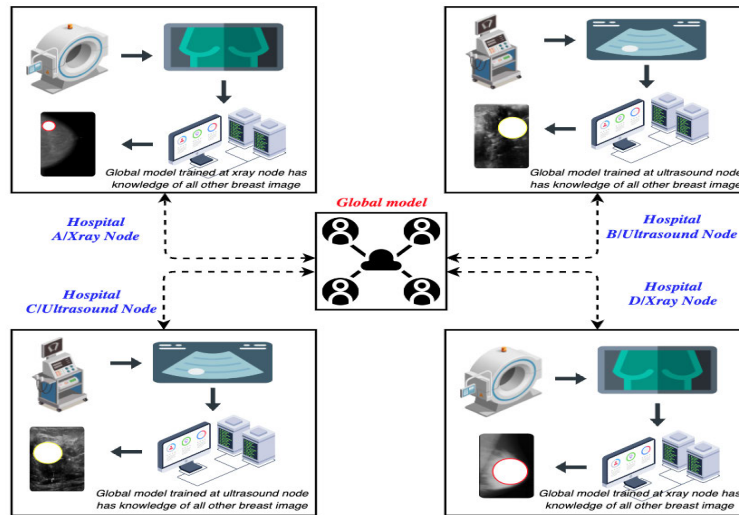

FIGURE 6. Proposed framework: Client nodes update the global meta-model, and knowledge aggregation occurs after every two iterations of the local model. The weights of the global model are shared with client models, allowing nodes to share knowledge without sharing data.

The range of values for all presented indicators extends from 0 (worst) to 1 (best). The goal is to determine the factory testing unit (FTU) mask segment for each histopathological image. Each mask is indicated by 0 for background or 1 for FTU.

- **TP (True Positive)** represents the number of FTU pixels correctly classified as FTU.
- **FP (False Positive)** denotes the number of background pixels misclassified as FTU (due to bias).
- **FN (False Negative)** is the number of FTU pixels misclassified as background.

- **TN (True Negative)** indicates the number of background pixels correctly classified as background.

The specificity is the TPs divided by the number of all positive outcomes:

$$\text{Specificity (PE)} = \frac{TP}{TP + FP} \tag{4}$$

Sensitivity is the number of TPs divided by the number of samples that should have been identified as positive:

$$\text{Sensitivity (SE)} = \frac{TP}{TP + FN} \tag{5}$$

Algorithm 3 Training strategy

Requirements: $|w|$ of model M, η: learning rate, e: epochs. unicast initialize - w_0 of basic training to hospital node.

for *per* $i \in \{R, \ldots R-1\}$ **do**

    to c-th client from round robin scheduling i for c

    */*run train on clients $c \in C$*/*

    augmented data $x'_{c,i}, y'_{c,i}$ by passing input data $x_{c,i}, y_{c,i}$, in Alg. 1.

    produces $x'_{c,i}, y'_{c,i}$ applying M_i model and continues e epoch training for all c at i generates loss $\sum_i^R L_i$ by passing $x'_{c,i}, y'_{c,i}$ through w_t in parallel.

    */*run update on server*/*

    updates $|w|$ via $w_t \leftarrow \eta.\nabla_{w_t} \left(\sum_i^R L_i \right) \triangleright meta$ *global model at the server-side update.*

    if *FedAvg* **then**

        send to c-th model M with average weight $\frac{1}{C} \sum_c^C |w|$

        for *each user c and each layer l* **do**

            if *layer l is not BatchNorm* **then**

                send to c-th model M with average weight $\frac{1}{C} \sum_c^C |w|$

            end

            if $C_i - P_i > 0$ **then**

                $PS_i \leftarrow (PS_i^{temp} - PS_i^{loss})$

            else

        end

    end

end

*/*run update on clients $c \in K$*/*

updates next round weight $w_{c,i+1}$ via $M_{c,i} \leftarrow M \triangleright local$ *model at the client-side update*

if N *is even* **then**

    $X \leftarrow X \times X$;

    $N \leftarrow \frac{N}{2}$

else

    if N *is odd* **then**

        $y \leftarrow y \times X$;

        $N \leftarrow N - 1$;

    end

end

if *frequency* **then**

    Perform tumor salient attention with the Alg. 2.

end

The Rand index, referred to as accuracy here, measures segmentation performance by evaluating all pixel pairs, calculating the ratio of correctly classified pairs to the total pairs:

$$\text{Rand index (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

The DC is calculated from the prediction accuracy and recall, scoring the overlap between the predicted segment and ground truth. It penalizes FPs, a common factor in high-end imbalanced datasets, such as medical image segmentation:

$$\text{Dice Coefficient (DC)} = \frac{2TP}{2TP + FP + FN} \quad (7)$$

Evaluating medical image segmentation involves challenges related to imbalanced datasets and different tumor sizes. In this context, an accurate assessment requires accounting for the TN results, especially with a small percentage of pixels in the ROI and the prevalence of background pixels. In image segmentation, each pixel should be treated as a data point. The AUC provides a consolidated measure of performance across all pixel classification thresholds. From this curve, representing two recall parameters, $TPR = \frac{TP}{TP+FN}$ and fallout $FPR = \frac{FP}{FP+TN}$, the AUC measures the two-dimensional area beneath the entire receiver operating characteristics (ROC) curve (integral operation) from $(0, 0)$ to $(1, 1)$. The dashed line presents the ROC curve in [51], depending on the baseline, to determine whether the model predicts the background retention and tumors correctly (typically, the threshold is 0.5)).

