IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 February 2024, accepted 27 February 2024, date of publication 7 March 2024, date of current version 21 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374650

RESEARCH ARTICLE

Federated Learning Approach for Breast Cancer Detection Based on DCNN

HUSSAIN ALSALMAN[1], MABROOK S. AL-RAKHAMI[2], (Member, IEEE), TAHA ALFAKIH[2], AND MOHAMMAD MEHEDI HASSAN[2], (Senior Member, IEEE)

[1]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Mabrook S. Al-Rakhami (malrakhami@ksu.edu.sa)

This work was supported by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Project IFKSUDR-D127.

ABSTRACT Breast cancer stands as one of the predominant health challenges globally, affecting millions of women every year and necessitating early and accurate detection to optimize patient outcomes. Currently, while deep convolutional neural networks (DCNNs) have shown promise in breast cancer detection, their application is often hampered by privacy concerns associated with sharing patient data and the limitation of training on small, localized datasets. Addressing these challenges, this manuscript introduces an effective federated learning approach tailored for breast cancer detection, leveraging DCNNs across diverse and large datasets without compromising data privacy. Our experimental findings underscore significant advancements in detection accuracy of 98.9% on three large scale datasets which are VINDR-MAMMO, CMMD, and INBREAST. Additionally, we tested the proposed federated learning performance, showcasing the potential of our approach as a robust and privacy-preserving solution for future breast cancer diagnostic strategies.

INDEX TERMS Breast cancer detection, federated learning, deep convolutional neural networks, DCNN, medical image analysis, healthcare data privacy.

I. INTRODUCTION

Breast cancer represents one of the main causes of mortality in the world [1]. According to the latest statistics, it is estimated that 1 in 8 women will be diagnosed with breast cancer throughout their lives,[1] making it a highly relevant public health problem. Early detection of this disease is crucial, as it significantly increases survival rates and allows the implementation of less invasive treatments [2].

In recent years, artificial intelligence (AI) has revolutionized the field of medicine, especially in the field of diagnosis [3]. Deep learning algorithms, specifically the deep convolutional neural networks (DCNN), have proven to be extremely efficient tools in the detection and diagnosis of diseases, including breast cancer [4], [5], [6]. By leveraging these technologies, it is possible to improve the accuracy of diagnosis and accelerate the healing process [7]. However, when using AI algorithms for medical diagnosis, concerns arise regarding the privacy and security of patient data. Therefore, it is essential to address these issues when developing AI-based solutions [8], [9].

As advances continue in the domain of breast cancer diagnosis through machine learning and other computational techniques, it's crucial to address and refine certain limitations present in existing research [10]. A significant challenge faced by many studies is their dependence on limited datasets [10]. The robustness and reliability of machine learning models are profoundly influenced by dataset size. While smaller datasets might be more manageable, they often fail to capture the diverse variations present in real-world medical scenarios [11]. Such limited datasets can lead to biases, resulting in overfitting, where the model excels with the training data but underperforms when presented with new, unseen cases.

Increasing the dataset size in deep learning models, particularly for breast cancer early detection, directly contributes

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

[1]https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html

to accuracy improvement in several keyways [12], including the following benefits:

- Reduction of Overfitting: DCNNs are known for their capacity to learn complex patterns. However, with smaller datasets, there's a significant risk that the model will learn noise and specificities of the training data instead of generalizing patterns [13], [14]. Increasing the dataset size provides the model with a more varied set of examples, reducing the likelihood that it will memorize the training data and improving its ability to generalize to new, unseen data.
- Enhanced Feature Learning: DCNNs learn features automatically from the data. With a larger dataset, the network has access to a more diverse set of features, including subtle variations in breast cancer presentations [15]. This diversity is crucial for early detection where differences in imaging might be minimal and not well-represented in smaller datasets.
- Improved Robustness: A larger dataset is likely to include a wider range of cases, including rare and atypical presentations of breast cancer [16]. This diversity can make the model more robust, as it learns to identify cancer across a broader spectrum of cases, thereby improving its diagnostic accuracy.
- Balanced Class Distribution: In breast cancer detection, datasets can suffer from class imbalance (e.g., more benign cases than malignant). Larger datasets can help balance this distribution, allowing the model to learn equally from both classes and improving its diagnostic precision and recall [17].
- Cross-Patient and Cross-Institutional Generalization: Breast cancer presentations can vary significantly between patients and institutions (due to different imaging equipment, techniques, etc.) [18]. Larger datasets, particularly in a federated learning context [19], can encompass data from multiple sources, enhancing the model's ability to generalize across different patient demographics and imaging protocols [20].
- Statistical Significance: Larger datasets provide more data points, which can lead to more statistically significant results [21]. This is crucial in medical applications where the reliability and reproducibility of results are paramount.
- Enabling Complex Model Architectures: Larger datasets support the use of more complex DCNN architectures without the risk of overfitting [22]. These complex models can potentially learn more nuanced features relevant for early breast cancer detection, which might not be possible with smaller datasets.

In the realm of medical diagnostics, the consequences of such discrepancies can be grave, given that a misdiagnosis can drastically alter a patient's treatment path.

Moreover, while several studies highlight their methodological advancements [23], [24], [25], the accuracy rates of some models can still be improved. Given the critical nature of breast cancer diagnosis, achieving peak accuracy is of utmost importance. Even a slight increment in accuracy can have a tangible impact on patient outcomes, ensuring timely and accurate treatment.

Additionally, with the increasing digitization of medical records and data, there's a pressing need to prioritize the privacy and security of patient information [26], [27]. Especially in sensitive areas such as cancer diagnosis, preserving the confidentiality of patient data is not just a technical necessity but also an ethical obligation [28]. As we advocate for larger and more diverse datasets, it becomes equally important to implement stringent data protection measures to ensure that patients' privacy rights are not compromised.

Recognizing these challenges, particularly in accuracy and data privacy and security, underscores the necessity of our proposed federated learning approach, which aims to leverage collaborative, decentralized data sources to enhance the robustness and accuracy of breast cancer detection models.

The present work aims to contribute to this area, presenting an efficient federated learning approach for breast cancer detection. Through this methodology, it is proposed to develop a highly effective DCNN for the early detection of breast cancer, while guaranteeing patient privacy using federated learning technology. This approach avoids the need to centralize data, allowing models to be trained on local devices, which in turn ensures greater privacy.

Furthermore, another significant contribution of this work is the utilization of a large-scale breast cancer dataset, which enhances the overall accuracy of the proposed approach. With the combination of these technologies, the aim is to offer a robust and secure tool that contributes to the fight against breast cancer, maximizing the accuracy of the diagnosis and protecting sensitive patient data.

The rest of this paper is organized as follow: Section II gives a brief review for related works. Section III elaborates the proposed federated learning approach in general. Section IV explains the DCNN model for breast cancer detection. Section V shows the experimental results of our proposed DCNN model while Section VI describes the experiments of federated learning Approach. Section VII provides more experiments on data privacy and security. Section VIII conducted experiments on computational efficiency. Section IX gives a brief discussion, and finally, we concluded our work with Section X.

II. RELATED WORKS

In this section, we reviewed some of the related works that focus on the development of breast cancer frameworks using federated learning and AI algorithms.

A study on Triple-negative breast cancer (TNBC) [29], known for its high metastatic potential and limited treatment options, utilized machine learning to determine early TNBC patient response to neoadjuvant chemotherapy (NACT) using whole-slide images and clinical data. Using federated learning, the authors addressed the challenges of small datasets and data privacy. The findings demonstrated that while local

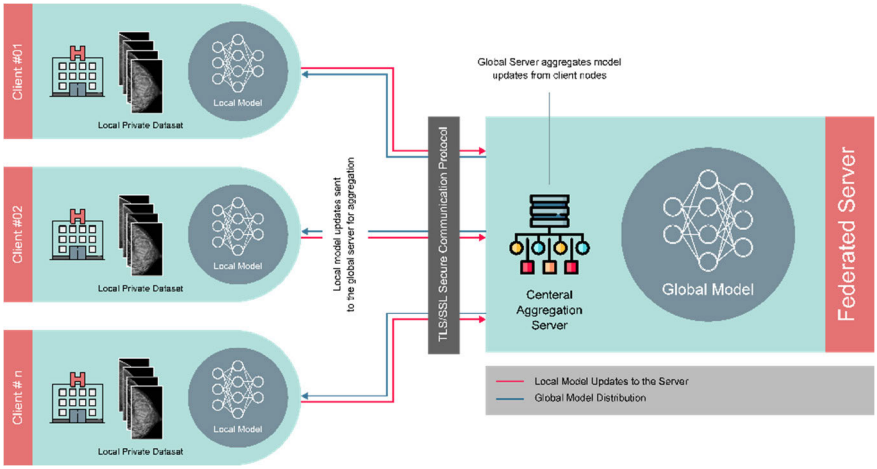

FIGURE 1. Simplified diagram of the federated learning approach.

ML models were effective, collaborative federated training enhanced performance, matching expert annotation levels. This approach emphasized the utility of federated learning in exploring large datasets for biomarker identification.

In another study [30], authors presented a new federated learning technique that focuses on memory-aware curriculum learning, enhancing local model consistency by penalizing forgotten sample predictions. The training process prioritizes overlooked samples post-global model deployment, integrates unsupervised domain adaptation for domain shifts, and maintains data privacy. Using ROC-AUC and PR-AUC metrics on three clinical datasets, the method outperformed conventional federated settings by 5% and 6%, respectively. This research underlined the benefits of curriculum learning in federated environments for breast cancer classification.