Due to the simplicity of the binary classification task, the model easily achieves accurate classification results under the optimization of the binary cross-entropy (CE) loss function [52], which performs instructions to overcome the disadvantage of excessive segmentation on tumor-free images (confusion). We added the loss functions combined with the standard stochastic gradient descent for associative learning to update the weight one time, and in each epoch, there are N data points, so the weight is updated N times. Source training and post-screening are necessary for the new source detection phase to ensure that the outputs are highly reliable to strengthen the boundaries of intermediate tumor salinity further. We propose expecting a multiscale structure analog index loss function (MS-SSIM) to assign a higher weight to blurred boundaries. The UNet monitors and captures blurred boundaries early because a greater regional distribution difference results in a higher MS-SSIM value. Two ROI regions of size N are cut from the results of segment P and the background truth mask G, respectively, which are $p = \{p_j : j = 1, \ldots N^2\}$ and $g = \{g_j : j = 1, \ldots N^2\}$. The MS-SSIM function of p and g is defined as follows:

$$\text{l}_{ms-ssim} = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_p\mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2} \right)^{\gamma_m} \quad (8)$$

where M denotes the sum of the scales, μ_p, μ_g, and σ_p, σ_g are the mean, and the standard deviation of p, g, and σ_{pg}, denoting their covariance. We determine the relative importance of the two components as β_m and γ_m in each scale, established according to [53]. Two minor constants, $C_1 = 0.01^2$ and $C_2 = 0.03^2$, are added to avoid the unstable situation of dividing by zero. This step is followed by the focal loss, which addresses class imbalances (e.g. misclassification and small p_t). The coefficient of variation is close to 1, and losses are not affected. When p_t approaches 1, the coefficient progresses to

0, and the losses for well-classified examples are weighted:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \tag{9}$$

$$l_{fl} = -(1 - p_t)^\gamma \log(p_t)$$

When $\gamma = 0$ FL corresponds to CE. As the γ increases, the effect of the modulation factor also increases. For example, for $\gamma = 2$, an example classified with, $p_t = 0.9$ has a loss of 100 less than CE, and with, $p_t = 0.968$, it has a loss less than 1000, which increases the importance of correcting misclassified examples.

The standard Jaccard loss function (IoU loss) works when the predicted limit boxes coincide with the basic truth box:

$$l_{IoU} = 1 - \frac{TP}{TP + FP + FN} \tag{10}$$

Thus, the combined loss is developed for segmentation according to a three-level hierarchical middle period: pixel, patch, and map levels, which can capture large- and small-scale structures with clear boundaries:

$$l_{seg} = l_{fl} + l_{ms-ssim} + l_{IoU} \tag{11}$$

V. IMPLEMENTATION DETAILS

This section selects and sets the settings for the system. The details are introduced in the following sections.

A. INDIVIDUAL BASELINE

For each task, the VGG backbone and UNet model variants were trained based on the FedAvg and FedBN scenarios. The global model training involved splitting the data into training and testing sets, using distinct portions of the DDSM, CBIS-DDSM, MIAS, INBREAST, and BUSI datasets to prevent overlap. These data are processed as described in Algorithm 1, Section IV-B, using the OpenCV library combined with Python programming. This involves converting DICOM images into 2D standard formats (PNG or JPG) for convenient use. The complete data for each client was partitioned into training and testing sets at 75% and 25%, respectively. Consequently, a single model was trained on a dataset encompassing two types of diagnostic imaging—radiology and ultrasound—addressing four distinct radiology types for breast cancer data.

B. CENTRAL AGGREGATION

In a multidistribution segmentation setup, we combine data types to create a hospital-based model capable of predicting relevant data. Acknowledging the intricate challenge, we recognize that aggregating various computational parameters through FedAvg may result in suboptimal performance compared to training individual models separately for each dataset or maintaining individual models to ensure FedBN stability. The centralized aggregation model is configured following the same steps as the baseline implementation, anticipating that the processing models are well-acquainted with all data and do not incur significant additional costs.

TABLE 2. Parameters for $VGG - s1$ and $VGG - s2$ with architecture models.

Architecture	$VGG - s1$	$VGG - s2$
UNet	39.396M	9.853M
UNet2+	47.176M	11.799M
UNet3+	26.972M	6.749M

C. NETWORK ARCHITECTURE

The versatility of this framework lies in its adaptability to any contemporary deep neural network by substituting the backbone with an alternative model architecture. In the experiments (Section VI-A), we explore two backbones, $VGG - s1$ and $VGG - s2$, for the fractional UNet model. The backbones differ in their model fractions, with $s2$ being smaller than $s1$ (Table 2). This modification adjusts the ratio of filter s1 to the depth of the encoder and decoder in the UNet model, transitioning from [64, 128, 256, 512] to filter $s2$ [32, 64, 128, 256]. For the postprocessing GMM classifier, each pixel feature is transformed into a 128-dimensional vector. In the implementation of this study, each class K is canonically represented by a GMM with $K = 2$.

D. TRAINING

The experiments involved horizontally partitioned image datasets distributed unevenly across hospitals (clients). All code was written in Python 3.10.1 using the PyTorch library (v. 2.1.0). The experimental computer configuration for simulation included 64 GB of RAM and one graphics processing unit (GTX 3090 with 24 GB), with the node architecture x86 and 64. The setup assumes all participating hospitals update the model in each global epoch (i.e., $f = 2$ during training). The UNet datasets and network architectures were selected based on performance considerations, ensuring proportionate participation in the study. The learning rate for all methods was initially set to 0.01 and sharply decreased with the number of local epochs ($e = 2$), according to CosineAnnealingLR, reaching 0.0001 before resuming an increase in the learning rate. The learning rate was chosen based on the combined performance, using the stochastic gradient descent optimization function for the models. Initial observations were made for the learning client, ensuring generalized learning with a minibatch size of $B = 10$, evaluating training and testing accuracy in Section VI for comparison:

$$\begin{cases} \eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + cos(\frac{T_{cur}}{T_{max}}\pi)\right), \\ \qquad T_{cur} \neq (2k + 1)T_{max} \\ \eta_{t+1} = \eta_t + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 - cos(\frac{1}{T_{max}}\pi)\right), \\ \qquad T_{cur} = (2k + 1)T_{max} \end{cases} \tag{12}$$

where T_{max} is the maximum number of repetitions and η is the learning rate. In each training iteration, a momentum (Sinkhorn) EM loop [53] (i.e., $t = 1$) is conducted on the current training batch and external memory to optimize

TABLE 3. Experimental setup.

Notice	Standard training	w. Local fine-tuning
Image size	128×128	
Learning rate	$\eta = 1.10\text{-}3$, decreased after 1 epochs until 1×10^{-5} at the hospital node	
Number of class	1	
Number of communication rounds	$R = 100$	
Number of hospital	$C = 5$	
Number of training batch	$B = 10$	
Number of training iterations	$e = 2$	
Scenarios	FedAvg or FedBN	
Test frequency	$f = 2$	

the GMM and backpropagation global gradient of CE loss over the current batch for discriminative training of feature extractors.

E. LOCAL FINE-TUNING
Similar to postprocessing, where a pretrained model is customized and at least one internal model parameter (i.e., weights) is fine-tuned, as demonstrated in [54] and [55], which are highlighted the importance of fine-tuning in FL models. Fine-tuning is crucial for enhancing the accuracy of the global model while maintaining accessibility for personalization. Consequently, the FedAvg and FedBN models were refined on local datasets, aiming to enhance the task-specific performance of each task block while preserving the block representation through the fine-tuning method.

VI. EXPERIMENTAL RESULTS
With advancements in imaging technology and artificial intelligence, the experiments demonstrated the incredible ability of the segmentation system to recognize breast tumors at an early stage and accurately assess their severity. Section VI also answers the research questions **RQs**, demonstrating the novelty and creativity of this research.

A. MODEL COMPARISON
In exploring advanced models, we conducted a series of tests, comparing the performance of various models, including UNet3+, UNet2+, and UNet+. We applied two training methods, FedAvg and FedBN, as presented in Tables 4. In this process, we used diverse data from five sources: DDSM, CBIS-DDSM, MIAS, INBREAST, and BUSI. We mixed a composite dataset (ALL) by combining data from these sources, which allowed us to evaluate the performance of the models on a diverse range of data collected from five hospitals rather than using just one type of data shared between hospitals.

The results of the experiment reveal that the UNet3+ model excelled with the highest accuracy. In addition, the performance of FedAvg and FedBN does not demonstrate a significant difference. When using the FedBN training method, the DC performance reached 81.4%, surpassing

the 80.3% reached by FedAvg. No significant deterioration occurred in performance between the two backbone models when using VGGs2, which may even have the advantage of a lower number of model parameters at about 70%. However, the UNet2+ model did not achieve satisfactory results, with 60% lower DC performance. These results refer to the advantages of the UNet3+ model in medical image segmentation and the performance difference between the FedAvg and FedBN methods and the proposed model (ALL).

B. COMPARISON WITH THE STATE-OF-THE-ART SCENARIO
This study examined a range of datasets, including DDSM, CBIS-DDSM, MIAS, INBREAST, and BUSI, to evaluate the performance of an ML model. Initially, this model performed well on the MIAS and BUSI datasets, with an accuracy of 86.9% and 90.3%, respectively, using the FedAvg method. However, when applying this model to other datasets, its performance is inferior, and the accuracy is low. To improve this, we performed local fine-tuning (+FT) beyond linking, retuning the model on each dataset.

Therefore, performance on the BUSI dataset was reduced by 2.5% despite improvements from 3.8% to 8.8% on other datasets. However, when applying another method, FedBN+FT increased from 4.3% to 11.0%, the BUSI data node had no change, and the method in the Meta+FedBN link increased from 4.3% to 6.0% compared to the baseline (except in BUSI, which decreased by 9.1%). The tough challenge with the INBREAST and BUSI datasets does not extend beyond the federated. All experimental results are listed in Table 5.

C. LOSS ASSESSMENT OF OPTIMAL SCENARIOS
During the comparison and research, we developed a hybrid loss using a three-level hierarchy: pixel, smashed, and map. This loss aims to capture large- and small-scale structures in the data with clear boundaries. We proposed Meta+FedBN to address the challenges associated with differences in data distribution across hospitals. This method aggregates the mean and variance statistics from activations across devices, differing from the Meta+FedAvg method. In practice, after applying Meta+FedBN to nodes at hospitals with diverse datasets, we observed a significant decrease in the total l_{seg} loss of Meta-FedBN, reaching 1026 at the $100th$ communication round. This outcome corresponds to a loss reduction of about 6.7% compared to the Meta+FedAvg method. We also found that the most substantial loss reduction occurs when applying l_{IoU}, especially when the predicted bounding boxes match the ground-truth boxes exactly. These findings indicate that Meta+FedBN performs better in this situation, as presented in Table 6.

D. DECISION THRESHOLDS
From the results in Section VI-A, FedBN outperforms FedAvg in DC performance. Consequently, we opted to compare the results using FedBN. Fig. 7 and Table 7

TABLE 4. Dice coefficient performance on the testing dataset using the FedAvg & FedBN scenarios and various architectural models.

Architecture	VGG-s1						VGG-s2					
	DDSM	CBIS-DDSM	MIAS	INBREAST	BUSI	ALL	DDSM	CBIS-DDSM	MIAS	INBREAST	BUSI	ALL
FedAvg-UNet	75.9	73.1	85.0	65.4	88.7	79.9	75.8	72.6	83.0	32.1	87.1	79.9
FedAvg-UNet2+	64.5	60.1	71.1	35.5	77.0	69.4	64.6	59.4	54.2	44.6	75.6	69.8
FedAvg-UNet3+	**76.7**	**75.2**	**86.9**	**75.3**	**90.3**	**81.0**	**76.7**	**75.2**	**88.0**	**75.9**	**89.9**	**80.3**
FedBN-UNet	75.5	73.5	84.1	63.7	88.7	80.0	75.5	72.5	81.9	32.1	86.7	80.0
FedBN-UNet2+	64.9	60.3	70.3	37.0	77.0	69.5	64.5	59.2	53.2	52.3	75.0	69.4
FedBN-UNet3+	**76.5**	**75.3**	**87.2**	**76.4**	**90.2**	**80.4**	**77.3**	**75.3**	**88.4**	**76.9**	**90.1**	**81.4**

TABLE 5. Performance DC on test dataset when using training strategy with a meta-global model.

Scenario	DC				
	DDSM	CBIS	MIAS	INBREAST	BUSI
Baseline in [13]	72.3	73.3	81.6	64.1	88.8
FedAvg in [14] + UNet3+	76.7 (+4.4%)	75.2 (+1.9%)	86.9 (+5.3%)	**75.7** (+11.6%)	**90.3** (+1.5%)
Meta-FedAvg	77.1 (+4.8%)	71.7 (-1.6%)	83.2 (+1.6%)	55.7 (-8.4%)	85.7 (-0.3%)
Meta-FedAvg + FT	**78.4** (+6.1%)	**79.8** (+6.5%)	**85.4** (+3.8%)	72.9 (+8.8%)	86.3 (-2.5%)
FedBN	76.5 (+4.2%)	75.3 (+2.0%)	87.2 (+5.6%)	**76.4** (+12.3%)	**90.2** (+1.4%)
Meta-FedBN	78.3 (+6.0%)	73.2 (-0.1%)	85.9 (+4.3%)	55.0 (-9.1%)	86.1 (+2.7%)
Meta-FedBN + FT	**79.4** (+7.1%)	**79.8** (+6.5%)	85.9 (+4.3%)	75.1 (+11.0%)	88.7 (-0.1%)

TABLE 6. Evaluate the overall semantic segmentation loss between the proposed FedAvg and FedBN scenarios.