Addressing challenges related to data availability and privacy in medical imaging, authors in [31] introduced a federated learning framework for breast cancer histopathology. By sharing model parameters rather than data, the approach safeguards data privacy and consolidates knowledge from multiple sources. Tested on the BreakHis dataset, the federated method yielded results comparable to centralized learning, emphasizing the method's feasibility and efficiency.

Highlighting AI's increasing role in breast cancer diagnosis, the study [32] described a federated learning system that extracts features without compromising privacy [33]. Distinctively, it utilizes transfer learning to process image regions of interest, applies SMOTE for balanced data classification, and leverages FeAvg-CNN + MobileNet in the FL framework. Comparing various deep learning and transfer learning models, the approach showcased superior classification performance of 100% recall and 99.8% AUC when using pre-trained MobileNet model to extract features, highlighting its potential in AI healthcare solutions.

Survival analysis, crucial in medicine, faces challenges due to incomplete and confidential data. Authors in [34] proposed FedSurF++, an extension of the Federated Survival Forest algorithm, that creates random survival forests across diverse federations. The method displayed comparable performance to neural network-based models, requiring only a single communication round, emphasizing its efficiency and privacy preservation. Tested on real-world datasets, the results indicate the algorithm's potential in large-scale, privacy-conscious survival analysis with maximum cumulative AUC of 95.6%.

Finally, a recent study [35] proposed a comprehensive disease diagnosis system using federated and deep learning. It encompasses image acquisition, encryption for confidentiality using the E-EIE method, optimized key generation, secure data storage using federated learning, and classification using the C2T2Net model. Parameter fine-tuning was done using the CTSO algorithm. Results from the BreakHis Database indicated high accuracy (95.68%) and performance across various metrics.

Despite the distinguished efforts in previous research, some of these studies have inherent limitations. For instance, they often rely on relatively small datasets [32], which can constrain the generalizability of their findings. Additionally, despite notable advancements, certain methods [34], [35] reported accuracy rates that may not be optimal for critical applications such as breast cancer diagnosis.

III. FEDERATED LEARNING APPROACH

As Figure 1 illustrates, the general architecture of the suggested federated learning technique for breast cancer detection with the DCNNs model.

It depicts the federated system with the interaction of multiple clients (Hospitals/ Medical centers), sharing their local models' outputs with the central server that acts as a data collector [36]. This server eventually redistributes the enhanced model of cumulative joint knowledge back among the individual nodes.

Each or client utilizes the DCNNs model to process the data related specifically to breast cancer, subjecting it to

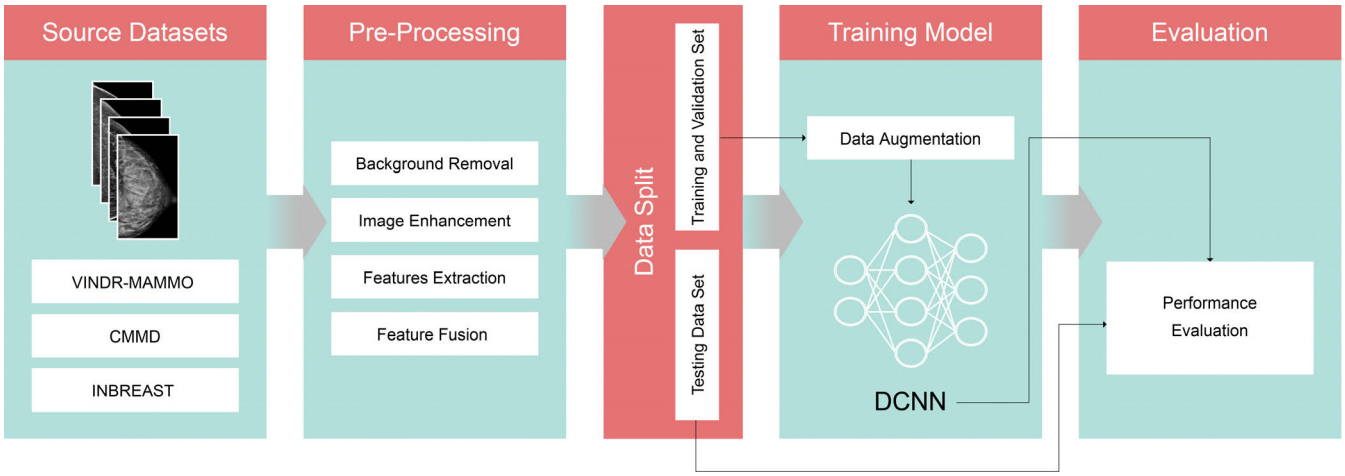

FIGURE 2. Proposed model for the breast cancer detection based on DCNNs.

completely independent training. Based on the original data from several comprehensive datasets, which will be explained in detail in the next sections, the data is segmented on the level of each individual hospital in an attempt to mimic diverse scenarios where one client possesses extensive data while others may have less data, resulting in an uneven distribution.

Three distinctive datasets will be used for the evaluation of all trained models, as explained in IV-B. This method allows a sensible direct comparison between the evaluated models since each will be consistently assessed through these datasets possessing more or less data for the training. It's noteworthy to clarify that all the deployed federated models will be trained using the Federated Average method (also known as FedAvg) [37]. To make it all simpler, each hospital will independently train its own model on its individual dataset. After a predefined number of turns, these models' parameters will be shared with the central server, which will subsequently blend them together. This iteration will proceed until the completion of the entire training.

The Federated Average method represents a usual implementation of the algorithms for federated optimization [38], including a fraction of customers, C, set to 1 and a fixed learning rate η, making every client (k) calculate $g_k = \Delta F_k$ (w_t), representing the mean gradient derived from the local data at the current model wt. The central server accumulates these gradients and uses the update based on the following equation:

Following that, each hospital performs local gradient descent on the present model employing its individual data, and the global server aggregates a weighted average of all the acquired models.

IV. OVERALL DCNN MODEL

Figure 2 represents the proposed model for the breast cancer detection based on DCNNs which is a comprehensive, step-by-step methodology encompassing the entire process from data sourcing to model evaluation. Initially, data is sourced from three reputable datasets (See Sec. V-A), which are: VINDR-MAMMO, CMMD, and INBREAST. These datasets offer a plethora of mammographic images crucial for the research.

Subsequent to the data collection is the pre-processing phase. This phase is instrumental in refining and preparing the raw images for optimal input into the neural network. Key steps involved in this phase (See: Sec. IV-A) include the removal of any background noise or unnecessary elements from the images, enhancing image clarity and sharpness, extracting critical features that could be pivotal for cancer detection, and fusing these features to craft a holistic image representation.

Following the pre-processing stage, the data undergoes a strategic splitting, segregating it into training, validation, and testing subsets (Sec. V-E). This ensures that the model learns from a diverse set of data and gets validated and tested on unseen data, which is crucial for understanding its generalization capabilities. An integral part of the training regimen is data augmentation. Given the inherent variability in medical images, data augmentation techniques (Sec. IV-B) such as rotations, zooms, and brightness adjustments are employed to artificially enhance the dataset's diversity. This augmentation is critical in equipping the DCNN to recognize and diagnose breast cancer across a wide array of mammographic images, thereby enhancing its robustness and accuracy.

The core component of the framework is the Deep Convolutional Neural Network (DCNN) (Sec. IV), which undertakes the actual task of learning and pattern recognition from the mammographic images. Designed with multiple layers, the DCNN delves deep into the images, identifying intricate patterns and features which are often imperceptible to the human eye, but are indicative of potential malignancies.

The final stage of the framework is the evaluation (Sec. V-D), where the trained DCNN model's efficacy is gauged. Comprehensive performance metrics, including accuracy, precision, F1-score, and recall, are employed to understand the model's diagnostic capabilities.

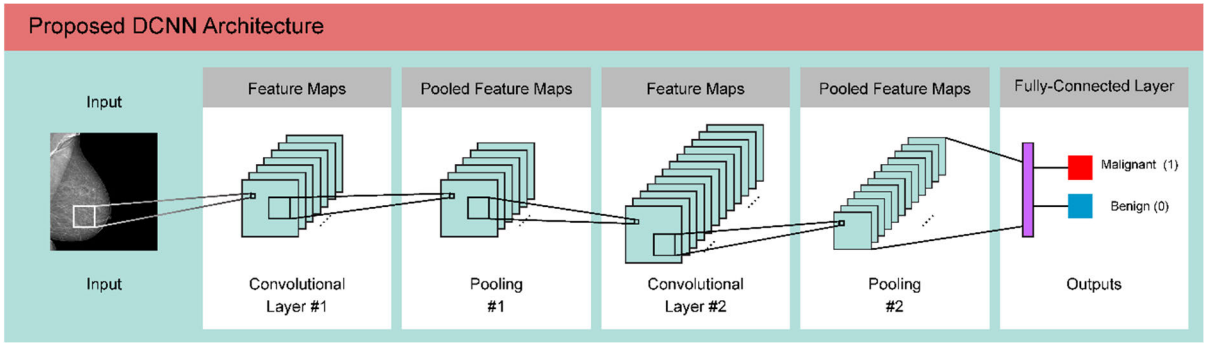

FIGURE 3. Proposed approach for Breast Cancer classification using DCNNs.

A. PRE-PROCESSING STAGE

In the pre-processing stage, we implemented effective techniques of image pre-processing that are straightforward to compute. These techniques allow the elimination of any undesired areas with the aim of improving local attributes. As a result, it is easier to detect any possible Regions of Interest (ROIs) and sub-regions within them.

We study several collected mammographic datasets as sources for our work and chose three distinct datasets of mammograms, as described in (Sec. V-A) these data sets would serve the following development of our proposed approach. This research, based on a rigorous analysis of available methods and literature, introduces various effective techniques for accomplishing tasks related to image background removal, image enhancement, and general pre-processing for DCNN when dealing with mammographic images.