Scenario	Meta+FedAvg				Meta+FedBN			
Server – ALL	l_{fl}	$l_{ms-ssim}$	l_{IoU}	l_{seg}	l_{fl}	$l_{ms-ssim}$	l_{IoU}	l_{seg}
Hospital 1 – DDSM	0.021	0.645	0.391	**1.058**	0.018	0.377	0.377	**1.043**
Hospital 2 – CBIS DDSM	0.035	0.588	0.454	**1.079**	0.026	0.580	0.443	**1.050**
Hospital 3 – MIAS	0.005	0.615	0.302	**0.923**	0.001	0.612	0.271	**0.885**
Hospital 4 – INBREAST	0.026	0.801	0.615	**1.444**	0.041	0.527	0.634	**1.204**
Hospital 5 – BUSI	0.036	0.658	0.270	**0.965**	0.029	0.653	0.266	**0.949**
Average	0.024	0.661	0.406	**1.093**	0.023	0.549	0.398	**1.026**

TABLE 7. Dice coefficient on the out-of-federation ALL dataset, considering the standard vs. postproduction GMM combination method.

Architecture	AUC	DC	ACC	SE	PE
FedBN UNet3+	98.2	80.4	94.3	90.8	86.0
FedBN UNet3+GMM	95.7	75.3	92.7	93.2	67.4
FedBN UNet3+FT	98.5	82.7	94.5	91.5	81.1
FedBN UNet3+FT+GMM	96.8	80.5	93.8	93.0	76.3

FIGURE 8. Gaussian mixture model analysis of the image histogram identifying the optimal threshold for separating the foreground from the background.
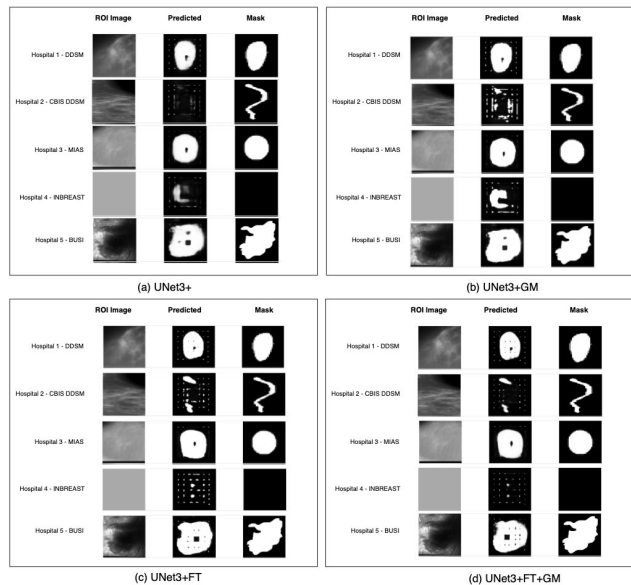

FIGURE 7. Ground truth segmentation masks with masks generated by hospitals using the UNet3+ models and meta federated batch normalization (Meta-FedBN).

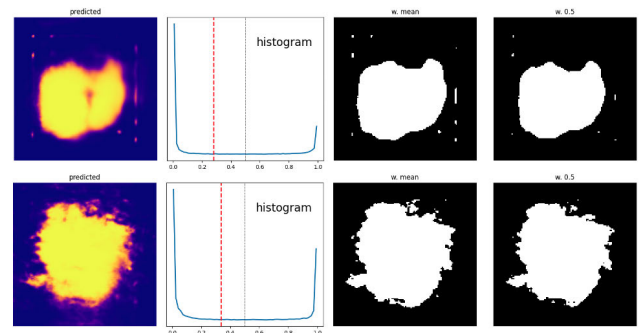illustrate that the fine-tuned FedBN model exhibits superior performance compared to the basic UNet3+ models when segmenting the DDSM, CBIS-DDSM, MIAS, INBREAST, and BUSI datasets. This outcome is manifested by the highest DC scores of 82.7% (without the GMM) and 80.5% (with the GMM), surpassing the 80.4% achieved by the base UNet3+ model. The UNet3+FT model, coupled with the GMM and leveraging the FedBN after fine-tuning, surpasses the baseline models by more effectively capturing tumor variability and prominence across datasets.