When applying these techniques, the images are cleaned up to eliminate any undesirable background elements and noise. Next, we focus on the elimination of the pectoral muscle, connecting the front of the chest with the upper arm bones.

At this stage, the noise is partially reintroduced in order to get closer to the real-life scenario conditions.

This comprehensive pre-processing procedure allows us to enhance the efficiency of the model of neural network model in practical, real-life situations. Once all undesirable elements were extracted and eliminated, we proceeded with enhancing the images to highlight the desired ROIs and areas within them as part of the pre-processing.

When we finalized the processing with the deployment of the selected techniques, the images were prepared for the training phase, verification, and testing for the DCNN. Outputs were organized individually for each phase of the proposed techniques' implementation, including background elimination, pectoral muscle extraction, and the final enhancements. These results were instrumental in examining the effectiveness of the suggested techniques side by side with the current approaches.

1) BACKGROUND REMOVAL

Input image background is erased by removing the zero-intensity pixels from both columns and rows. Next, we used grayscale thresholding, taking advantage of Otsu's technique [39], to the refined image. The image's intensity was adjusted to a medium value between the lowest and highest values of its initial intensity.

2) IMAGE ENHANCEMENT

Good-quality image enhancement relies on accurate observations and is crucial in segmenting anomalous areas for the purpose of classification of a disease [40]. The image enhancement methods can significantly improve the mammography quality regarding the contrast and noise with one essential purpose: to aid computerized breast cancer detection techniques in pinpointing lesions in this area that are harder to detect and enhance the clarity of areas with insufficient contrast. In mammogram images, the areas with poor contrast pose a considerable challenge since they can conceal abnormalities within the tissue, leading either to false detection or, on the contrary, undetected lesions.

In order to improve the input mammography images, we utilize the technique Contrast-Limited Adaptive Histogram Equalization (CLAHE) [41]. CLAHE is an image enhancement technique that aims to improve the local contrast and visibility of subtle details in mammography images, which is crucial for accurate breast cancer detection using DCNNs. Mammography images are preprocessed before feeding them into the DCNN model. Preprocessing involves steps like resizing, normalization, and noise reduction to prepare the images for analysis. Then, CLAHE enhances the contrast of smaller, localized regions within the image rather than globally. This is important because mammograms often contain regions with varying levels of contrast and illumination. It adaptively divides the image into smaller tiles or patches and performs histogram equalization separately for each tile. This adaptive approach prevents over-amplification of noise and ensures that subtle structures within each region are better visualized.

3) FEATURES EXTRACTION

Next, the proposed technique involves the extraction of various statistical features from the preprocessed images. These are utilized as inputs for breast cancer detection in the

PSO-MLP [42] and ACOMLP systems [43]. The respective extracted features and characteristics include mean, kurtosis, skewness, variance, contrast, entropy, correlation, homogeneity, and energy.

The mean feature refers to the overall image brightness and is computed using the following equation:

$$mean\mu = \frac{1}{m*n}\sum_{i=0}^{m}\sum_{j=0}^{n} p\,(i,j) \tag{1}$$

$\sum_{i=0}^{m}\sum_{j=0}^{n} p(i,j)$ represents the aggregate of pixel values, while $m*n$ denotes the dimensions of the image.

Kurtosis refers to the peak tendency considering normal distribution. It is calculated using the equation.

$$kurtosis = -\frac{1}{m*n}\sum_{i=0}^{m}\sum_{j=0}^{n}\left[\frac{p\,(i,j)-\mu}{\sigma}\right]^4 - 3 \tag{2}$$

Skewness indicates the distribution of pixel values, measuring the dark or lustrous regions. Its calculation follows equation (3),

$$skewness = -\frac{1}{m*n}\sum_{i=0}^{m}\sum_{j=0}^{n}\left[\frac{p\,(i,j)-\mu}{\sigma}\right]^3 \tag{3}$$

While the expected deviation is determined using equation (4).

$$standarddeviation\sigma = \sqrt{\frac{1}{m*n}\sum_{i=0}^{m}\sum_{j=0}^{n}(p\,(i,j)-\mu)^2} \tag{4}$$

Variance quantifies the level of contrast within the image and is determined using equation (5).

$$variance = \sigma^2 \tag{5}$$

Contrast means the distinction between the highest and lowest pixel intensity within an image, and it is computed through the following equation.

$$contrast = \frac{(maximum\ intensity - minimum\ intensity)}{(maximum\ intensity + minimum\ intensity)} \tag{6}$$

Entropy defines the character of randomness and texture description. To determine this value, we use the following equation.

$$entropy = \sum_{i=0}^{m}\sum_{j=0}^{n} p\,(i,j)\,log_2(p\,(i,j)) \tag{7}$$

To calculate the correlation, a filter mask $Fil\,(x,y)$ is applied to the image. For each region, evaluate the sum of products to identify the objects within the image, regardless of their position, taking advantage of the equation.

$$Fil\,(x,y)\,p(i,j) = \sum_{i=0}^{m}\sum_{j=0}^{n} Fill\,(x,y)\,p\,(i+x,j+y) \tag{8}$$

The homogeneity feature tells us how the pixel values differ within the image. To calculate it, use the following equation.

$$homogeneity = \sum_{i=0}^{m}\sum_{j=0}^{n}\frac{1}{1+(i-j)^2}p\,(i,j) \tag{9}$$

Finally, the energy sums the distribution of gray in the image. It is derived from the image-normalized histogram, and we calculate it using equation (10).

$$energy = \frac{1}{m*n}\sum_{i=0}^{m}\sum_{j=0}^{n}(p\,(i,j))^2 \tag{10}$$

4) FEATURE FUSION

How to select the most optimal and representative features from the initially extracted set of features remains a subject of active research. The literature suggests numerous algorithms, and many have already been applied in the field of clinical imaging, including techniques like Particle Swarm Optimization (PSO) [44] and Genetic Algorithm (GA) [45]. These methods try to identify the most representative subset of features rather than operating with the entire space of features. A clear benefit of such feature selection is their ability to finetune system accuracy while also considerably reducing the computational demands of the process [46]. Nevertheless, in some cases, some key features are left behind during the selection, which can negatively affect the accuracy of the system. To address this issue, the researchers have introduced several methods of feature fusion deployed to raise the number of predictors and bolster the system's accuracy [47]. The serial and parallel fusion strategies are among the most recognized techniques in this field [48].

To create a fusion of the chosen features, we used a sequential probability-based method [49], where probabilities for both chosen vectors are first calculated, and then only a single feature is utilized based on its higher value of probability. Then, we conduct a comparison based on the feature with the highest probability and blend the two features into a single matrix. This comparison primarily addresses the problem of redundancy in both vectors' features.

Consider two feature spaces, specified as A and B, defined within the sample space of patterns marked Ω. For any sample $\xi \in \Omega$, the respective feature vectors are $\alpha \in$ A and $\beta \in$ B. The serial blended feature for ξ is expressed as $\gamma = \alpha\beta$. Logically, in case the feature vector α is of n dimensions and β is of m dimensions, the feature resulting from their combination, defined as γ, is of $(n+m)$ dimensions. Any pattern sample vectors of serial combined features collectively constitute a serial combined feature space of $(n+m)$ dimensions.

In the next step, we have subjected the final fused features to machine learning algorithms for classification. Following the fusion, the vector's dimensions were 4788 × 704, where 704 represents the feature quantity, and 4788 indicates how many images were involved.

B. DCNN MODEL

The initial convolutional layer of the DCNN, as shown in Figure 3, processes the pixel values of the input image. Each of the following convolutional layers seeks the connections between the neighboring pixels through the kernels that extract diverse features out of the image. A typical CNN architecture contains a convolutional layer, employing a set

of learning filters for the image. These filters can identify diverse elements such as edges, different textures, and specific patterns. The outcome of the convolutional layer has the form of a feature map, accentuating these significant features in the examined image.

Typically, a pooling layer is used next. This can include either average or maximum pooling function. It reduces the measurements of the map while preserving its most important features. This process helps with the simplification of the computation by downsizing the total size of data and pulling out exclusively the most useful information.

Once we have the pooling layer, a series of more pooling and convolutional layers can be stacked in sequence to grasp the more intricate features of the analyzed image. Such a hierarchy facilitates the neural network's understanding of the abstract models of the input data. Activation operations, such as the Rectified Linear Unit (ReLU) [50], incorporated after each convolutional layer, introduce non-linearity and transform negative values to zero while maintaining each positive value.

The final part of the network typically comprises fully connected layers responsible for processing the features and using them to establish predictions. The layers that are completely connected set links between all neurons in the preceding and current layers, promoting the network's capability to comprehend complicated relationships that manage the interconnection between the output labels and the relevant features. The pooling layer is used next. This can include either average or maximum pooling function. It reduces the measurements of the map while preserving its most important features.

This process helps with the simplification of the computation by downsizing the total size of data and pulling out exclusively the most useful information.

Once we have the pooling layer, a series of more pooling and convolutional layers can be stacked in sequence to grasp the more intricate features of the analyzed image. Such a hierarchy facilitates the neural network's understanding of the abstract models of the input data. The ReLU Activation function incorporated after each convolutional layer, introduce non-linearity and transform negative values to zero while maintaining each positive value.

The final part of the network typically comprises fully connected layers responsible for processing the features and using them to establish predictions. The layers that are completely connected set links between all neurons in the preceding and current layers, promoting the network's capability to comprehend complicated relationships that manage the interconnection between the output labels and the relevant features.