This finding is crucial when the training model relies on centrally aggregated data and may not generalize well to diverse labels. The combination with the GMM retains many characteristics due to its flexible threshold area, resulting in a longer and more general convergence time. In the absence of the GMM, using a fixed threshold of 0.5, the SE drops as low as 90.8%, lower than that for the UNet3+GMM
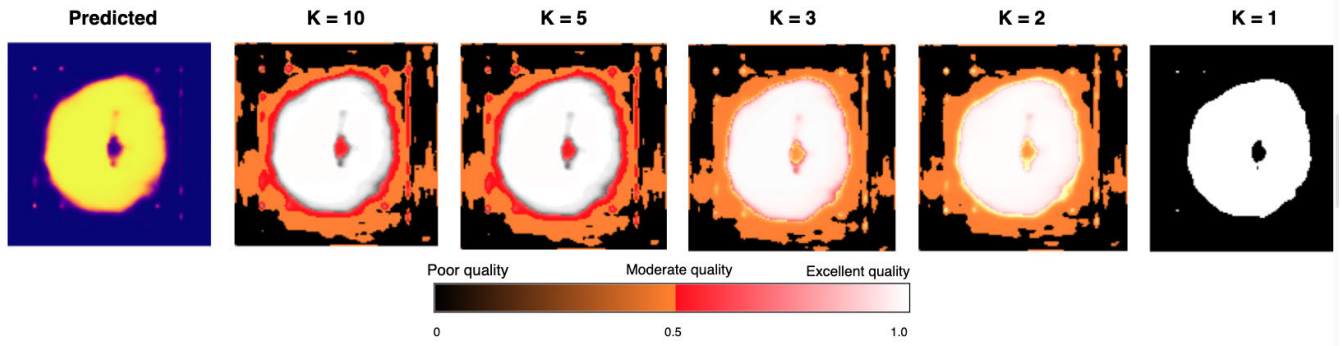
FIGURE 9. Eliminated saliency maps from the original dataset. The left column depicts the output test image, the middle column displays the saliency with each component after the Gaussian Mixture Model (GMM), and the right column presents the primary tumor mask.
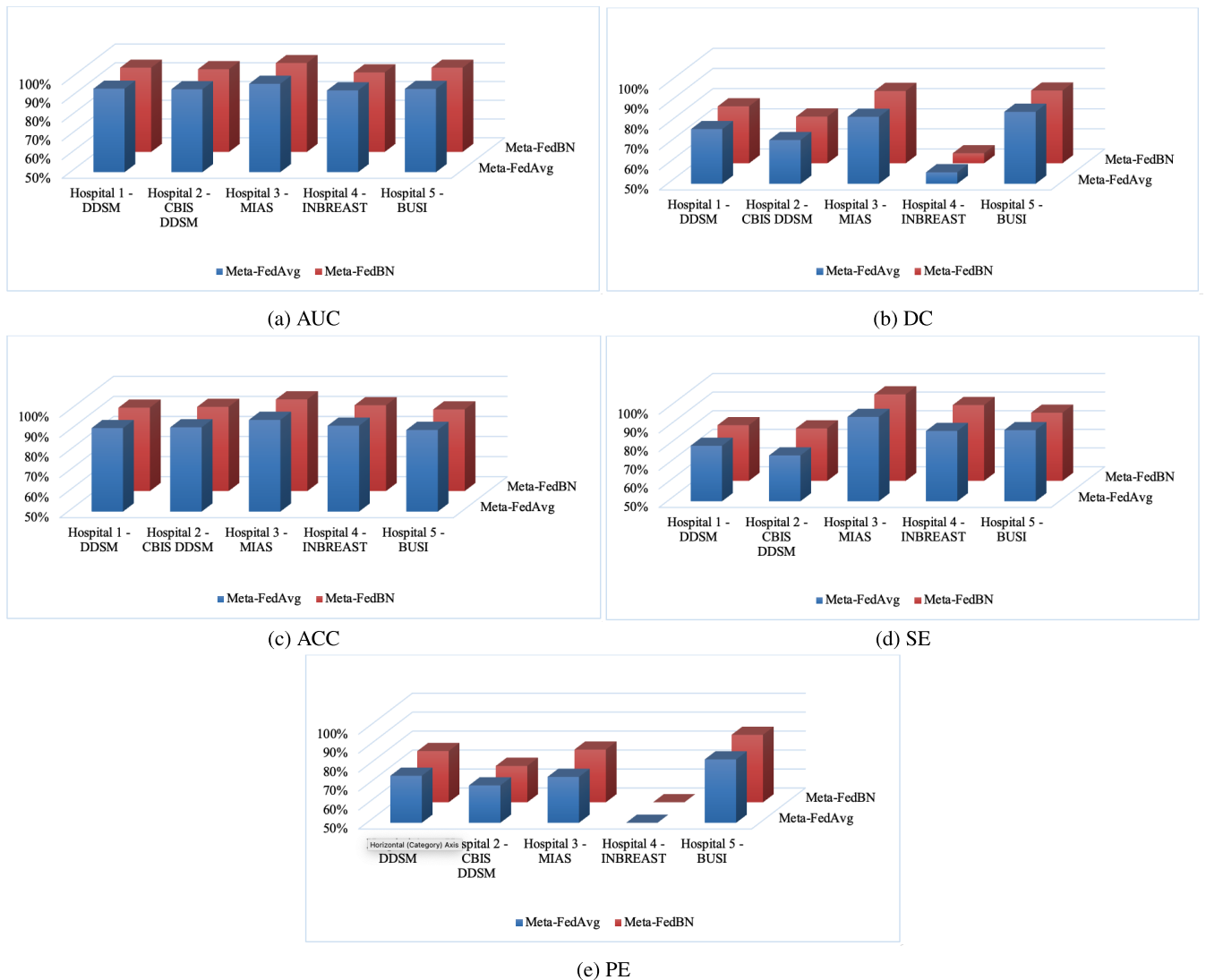

(a) AUC

(b) DC

(c) ACC

(d) SE

(e) PE

FIGURE 10. Performance evaluations based on five key metrics in meta federated averaging and batch normalization (Meta-FedAvg and Meta-FedBN) across hospitals.

and UNet3+FT+GMM at approximately 93%, displaying better sensitivity to detect FNs, especially in Hospital 2 (CBIS-DDSM) and Hospital 4 (INBREAST).

Furthermore, the assessment of other evaluation scores, such as AUC, accuracy, sensitivity, and specificity,

demonstrates the improved performance of the best-performing model. This result emphasizes the superiority of the proposed method for tumor segmentation across medical datasets, leveraging fine-tuning and the integration of methods for optimal results. In an extension of the

(a) Meta-FedAvg

(b) Meta-FedBN

FIGURE 11. Total time training comparison of the *VGG* − *s*1 and *VGG* − *s*2 models in meta federated averaging (Meta-FedAvg) and meta federated batch normalization (Meta-FedBN) with segmentation loss across communication rounds.

proposed approach, we used information from the mixture model instead of employing a fixed decision threshold of 0.5 to classify the data. The GMM determines the decision threshold ($K = 2$) based on the mean value of the components of the Gaussian distribution.

To illustrate this, we generated a histogram representing the data distribution. The tumor retains essential information regarding the primary tumor and its surrounding parts, facilitating the estimation of the tumor spread and determining its prominence in the prediction process. By employing an average threshold based on the number of tumor components, the data distribution was concentrated between 0.3 and less than 0.5, serving as an excellent example of applying distribution-based decision thresholds, enabling more flexible and accurate tumor decisions.