In the end, the CNN output layer is completely dependent on the exact task in question. For instance, in the case of classification assignments, a SoftMax layer is recommended to generate class-specific probabilities, whereas, in the case of segmentation assignments, it makes more sense to utilize a convolutional layer equipped with a pixel-wise activation feature. It's crucial to enable the variation of the architectural arrangement and layer sequencing of CNN to ensure its adaptation to the exact demands of the assignment and the dataset. We often try varying hyperparameters, architectures, and methods of regularization to make the performance more optimal and acquire precise and trustworthy outcomes. The outcomes are then forwarded to the following layers and processed through the kernel.

The combination of depths and measurements of these layers within the CNN determines the quantity of its parameters. Every parameter is associated with a weight obtained in the course of the training. The count of convolutional layers is affected by the dimensions of the filters as well as the layers' depth. If we increase the filter count or their size, we also get more parameters.

Since each link between the neurons depends on a parameter of weight, the parameter count within the fully connected layers is defined by the dimensions of the layers. Involving additional convolutional layers results in more parameters. However, there are other aspects in play, too. The parameter count within a neural network structure can also be affected by the dimensions and depth of convolutional and fully connected layers.

Including numerous convolutional layers enables CNNs to understand even more abstract and intricate forms of the input data. The layers located lower capture features on the low level, whereas the advanced semantic data is absorbed by some of the deeper levels. Adding more depth or increasing the size of the layers generally leads to an increased number of parameters. However, other structural choices like strides, pooling, or padding also play their role.

Convolutional layers filter the input information, facilitating the extraction of noteworthy features, including shapes, textures, and edges. Understanding these features is fundamental for the following layers to comprehend and sort the data effortlessly. The convolutional layers have the capacity to identify features independently from their precise spatial placing, making CNNs strong in translations or modifications in the input information.

1) FULLY CONNECTED LAYER

The fully connected layer, also referred to as the dense layer, is a neural network that sets links with every neuron in the preceding layer. This layer plays a key role in diverse deep-learning applications, e.g., recognition of speech, processing of natural language, or image classification [51].

In a dense layer, one vector serves as input, while another vector is utilized as output. Every neuron of this layer sums the weighted input values and uses an activation feature afterward to generate the output. During the training process, the gradient descent algorithm and backpropagation assist with determining the right biases and weights for the dense layer. In each training repetition, these are adjusted to minimize losing the function. While fully interconnected layers demonstrate adaptability and robustness, they also introduce

a large number of parameters. If not handled properly, this can easily lead to overfitting. Deep learning models often enforce strategies such as dropout and weight decay to address this issue [52].

2) LENET ARCHITECTURE

LeNet Architecture [53] is a preferred tool for image classification assignments, offering numerous advantages that other DL classifiers lack. These include spatial invariance, parameter efficiency, activation features, weight sharing, hierarchical feature learning, and pooling layers. Regarding ensembling and image classification, a revised LeNet CNN is used to enhance accuracy through numerous models. This revised LeNet architecture involves further pooling and convolutional layers to enhance the depth of the model and catch more intricate features. Additionally, the classifier implements normalization of batch [54] and constant output to avoid overfitting and strengthen generalization.

In any case, it's important to acknowledge the limitations of existing models for cancer detection and classification utilizing LeNet architecture. For example, while LeNet can be effective when it comes to the classification and diagnosis of cancer, it has restricted flexibility and capacity. It's also vulnerable to overfitting and has demanding pre-processing and hardware requirements, all posing considerable challenges [55]. Addressing these challenges, the current model strives to refine the architecture by applying the following adjustments:

Hyperparameter optimization boosts the performance of the model, and it contributes to the decision on the most suitable setup for the exact assignment. A meticulous study of the hyperparameter field and a deep understanding of the reciprocations associated with various decisions are needed. Via optimization of the hyperparameters, modified models can reach prime performance on the provided dataset, making it a vital step in the process of DL model optimization. Here are the key considered hyperparameters:

- The 0.01 learning rate defines the size of the step per every iteration in the training process and contributes to accomplishing optimal performance.
- A dropout rate of 40% (0.4) hold control over the model's regularization and avoids overfitting.
- A batch size of 32 defines how many samples pass through the processing before the weights of the model are edited. This accelerates convergence and also improves generalization, affecting the model's overall performance.
- There are three hidden units, which has a great influence the capability and effectiveness of the model, being also. crucial to establishing the proper balance between the overfitting and complexity of the model.
- There are ten training reiterations, defining the number of repetitions the model completes over the whole dataset. Again, this results in a symmetry between overfitting and underfitting of the model.
- The ReLU activation features in the convolutional layer model and SoftMax in the fully connected layer model can be subjected to experimentation to assess which function works the best in the particular scenario.
- The set of filters in the Conv layer is adjusted to 32, 64, and 128, allowing adjusting to specify the model's capability and complexity.
- The size of the kernel in the convolutional layer is 3, 5, and 4 with enabled adjustments to address various spatial designs in the input information.

The preferred optimizer plays a vital role in the performance of the model, too, since this feature adjusts the network's characteristics, including its learning rate and weights. Typically, the optimizer of choice is either ADAM [56], which is our choice, too, since it's suitable for breast cancer datasets, RMSprop [57], or Stochastic Gradient Descent (SGD) [58] with momentum. These optimizers may also have their own unique hyperparameters (e.g., decay rates) that impact their accuracy, help reduce losses, and enhances their overall performance.

When we approach the model in this specific manner, it helps us evaluate how effective it is when dealing with the data it hasn't seen before, and it's a good prevention for overfitting. The machine learning model is fed with pre-processing images as an input. The LeNet model requires the images to be processed in a format suitable for its effective processing. This adjustment typically involves amending dimensions of the images to a specific size, standardizing the values of pixels, and other potential transformations based on the demands of the architecture. What sets LeNet apart from other networks is how it incorporates the sharing of weight and picks the activation features [59]. LeNet is a traditional convolutional neural network architecture, but it possesses several distinctive features, making it notably efficient with the image classification-related assignments:

- Activation: In its dense layers, LeNet employs the sigmoid activation feature. This function compresses the output of every neuron so that it's above 0 and below 1. Such a non-linear transformation allows the modelling of complex interconnections between diverse features.
- Simplicity: The LeNet architecture is defined by a straightforward structure consisting of sequences of convolutional layers and pooling layers, which are then succeeded by dense layers.
- Weight sharing: To enhance its effectiveness and overall performance, the LeNet architecture employs a weight sharing technique, where the identical set of weights is shared across various spatial zones in the input. This helps the network to detect and extract common features across the entire image, reducing the number of general parameters.
- Combination of pooling and convolution: The LeNet architecture uses convolutional layers to pull out the spatial features. These layers utilize filters to process the input, identifying locally specific patterns and mapping

the features. Additionally, the architecture also takes advantage of the pooling operations, particularly max pooling, which are responsible for the downsizing of the maps and reduction of the spatial measurements of the features. This is carried out through the selection of the limit value for each pooling area, reducing the spatial proportions of the features.

Generally, the LeNet architecture's distinctive traits, such as weight sharing, use of the sigmoid activation, simple structure, and the combined pooling and convolutional processes, prove it to be a leading architecture for the classification of images and performance improvement.

3) BATCH STANDARDIZATION

Batch standardization, also known as normalization of the batch [60], is a method utilized by the LeNet model to enhance the effectiveness of the operated neural networks. This includes different classification assignments, such as breast cancer identification with the help of ultrasound imaging techniques.

Using batch standardization in this type of scenario comes with various advantages. It can, for example, improve and speed up the confluence of the process of training by minimizing changes in the network parameters (inner covariate shifts), demonstrated by the changed input distribution across the layer because of the shifted weights of the other layers. Batch standardization [61] is also able to decrease inner variability through standardizing data per layer and assist with prompt stabilization of gradients during the recovery phase.

Batch normalization comes with regularization, enhanced accuracy, and even better flexibility. Taking all these benefits into account, using the LeNet architecture that incorporates batch standardization for the detection of breast cancer through ultrasound can lead to better results and overall effectiveness on this assignment.

4) REPLACEMENT OF THE POOLING LAYERS

In order to enhance data retention, improve the extraction of features, and possibly even the overall performance of the model, introducing the stride 2 convolutional layers in place of the pooling layers within the LeNet architecture can be considered [62]. This consideration is backed up by the potential constraints of pooling procedures since these operations can leave some data out of the maps of features. This is particularly probable in the case of large kernel sizes or high strides.

Using stride 2 convolutional layers instead of the pooling layers can result in the retention of more data and valuable details in the maps. As a result, it also reduces the risk of losing important data during the following downsizing process.

Unlike pooling layers, convolutional layers are capable of learning and comprehending more abstract and complex features. This is particularly true in cases where the layers can work with a large variety of parameters and more depth. A model equipped with convolutional layers is able to pull out more useful features from the original images than if it had pooling layers, which eventually leads to more accurate results of the classification assignment.

This replacement also has other advantages. It can, for example, lead to lower costs of computation (again, especially in the case of high strides and large kernels), which are expensive to process through pooling procedures. This burden can be diminished by substituting the pooling layers with the stride 2 convolutional layers, resulting in a model that requires less time for training and evaluation. In general, substituting the pooling with convolutional layers profoundly improves data retention and leads to more efficient extraction of features, overall performance, and cost efficiency of the deployed LeNet architecture.

5) ACTIVATION OF RELU

The network, in general, acquires non-linearity through the activation of individual layers within a network, which drives the activation of the prior layer. The suggested architecture proposes a mesh composed of five groups of convolutional stack standardization. Substituting the feature of sigmoid activation with the alternative ReLU function practically modifies the architecture for cancer detection based on ultrasound. This conversion comes with a number of advantages, such as better performance, improved abstraction, quicker computation, expanded learning capabilities, and elimination of the issues with loss of gradients.