In the final phase as shown in Fig. 9, we assessed the saliency heatmap in various forms within the prediction output components. With increasing detail, we could better discern the shape and distribution of the tumor across multiple layers, ranging from low to high quality. Throughout this analysis, we employed a specific color code:

- Orange represents weakly involved areas.
- Red indicates the extent of the tumor boundary relative to the background.
- White symbolizes the tumor's most vigorous intensity.

Using $K = 5$ in this context, we found that the histogram was well-suited for calculating the tumor shape and determining its spread. In contrast, employing $K = 1$ allowed us to focus on determining the primary tumor. This adaptive approach enabled personalized decisions based on the data distribution at each hospital. To ensure personalized control over the output of the UNet3+ model, we implemented supply-port combination losses. This strategic incorporation enhanced decision-making capabilities, ensuring consistency in the model output across scenarios.
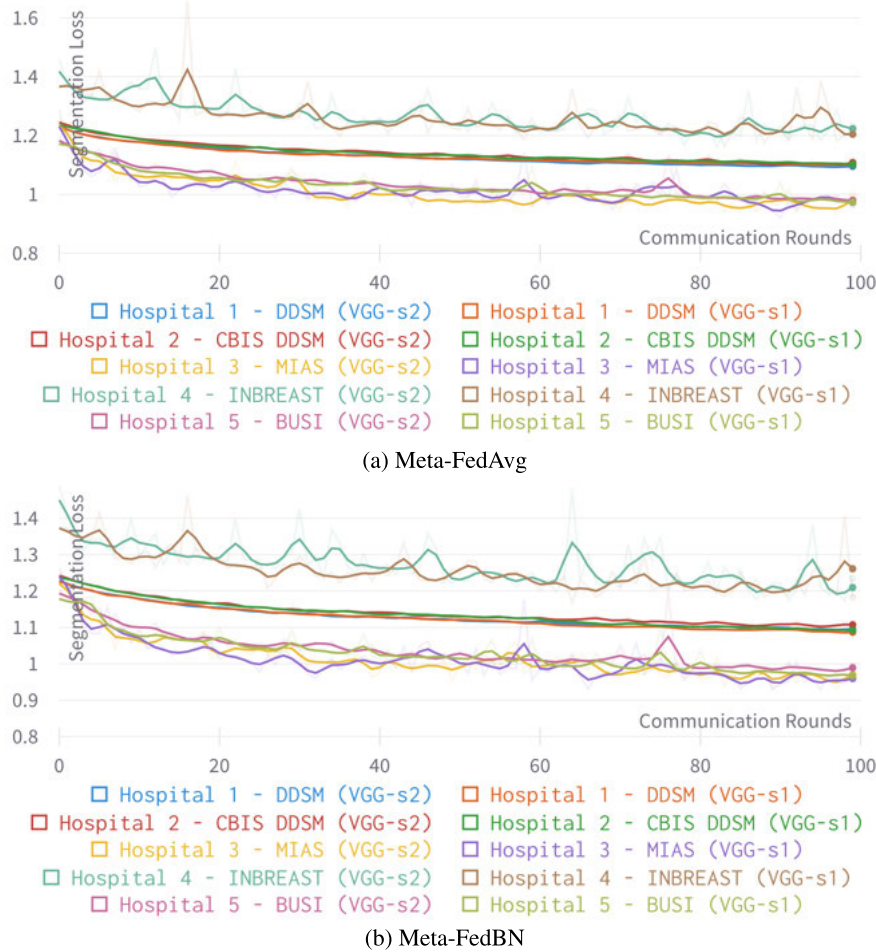
(a) Meta-FedAvg

(b) Meta-FedBN

FIGURE 12. Convergence comparison of the $VGG - s1$ and $VGG - s2$ models in meta federated averaging (Meta-FedAvg) and meta federated batch normalization (Meta-FedBN) with segmentation loss across communication rounds.

E. PERFORMANCE EVALUATIONS BASED ON AUC, DC, ACC, SE, AND PE

This section delves into additional performance evaluations based on five key metrics (area under the curve, Dice coefficient, accuracy, sensitivity, and specificity) across five diverse data types. The analysis compares three scenarios: traditional, FedAvg, and FedBN, as illustrated in Table 8, providing crucial insight for a comprehensive understanding of the study outcomes. The findings highlight significant performance variations between data types and scenarios. Notably, the BUSI dataset and ultrasound image data exhibit impressive predictions, surpassing 90%, whereas x-ray image data face challenges, performing at levels of around 75% to 80%. These disparities underscore the complexity of leveraging varied data types, emphasizing the need for thoughtful selection aligned with the research objectives. In addition, the study emphasizes the substantial influence of the INBREAST dataset on the overall tumor prediction and segmentation performance, given its unique characteristics. Recognizing such variations arising from the distinct dataset characteristics is crucial when training models.

The evaluation extends to two cases, Meta-FedAvg and Meta-FedBN, assuming each dataset represents a hospital on each $VGG - s1$ and $VGG - s2$ models. This approach mirrors real-world scenarios where hospitals possess diverse datasets. The analysis in Fig. 10 provides insight into data differences, facilitating a realistic assessment of the method's applicability. Privacy concerns are acknowledged, emphasizing that the results serve as a valuable reference for researchers and hospitals seeking solutions to connect data and enhance predictive capabilities.

Upon reviewing experimental results and considering FedAvg and FedBN scenarios with two $VGG - s1$ and $VGG - s2$ models in Fig. 11, no significant time differences between hospitals were observed. Variations primarily hinge on training data from the DDSM and CBIS-DDSM datasets, peaking at 500 seconds per communication round. A time difference of approximately 200 seconds in architectural models indicates the influence of the training conditions on the performance and training time.

This disparity underscores the need for future research to optimize training processes based on each hospital's data

TABLE 8. Scenarios with multiple performance metrics on the Database for Screening Mammography dataset.