The problem of vanishing gradient occurs when the activation gradients are too small, hindering the modification of weights in the course of training. The gradients associated with the ReLU activation function are less inclined to this problem compared to the sigmoid gradients, which allows the network to learn the patterns and features of the input images. This conversion seeks to improve accuracy, enhance understanding, and boost the specificity of the results.

6) DROPOUT LAYERS

Enhancing the LeNet architecture deployed in breast cancer diagnostics with dropouts is another useful technique for improving its performance. Dropout is a method of regularization that unsystematically sets some activations within the network to zero (it literally ''drops out'') in the course of the training [22], [51]. Typically, the ratio of dropouts is around 40% (or 0.4) across all the relevant layers (Dropout (0.4)). There are two dropout layers incorporated behind the convolutional layers, plus two more inserted in front of the final fully connected layer. Hence, there are four deployed dropout layers in total. This method is helpful in preventing the model's overfitting in relation to the training data by reducing the mutual adaptation of the utilized neurons. As a result, the model is perfectly equipped to extend to new, previously unseen information, which is critical for precise cancer detection, as well as for its powerful learning ability [63].

By unsystematically dropping activations out in the training phase, the network learns to assess a broader range of attributes and patterns within the provided input. Such an experience improves the robustness of the model and its adaptability to changing input information, including the variable demographic factors or conditions of imaging.

Overall, utilizing the dropout layers can decrease the sensitivity of the model to the original conditions of the network, such as different biases and weights. This results in increased reproducibility of the obtained outcomes and decreased hazard of the model getting blocked at the minimum in the training phase. Through decreasing the mutual adaptation of the neurons and teaching the network to focus on a wider range of characteristics and elements, the dropout utilization method can boost the efficiency of the training and strengthen it against the risk of getting stuck on the level of local minimum [64].

C. AUGMENTATION OF DATA

In order to get the most out of our restricted training samples and enhance the accuracy of our model, we used a range of random modifications on our data [65], including the rescaled size, rotation (255 degrees), zooming, and scrolling in horizontal direction.

The augmentation of the data is also predicted to avert overfitting - a familiar problem in the field of machine learning, occurring in cases when the model is introduced to a very limited set of examples, and it learns the patterns without being adapted to unseen data.

This approach improves the model's capability to conclude generalized assessments. Furthermore, since the augmentation generates identical quantity of images for every class, it also contributes to creation of well-balanced datasets and, thus, unbiased comparison of the outcomes.

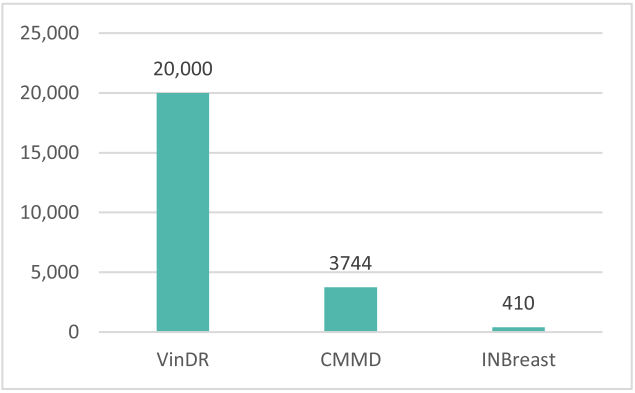

FIGURE 4. Datasets statistics.

D. PARAMETERS DEFINITION

There are several parameters we should focus on during the training phase in order to maximize the quality of our suggested network's performance in solving the problem in question. These include:

- Volume of the batch: The batch size defines how many training images will pass the process (forward/backward). Remember, though, that the batch size also affects the memory demands of the process.
- Iterations: This parameter refers to the number of repeated forward/backward passes in every step of the process, utilizing a given batch of images according to the general size of the batch.
- Epochs: This is the parameter that defines how many times an image rotates in the course of the training. Every single epoch stands for one appearance of the image, respectively, for the full run of the entire set of training samples. Numerically, this is calculated as:
- Loss: This function assesses the loss between the prognosis and the tagging of the absolute truth for every batch.
- Pace of learning, also referred to as a learning rate, is a parameter that represents the size of the step for revising the model's weights regarding the SGD.
- Optimizer: There are a wide range of optimizers to be deployed in order to identify the most suitable combination of parameters for your model, including ADAM, SGD, and RMSprop.

E. ENHANCED MODEL INTERPRETABILITY WITH GRAD-CAM

To further advance the clinical applicability and trustworthiness of our DCNN model, we have integrated the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [66]. This method provides a visual explanation of the model's decision-making process by producing heatmaps that highlight critical areas influencing the model's predictions [67]. Upon processing mammographic images through our DCNN, the Grad-CAM technique was applied to the final convolutional layers to generate a heatmap overlay. The highlighted regions correspond to features such as calcifications, masses, and architectural distortions, which are pivotal in diagnosing breast cancer. These heatmaps serve as a valuable interpretative tool, offering clinicians a visual representation of the model's focus areas, which supports the BI-RADS assessment categories used in clinical practice. The integration of Grad-CAM not only enhances the interpretability of our DCNN model but also provides a valuable bridge between AI functionality and clinician expertise.

V. EXPERIMENTAL RESULTS

The following section explains which dataset and metrics for evaluation were applied in our study. Furthermore, we examine the experimental outcomes of the suggested model of architecture.

A. DATASETS

The following section briefly summarizes and describes the datasets we have applied in testing our proposed approach for

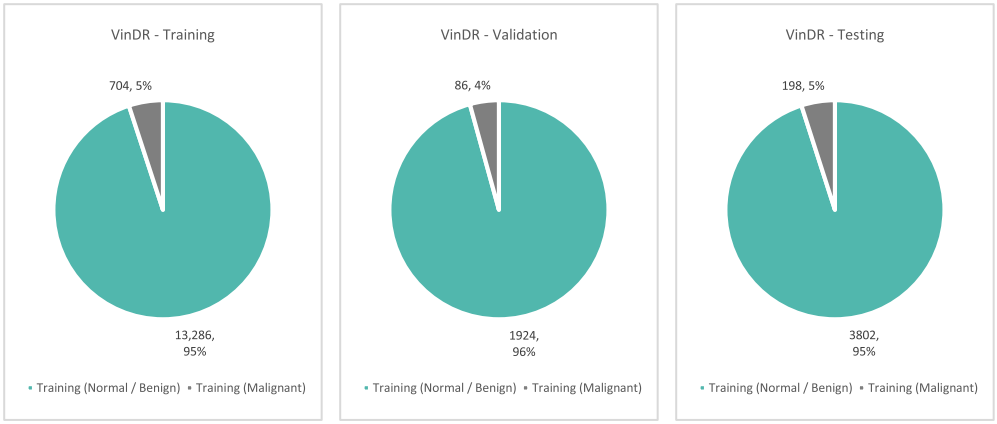

FIGURE 5. VinDr-Mammo dataset distribution.

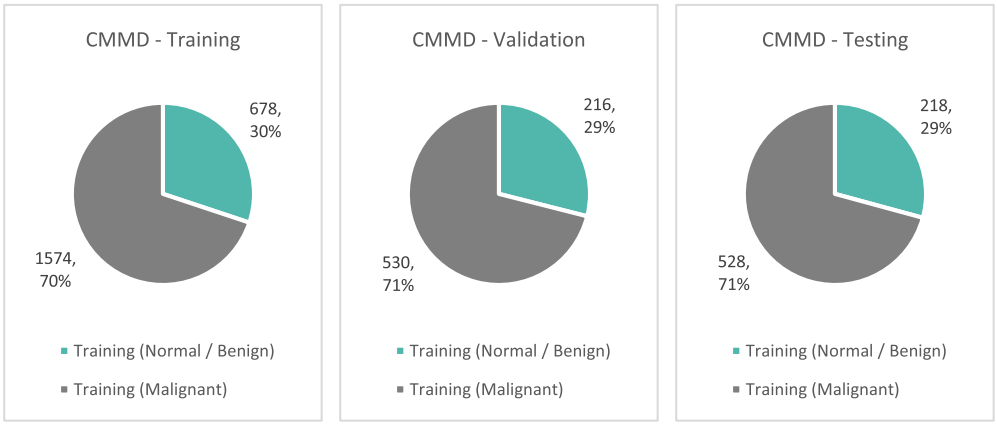

FIGURE 6. CMMD dataset distribution.

breast cancer detection. Figure 4 shows the instances count of the three datasets used in our work.

1) VINDR-MAMMO DATASET

We started our training with the VinDr-Mammo [68] dataset, which is a public dataset used of more than 20,000 image items coming from approximately 5000 diverse patients. These images come from two medical facilities located in Vietnam: Hospital 108 and Hanoi Medical University Hospital.

The hospitals possess mammography technologies produced by Giotto, Siemens, and Planmed. The VinDr-Mammo dataset provides a good example of a considerable class imbalance, since just 988 (5%) of the images are tagged as being malignant, which is comparable to a regular screening where the vast majority of the results are either standard or benign.

The researcher for dataset have already divided the data in two categories: the training dataset comprises of 16,000 images and the group for testing contains 4000 images as shown in figure 5. This ensures the consistency of the reported outputs.

2) CMMD DATASET

Another publicly available dataset we used for our training was CMMD, which stands for Chinese Mammography Database [69], one of the newest available datasets containing 3744 images coming from 1775 patients. This dataset comes from the South China University of Technology. All the images were produced by the GE Senographe DS Digital mammography technique. This set contains a higher ratio of malignant findings compared to the benign images, which corresponds to the standard narrative of clinical diagnosis: first, some suspicious finding is detected, followed by its confirmation through additional imaging.

Figure 6 shows the splitting of this dataset into two groups. The set for training contains 2998 images, which is 80% of the entire database, whereas 746 images (20%) are selected for testing. The images were stratified by class, and we ensured that if the set contained numerous images from the same patient, they all stayed in the same group.

Next, we split the training groups from both involved datasets into two sub-groups: one will be used in the training, and the other will serve the following validation. The validation will fine-tune hyper-parameters and to search for the

"most representative" model epoch throughout the training. The divisions were carried out based on the class. Once again, we maintained numerous pictures coming from the same patients in one of the groups. For every evaluated method addressing the imbalance of the class, our models were individually trained using each of these datasets. The efficiency and overall performance of the models were evaluated on each set individually. This came with some benefits. For instance, testing a model that was trained using the VinDr-Mammo training set and testing it using the CMMD test set gave us the opportunity to evaluate the model's ability to generalize on the datasets it has not encountered during its training.

3) INBREAST DATASET

The third and last dataset involved in our study is the well-known INBreast [70] public dataset. Nevertheless, we used this dataset exclusively for testing of the model's performance throughout the experiments as shown in figure 7. This allowed us to compare our study with other related studies and experiments in the field. The INBreast dataset is built around 410 FFDM pictures obtained by a Siemens mammography machine.

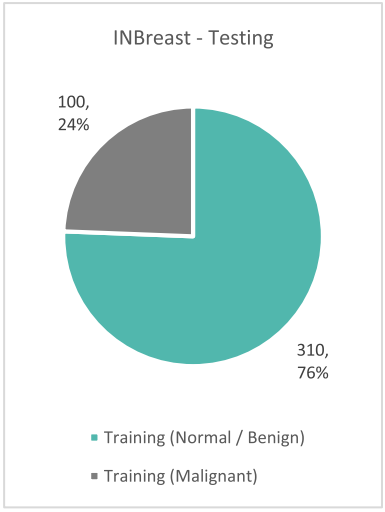

FIGURE 7. INBreast dataset distribution.

Based on the parameters of the classification procedure, we selected 100 epochs. During this evaluation, we reviewed the development of precision and loss graphs against the number of epochs, with every epoch consuming 300 seconds from the used GPU. Next, we conducted tuning with a small rate of learning. Besides that, we also deployed the ADAM optimizer [56] in an attempt to positively diminish the loss.

B. IMBALANCED CLASSIFICATION

One of the greatest challenges associated with the analysis of medical imaging is coping with data that is not balanced. This issue is particularly evident in this specific field, but some tools and methods exist to minimize its effects.

Our study utilized the Focal Loss feature [71], which forces the model to relieve the weight of the simple instances to acknowledge hard samples. It is characterized by adding a modulating element to the cross-entropy losses and class-balancing parameters. It's defined by the following equation:

$$Fl\,(p_t) = a_t(1-p_t)\gamma \log(p_t) \tag{11}$$

where p_t refers to the true class' anticipated probability; a_t stands for the class-balancing factor applied to the true class; γ defines the focusing parameter in the control of the down-weighting for examples with the correct classification.

Evaluation Metrics

To review and evaluate the effectiveness of the presented approach, we used the following parameters:

- Accuracy: This quantifies the number of exact predictions, dividing it by the absolute count of samples.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{12}$$

- Sensitivity: This metric measures the ratio of true-positive results. If the sensitivity is close to 100%, it indicates a high likelihood that a patient has indeed a positive diagnosis.

$$Sensitivity = \frac{True\ Positive}{False\ Negative + True\ Positive} \tag{13}$$

- Precision: Precision computes the fraction of applicable examples among all the retrieved samples.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{14}$$

- Specificity: This parameter specifies how often an outcome indicates a true-negative finding. If this metric is close to 100%, it means a great likelihood of the absence of the disease for the evaluated patient.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{15}$$

C. TRAINING

As we investigated various ways to optimize the training process and achieve outstanding results, we agreed to incorporate a convolutional layer atop the current layer with the highest available accuracy. Figure 8 displays the levels of accuracy for the used datasets, which is ideal for rapid assessment of the architecture.

Accuracy, Precision, Recall and F1-Score values according to the number of layers. The F1-Score measurement strives for well-balanced precision and recalls. The figure allows us to see that the F1-Score for the three-layer architecture shows satisfying accuracy and recall. This confirms a three-layer model is a good choice for further training.

Subsequently, Figure 9 shows outcomes for every individual dataset according to the metrics resulting from the tests on all the layers. We thoroughly compared results for each dataset individually. This evaluation allows us to assess

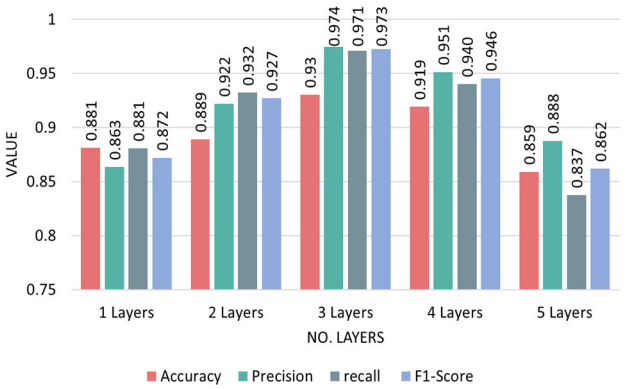

FIGURE 8. Accuracy, Precision, Recall and F1-Score values according to the number of layers.

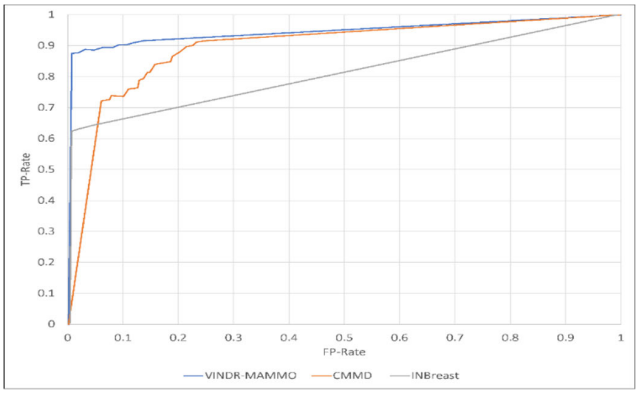

FIGURE 10. ROC Curve analysis for VINDR-MAMMO, CMMD and INBreast datasets.

the individual efficiency of each dataset (as opposed to the search for a single superior dataset), given the small volume of images from the INBreast dataset.

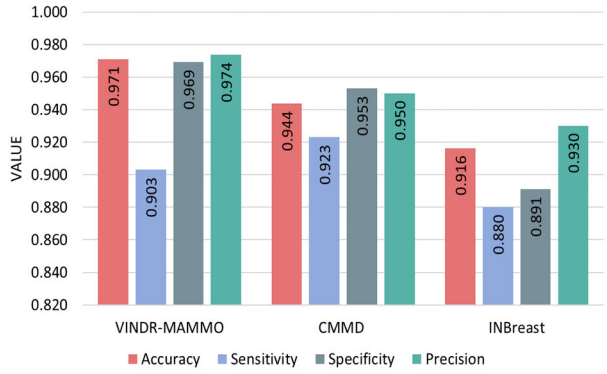

FIGURE 9. Comparison of training metrics for each dataset.

Figure 10 also illustrates the ROC curve used to assess the binary classification performance. Across all the involved datasets (VINDR-MAMMO, CMMD and INBreast), the ROC curve shows sensitivity in contrast with specificity in distinguishing malignant cancer from benign findings. Our suggested strategy generates a greater number of true-positive results compared to false-positive results. This indicates the deployed threshold of discrimination is able to precisely classify the dermoscopy input images, especially in combination with the VINDR-MAMMO dataset.

D. VALIDATION

The outcomes obtained from our presented model underline the possibility of achieving powerful performance by leveraging the initially suggested architecture for the datasets of image classification. This implies that the model is capable of effective generalization across a wide spectrum of classification jobs, even in the cases when the input images were absent from the dataset it was trained on.

Figure 11 illustrates the confusion matrix depicting the model's performance during the classification evaluated through the dataset for validation.

The model's assessment should take into account how many real melanomas were misclassified as benign. This is a critically concerning situation that would indicate the model's inability to catch genuine melanomas, which could lead to life-threatening consequences for patients.

Figure 11(a) presents the confusion matrix derived from the VINDR-MAMMO dataset. This matrix portrays the model's performance in differentiating between malignant and benign cases. Among the 265 malignant samples, the model demonstrated a substantial proficiency by correctly classifying 263 cases while only 2 were misinterpreted as benign. Similarly, of the 265 benign instances, 259 were accurately identified, and 6 were mistakenly predicted as malignant. The vivid differentiation between the true labels and the predicted ones underlines the model's capability to discern with high specificity and sensitivity in this dataset.

Figure 11(b) delineates the results from the CMMD dataset. The matrix accentuates that out of 530 malignant instances, 512 were accurately pinpointed, and 18 were erroneously categorized as benign. On the contrary, among the 216 benign images, 192 were correctly classified, with 24 misjudged as malignant. This assessment accentuates the model's robustness in handling datasets with an imbalanced proportion of classifications.

Figure 11(c) illustrates the model's performance on the INBreast dataset. The matrix indicates that, from the 100 malignant samples, 70 were aptly classified, but 30 were misinterpreted as benign. Among the 310 benign samples, 305 were identified correctly, leaving a marginal 5 cases misclassified as malignant. The relatively higher number of misclassified malignant cases calls attention to potential challenges in distinguishing subtler differences within this specific dataset.