Scenarios	AUC	DC	ACC	SE	PE
DDSM					
Traditional	96.6	72.3	91.4	81.0	83.1
FedAVG	96.8	76.7	91.4	81.7	83.6
FegBN	96.8	76.5	91.5	80.5	83.1
CBIS-DDSM					
Traditional	97.1	73.3	93.1	78.8	77.0
FedAVG	97.2	75.2	93.1	78.9	78.4
FegBN	97.3	75.3	93.1	78.0	78.1
MIAS					
Traditional	99.0	81.6	99.0	99.1	93.1
FedAVG	99.6	86.9	96.6	99.3	93.1
FegBN	99.6	87.2	96.7	99.3	96.3
INBREAST					
Traditional	98.2	64.1	97.6	98.7	97.0
FedAVG	98.4	75.7	97.7	98.0	97.1
FegBN	98.5	76.4	97.7	98.9	94.1
BUSI					
Traditional	98.1	88.8	93.2	96.1	99.6
FedAVG	98.5	90.3	93.3	97.0	1.00
FegBN	98.6	90.2	93.2	96.6	1.00
ALL					
Traditional	-	-	-	-	-
FedAVG	98.2	81.0	94.4	90.4	84.0
FegBN	98.2	80.4	94.3	90.8	86.0

distribution structure and learning opportunities. The process graph for two $VGG - s1$ and $VGG - s2$ models depicted in Fig. 12 presents the semantic segmentation loss, with minimal variations observed between models and scenarios. The loss values consistently range between 0.9 and 1.3, indicating the stability of the training process. This stability suggests the model's robustness during training, its ability to learn under diverse conditions, and its resilience against overfitting or performance degradation. Overall, these findings provide a positive outlook for the reliability of the proposed model in real-world applications.

VII. CONCLUSION

This study introduces an innovative FL architecture for breast tumor segmentation in medical images, combining physician visual observation with X-ray and ultrasound images. Using the modified UNet model and incorporating special attention blocks through GMM will improve the segmentation quality. Applying GMM supports the segmentation of salient images, focusing on important regions during the segmentation process. Experiments in the FedAvg and FedBN scenarios, as well as diverse evaluation methods on five standard datasets with hyper-global and complete models, showed superior segmentation performance compared to the baseline method. This approach addresses challenges in DL for medical image processing, especially in ultrasound and mammography segmentation, offering a promising strategy for physicians and technologists. It provides a tool to support and anticipate early tumor detection. The next study will

focus on developing a software application, with software projects collected directly at different cancer hospitals, in accordance with the architecture of FL, with the aim of improving the quality of life for the community.

REFERENCES

[1] X. Chen, S. Sun, N. Bai, K. Han, Q. Liu, S. Yao, H. Tang, C. Zhang, Z. Lu, Q. Huang, G. Zhao, Y. Xu, T. Chen, X. Xie, and Y. Liu, "A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy," *Radiotherapy Oncol.*, vol. 160, pp. 175–184, Jul. 2021.

[2] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, and Van B. Ginneken, "The medical segmentation decathlon," *Nature Commun.*, vol. 13, no. 1, p. 4128, 2022.

[3] A. Prochazka, O. Vysata, and V. Marik, "Integrating the role of computational intelligence and digital signal processing in education: Emerging technologies and mathematical tools," *IEEE Signal Process. Mag.*, vol. 38, no. 3, pp. 154–162, May 2021, doi: 10.1109/MSP.2021.3058634.

[4] F. Z. Francies, R. Hull, R. Khanyile, and Z. Dlamini, "Breast cancer in low-middle income countries: Abnormality in splicing and lack of targeted treatment options," *Amer. J. Cancer Res.*, vol. 10, no. 5, pp. 1568–1591, 2020.

[5] L. P. Huong, R. J. Stafford, B. Isabelle, B. Patrick, A. Wendy, and L. Stacy, "Chapter 28—Breast cancer," in *Oncologic Imaging: A Multidisciplinary Approach*, M. P. Silverman, Ed., 2nd ed., Philadelphia, PA, USA: W. B. Saunders, 2023, pp. 476–502.

[6] A. Bhowmik and S. Eskreis-Winkler, "Deep learning in breast imaging," *BJR Open*, vol. 4, no. 1, Jan. 2022, Art. no. 20210060.

[7] L. Yang, S. Wang, L. Zhang, C. Sheng, F. Song, P. Wang, and Y. Huang, "Performance of ultrasonography screening for breast cancer: A systematic review and meta-analysis," *BMC Cancer*, vol. 20, no. 1, p. 499, Dec. 2020, doi: 10.1186/s12885-020-06992-1.

[8] A. A. Mohamed, Y. Luo, H. Peng, R. C. Jankowitz, and S. Wu, "Understanding clinical mammographic breast density assessment: A deep learning perspective," *J. Digit. Imag.*, vol. 31, no. 4, pp. 387–392, Aug. 2018.

[9] S. Balaji, T. Arunprasath, M. P. Rajasekaran, G. Vishnuvarthanan, and K. Sindhuja, "Computer-aided diagnostic system for breast cancer detection based on optimized segmentation scheme and supervised algorithm," *Automatika*, vol. 64, no. 4, pp. 1244–1254, Oct. 2023, doi: 10.1080/00051144.2023.2244307.

[10] M. F. Mridha, M. A. Hamid, M. M. Monowar, A. J. Keya, A. Q. Ohi, M. R. Islam, and J.-M. Kim, "A comprehensive survey on deep-learning-based breast cancer diagnosis," *Cancers*, vol. 13, no. 23, p. 6116, Dec. 2021.

[11] Y. N. Tan, V. P. Tinh, P. D. Lam, N. H. Nam, and T. A. Khoa, "A transfer learning approach to breast cancer classification in a federated learning framework," *IEEE Access*, vol. 11, pp. 27462–27476, 2023, doi: 10.1109/ACCESS.2023.3257562.

[12] O. Ginsburg, C. H. Yip, A. Brooks, A. Cabanes, M. Caleffi, J. A. D. Yataco, B. Gyawali, V. McCormack, M. M. de Anderson, R. Mehrotra, and A. Mohar, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. 10, pp. 2379–2393, May 2020.

[13] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2016, pp. 1273–1282.

[14] T. Alam, W.-C. Shia, F.-R. Hsu, and T. Hassan, "Improving breast cancer detection and diagnosis through semantic segmentation using the UNet3+ deep learning framework," *Biomedicines*, vol. 11, no. 6, p. 1536, May 2023, doi: 10.3390/biomedicines11061536.

[15] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-24574-4_28.

[18] M. Xian, Y. Zhang, H. D. Cheng, F. Xu, B. Zhang, and J. Ding, "Automatic breast ultrasound image segmentation: A survey," *Pattern Recognit.*, vol. 79, pp. 340–355, Jul. 2018.