Lastly, Figure 12 represents the Grad-CAM heatmaps for the proposed breast cancer detection using DCNN. In the top row, original mammographic images. While the bottom row

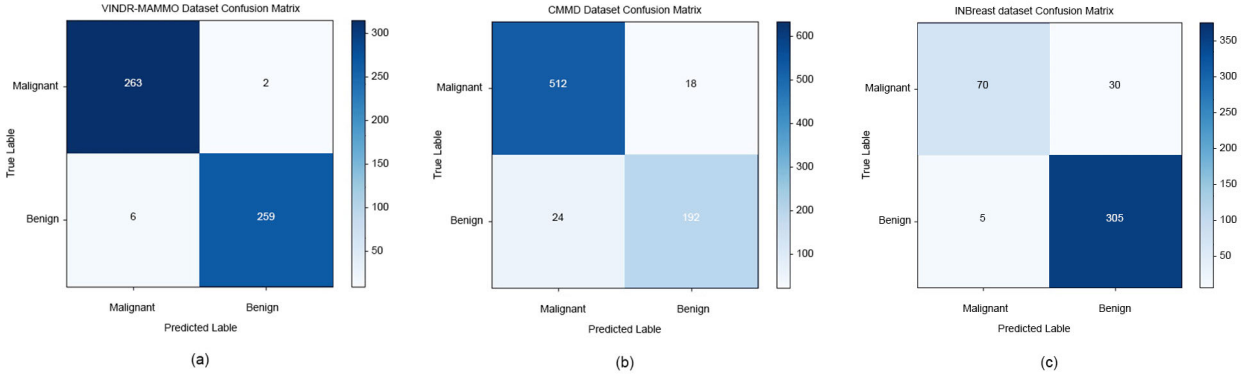

FIGURE 11. Confusion matrix of VINDR-MAMMO, CMMD and INBreast datasets.

shows the corresponding Grad-CAM heatmaps overlaid on the images, highlighting suspicious regions with higher intensity indicating areas with greater influence on the model's output. The regions marked are associated with suspicious calcifications and architectural distortions, as classified by the DCNN, with BI-RADS 5 indicating a high suspicion of malignancy.

VI. EXPERIMENTS OF FEDERATED LEARNING APPROACH

Considering the continuous challenges in the transmission, management, and discretion of breast cancer patient data, we conducted a number of simulations as part of this study to examine the efficacy of a federated model, working with data shared across various institutions, and putting it into contrast with a local model.

To be more precise, we executed four experiments, each in both centralized and federated architecture types. Here's the list and description of the performed experiments:

- 1st: One client with a local model that was trained on 100% of the available data.
- 2nd: Two clients, each with their respective local models, with the training data distributed in a 40%/60% ratio for each of them.
- 3rd: A federated model including the two clients with the previously mentioned data distributions.
- 4th: A federated model revolving around five clients and one global server, each equipped with a local model, where the training dataset is divided into 15%, 15%, 20%, 20%, and 30% portions.

Various metrics and parameters, including loss, accuracy, sensitivity and specificity, were used to assess the performance within these experiments. In total, 1000 epochs were set, meaning the dataset must be subjected to a thousand iterations during the training phase. Evaluations were conducted after each epoch using the test dataset.

In the case of the federated scenario, the training involved 100 rounds with 10 epochs. Next, in the federated architecture, local models were aggregated and updated across the whole network. The server aggregated the results after every 10 epochs of local training, and that's when the evaluations were also conducted. At the end, we analyzed the impact of varying data volumes across diverse clients and whether the federated method offered improvements in various scenarios. Table 1 represents a summary for the experimental values for different performance metrics.

TABLE 1. Local model vs. federated learning performance comparison of four experiments.

Experiment	Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)
1st	Local	81.0	85.1	75.4
2nd	Local	87.5	80.0	90.1
3rd	Federated	91.0	88.2	92.7
4th	Federated	98.9	95.0	98.0

A. 1st EXPERIMENT

Figure 13 illustrates that utilizing the entire dataset delivered positive results. The graph of Accuracy indicates that this model boasts an 81% likelihood of successful predictions. In Figure 14, the graph of Loss shows a consistent reduction in error values throughout the training process, without any apparent signs of underfitting or overfitting. The sensitivity is set at 85.1% and specificity at 75.4%. These values reflect the performance of a model trained on a comprehensive dataset, which, while robust in detecting true positives, may not be as adept at identifying true negatives due to potential overfitting or lack of diverse data representation.

B. 2nd EXPERIMENT

In the 2nd experiment, the dataset was split between two clients, where the first one had 40% and the second client received 60% of the training data. The training involved both clients. As displayed in the graph of Accuracy (Figure 15), starting from around epoch 400, there's apparent overfitting with the first client's model. The accuracy increases to 87.5%, accompanied by a sensitivity of 80.0% and a specificity of 90.1%. This suggests that although the model has improved in overall accuracy, the imbalanced distribution of data between

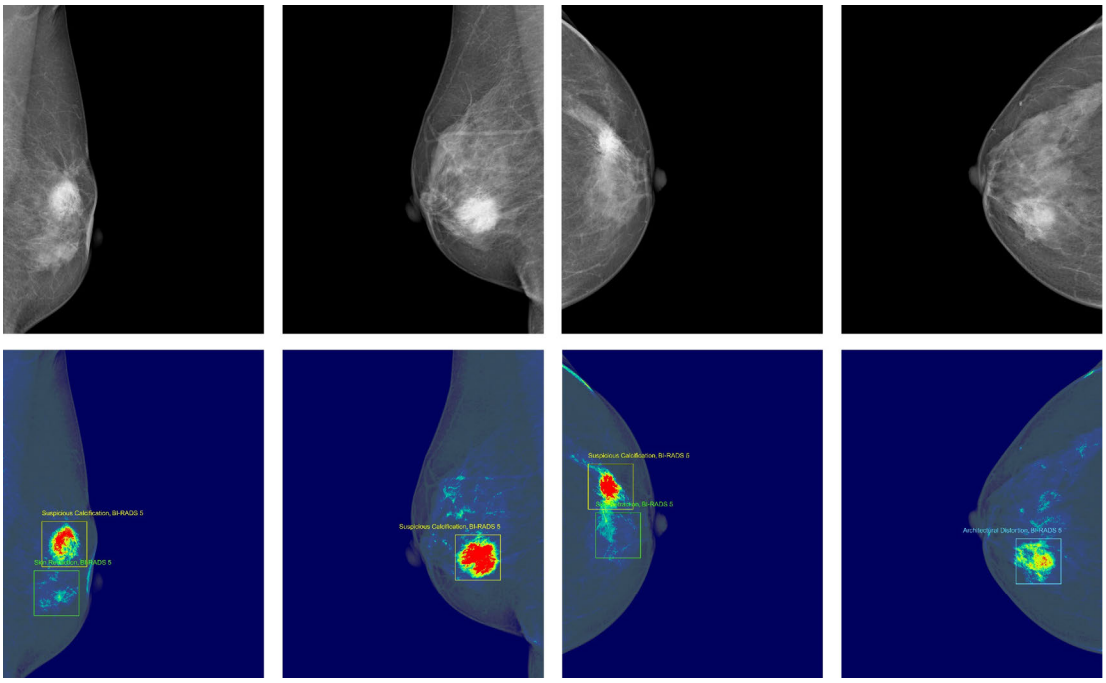

FIGURE 12. Grad-CAM Heatmaps for Breast Cancer Detection using DCNN.

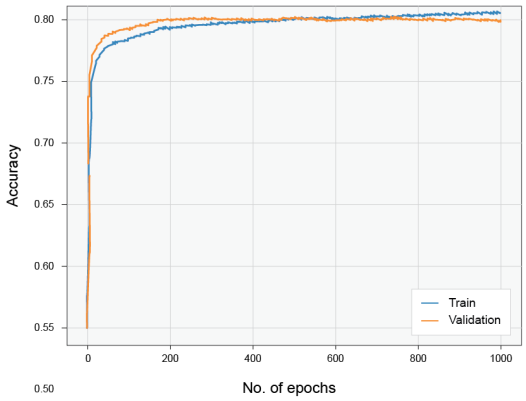

FIGURE 13. Accuracy of single client experiment with a local model trained on 100 % of the training data.

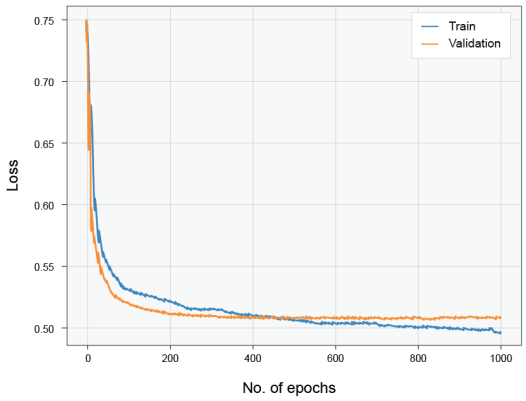

FIGURE 14. Loss graph of single client experiment with a local model trained on 100 % of the training data.

the two clients could lead to a better performance in recognizing negatives over positives. This is indicative of a model that may be more conservative in predicting the presence of cancer, thus reducing false positives but possibly at the expense of missing some true positive cases.

This occurs because of the diminished training data, in combination with a complex architecture. The model focuses excessively on the training data, harming its capacity to generalize with the testing set. This is affirmed in Figure 16, where we see the loss rising instead of decreasing. In this scenario, the metrics are less favorable compared to the client with the complete dataset due to the occurring overfitting. In any case, though, the overfitting isn't too severe in this case, and it does not interfere with the model's functioning.

Regarding the second client, overfitting isn't very noticeable because of the larger volume of data. Looking at Figures 17 and 18, we see a similar result as in the case of the client with the full dataset, thanks to the larger volumes of data.