[19] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang, "Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 240–249, Jan. 2019.

[20] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.

[21] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, "Connected-UNets: A deep learning architecture for breast mass segmentation," *Npj Breast Cancer*, vol. 7, no. 1, p. 151, Dec. 2021, doi: 10.1038/s41523-021-00358-x.

[22] C. Cao, X. Liu, Y. Yang, Y. Yu, J. Wang, Z. Wang, Y. Huang, L. Wang, C. Huang, W. Xu, D. Ramanan, and T. S. Huang, "Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 2956–2964.

[23] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6450–6458.

[24] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *J. Biomed. Informat.*, vol. 99, Nov. 2019, Art. no. 103291.

[25] N. Dong and I. Voiculescu, "Federated contrastive learning for decentralized unlabeled medical images," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., Cham, Switzerland: Springer, 2021, pp. 378–387.

[26] A. Jiménez-Sánchez, M. Tardy, M. A. González Ballester, D. Mateus, and G. Piella, "Memory-aware curriculum federated learning for breast cancer classification," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107318.

[27] M. Y. Lu, R. J. Chen, D. Kong, J. Lipkova, R. Singh, D. F. K. Williamson, T. Y. Chen, and F. Mahmood, "Federated learning for computational pathology on gigapixel whole slide images," *Med. Image Anal.*, vol. 76, Feb. 2022, Art. no. 102298, doi: 10.1016/j.media.2021.102298.

[28] J. Wicaksana, Z. Yan, D. Zhang, X. Huang, H. Wu, X. Yang, and K.-T. Cheng, "FedMix: Mixed supervised federated learning for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 7, pp. 1955–1968, Jul. 2022, doi: 10.1109/TMI.2022.3233405.

[29] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proc. 5th Int. Workshop Digital Mammography*, M. J. Yaffe, Ed., 2001, pp. 212–218.

[30] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin, 2016, "Curated breast imaging subset of digital database for screening mammography (CBIS-DDSM)," The Cancer Imaging Archive, doi: 10.7937/K9/TCIA.2016.7O02S9CY.

[31] J. Suckling, "The mammographic image analysis society digital mammogram database exerpta medica," in *Proc. Int. Congr. Ser.*, vol. 1069, 1994, pp. 375–378.

[32] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: Toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, pp. 236–248, Feb. 2012.

[33] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863, doi: 10.1016/j.dib.2019.104863.

[34] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[35] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.

[36] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, Apr./May 2018.

[37] H. Guo, Y. Mao, and R. Zhang, "Mixup as locally linear out-of-manifold regularization," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Honolulu, HI, USA, 2019, pp. 3714–3722.

[38] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.

[39] V. Patel and K. Mistree, "A review on different image interpolation techniques for image enhancement," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 12, pp. 129–133, Dec. 2013.

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2014.

[41] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[42] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with Atrous separable convolution for semantic image segmentation," in *Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII*. Berlin, Germany: Springer-Verlag, 2018, pp. 833–851, doi: 10.1007/978-3-030-01234-2_49.

[44] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," in *Proc. Med. Imag. Deep Learn.*, 2018.

[45] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[46] Y. Yang, Y. Zhuang, and Y. Pan, "Multiple knowledge representation for big data artificial intelligence: Framework, applications, and case studies," *Frontiers Inf. Technol. Electron. Eng.*, vol. 22, no. 12, pp. 1551–1558, Dec. 2021.

[47] W. Wang, C. Han, T. Zhou, and D. Liu, "Visual recognition with deep nearest centroids," 2022, *arXiv:2209.07383*.

[48] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," in *Proc. ICLR*, Feb. 2021, pp. 1–27.

[49] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," 2021, *arXiv:2102.07623*.

[50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.

[51] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Pacific Grove, CA, USA, 2003, pp. 1398–1402, doi: 10.1109/ACSSC.2003.1292216.

[52] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[53] G. Mena, A. Nejatbakhsh, E. Varol, and J. Niles-Weed, "Sinkhorn EM: An expectation-maximization algorithm based on entropic optimal transport," 2020, *arXiv:2006.16548*.

[54] Y. Jiang, J. Konečný, K. Rush, and S. Kannan, "Improving federated learning personalization via model agnostic meta learning," 2019, *arXiv:1909.12488*.

[55] P. Pu Liang, T. Liu, L. Ziyin, N. B. Allen, R. P. Auerbach, D. Brent, R. Salakhutdinov, and L.-P. Morency, "Think locally, act globally: Federated learning with local and global representations," 2020, *arXiv:2001.01523*.

NGUYEN TAN Y is currently pursuing the Ph.D. degree with the Faculty of Electrical and Electronic Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. He is a Lecturer at Phan Thiet University, Phan Thiet City, Binh Thuan. His primary research focuses on applications using machine learning, image processing, and computer vision in the medical field.

DUY-DONG LE received the master's degree in computer science from the University of Information Technology, Vietnam National University Ho Chi Minh City (UIT-VNUH). He is currently pursuing the Ph.D. degree in computer science with Industry University Ho Chi Minh City (IUH). He works as a Lecturer with the Undergraduate Training Department, Vinh Long Campus, University of Economics Ho Chi Minh City (UEH). His areas of research expertise revolve around deep learning and the Internet of Things.

PHAM DUC LAM is currently a Lecturer with Nguyen Tat Thanh University, Vietnam. His research interests include health applications that apply technologies, such as image processing, embedded systems, and the Internet of Things.

NGUYEN HOANG NAM received the Ph.D. degree from National Chiao Tung University, Taiwan, in 2017. He is currently a Researcher with the MERLIN Research Group and a Lecturer with the Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. His research interests include health applications using artificial intelligence, image processing, and computer vision.

VO PHUC TINH received the bachelor's degree from the Faculty of Electrical and Electronic Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. He is currently pursuing the Ph.D. degree with DIISM, University of Siena, Siena City, Italy. His primary research is in edge computing with artificial intelligence, the Internet of Things, and embedded systems.

TRAN ANH KHOA received the Ph.D. degree from the University of Siena, Siena, Italy, in 2017. He is currently a Researcher with the MERLIN Research Group, Faculty of Electrical and Electronic Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. His research interests include health applications using artificial intelligence and the Internet of Things.

● ● ●