C. 3rd EXPERIMENT

In the third of our experiments, we simulated a federated architecture involving one global server and two clients who shared training data in a 40% and 60% ratio.

The experiment shows further improvement with an accuracy of 91.0%, sensitivity at 88.2%, and specificity at 92.7%. The federated approach here begins to exhibit its strengths, as the aggregation of insights from two clients enhances

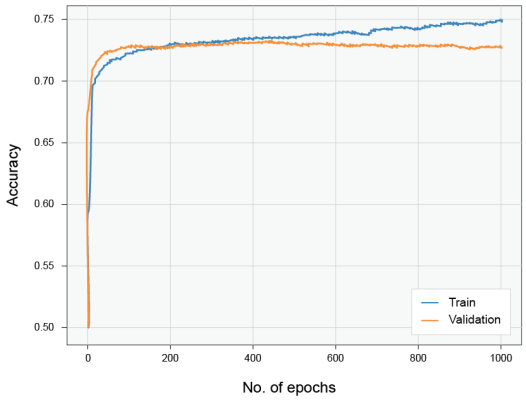

FIGURE 15. Accuracy experiment of the local client with the 40% and training data.

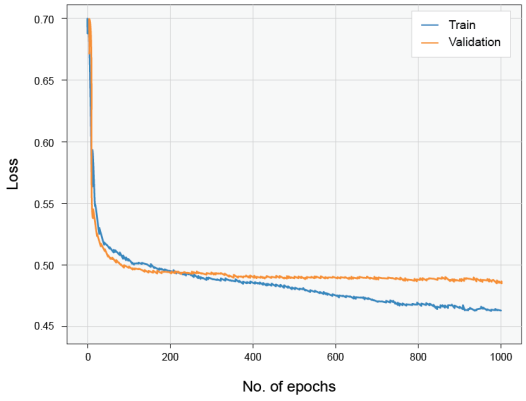

FIGURE 16. Loss graph experiment of the local client with the 40% and training data.

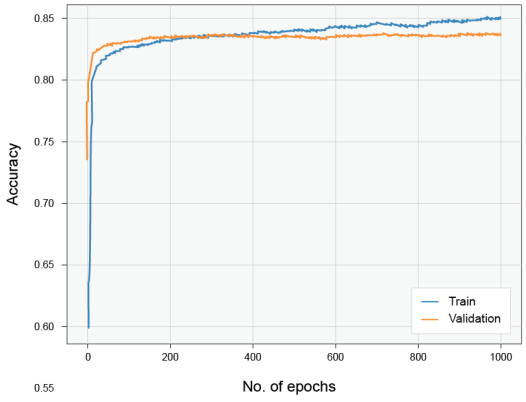

FIGURE 17. Accuracy experiment of the local client with the 60% and training data.

the model's ability to generalize, thereby achieving a more balanced and higher performance in both sensitivity and specificity. Figure 19 shows that they acquired nearly identical outcomes as those acquired through the centralized scheme with 100% availability of the data for training. Furthermore, this experiment also resulted in an intricate observation: the curves that were relative to the testing sets

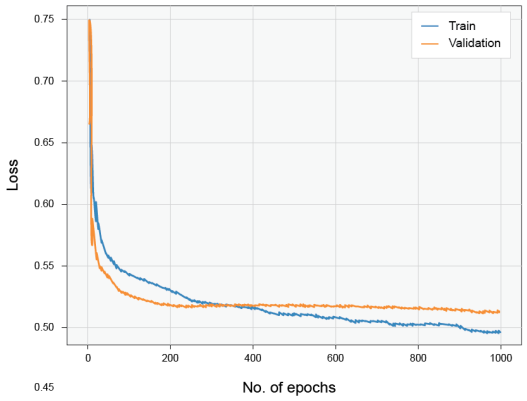

FIGURE 18. Loss graph experiment of the local client with the 60% and training data.

surpassed those related to the training dataset. This suggests that learning is progressing properly with no signs of underfitting or overfitting, as depicted in Figure 20.

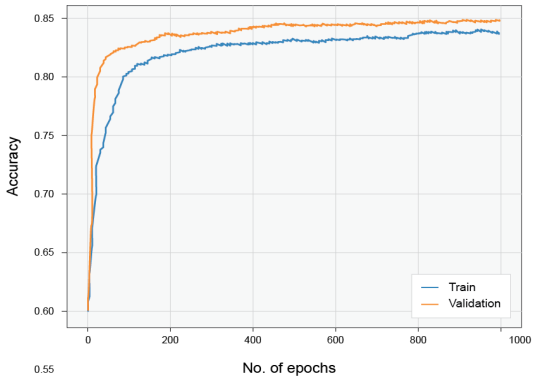

FIGURE 19. Accuracy of federated model experiment of the two clients with the different data distributions.

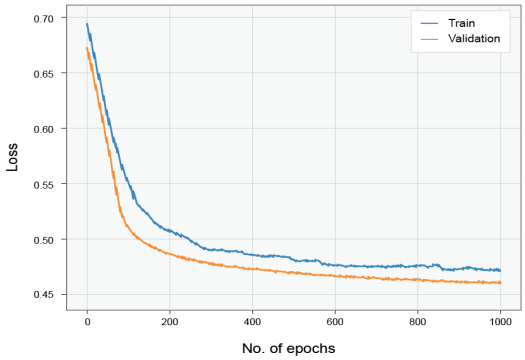

FIGURE 20. Loss graph of federated model experiment of the two clients with the different data distributions.

D. 4th EXPERIMENT

This time, we simulated an architecture with five clients. The training dataset is split in the following manner: two clients

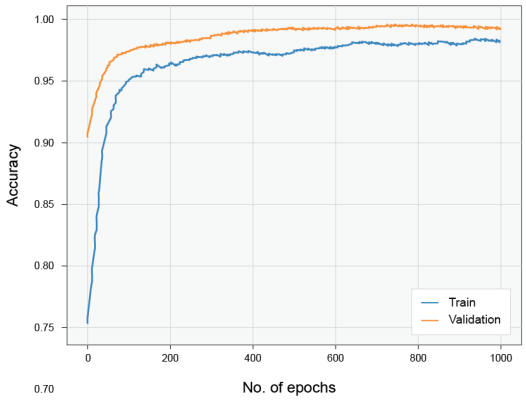

FIGURE 21. Accuracy of federated model of the global server that aggregate the client's nodes results.

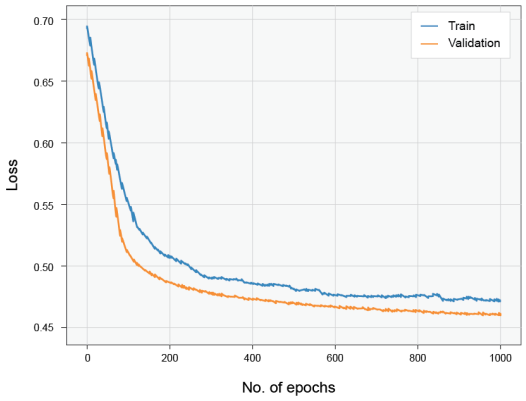

FIGURE 22. Loss graph of federated model of the global server that aggregate the client's nodes results.

received 15%, two more clients worked with 20%, and the last remaining client had 30% of the data. Each of them trained independently so that we could compare the outcomes. This simulation is based on a real-life scenario where some clients (such as large hospitals) possess a vast amount of data while others (like small health centers) have significantly less data.

Upon studying Figures 21 and 22, it becomes clear that the outcomes improve significantly by combining local models with a federated approach on the global server. The accuracy jumped clearly and the loss value is also notably small. This experiment demonstrates the most significant performance leap, with the accuracy reaching 98.9%, and sensitivity and specificity at 95.0% and 98.0%, respectively. The federated model, with data contributions from five clients, benefits from a rich diversity in the training dataset. This leads to a highly accurate model that excels at both identifying cancer cases and minimizing false alarms. The comprehensive data distribution ensures that the model is well-rounded, making it extremely reliable for clinical applications.

Comparable observations resulted from the federated architecture involving two clients, where everyone contributes individual knowledge that, when united, improves generalization and leads to more precise predictions. It's probably noteworthy that, in this case, there are also some client nodes with limited data, so the results resemble, or even surpass, those of the local subject with the complete dataset.

The progression in performance metrics across the experiments underscores the effectiveness of the federated learning approach in enhancing both the accuracy and reliability of breast cancer detection models. It showcases the potential of federated learning to utilize distributed data sources for developing robust diagnostic tools in healthcare.

VII. EXPERIMENTS ON DATA PRIVACY AND SECURITY

In this section, we have deployed Homomorphic Encryption (HE) [72] within our Federated Learning framework as a pivotal measure to ensure the security and confidentiality of data, particularly critical in the context of breast cancer early detection.

Mainly, HE allows for specific types of computations to be carried out directly on encrypted data, producing an encrypted result that, when decrypted, matches the result of operations performed on the plaintext [73]. For instance, if E represents the encryption function, D the decryption, and $\oplus$ an operation (like addition or multiplication), then for any two data points a and b, HE ensures that $D(E(a) \oplus E(b)) = a \oplus b$. This property is invaluable in federated learning, where data privacy is paramount. By using HE, models can be trained on encrypted data from multiple sources without exposing sensitive information. HE allows for computations to be performed on encrypted data, enabling the participating nodes in our federated network to contribute to the collective learning process without exposing sensitive patient data. This approach not only adheres to stringent data privacy regulations but also maintains the integrity of medical data.

By integrating HE, we aim to provide quantifiable results that demonstrate the efficacy of this privacy measure. The experimental results presented in this section are designed to empirically evaluate the impact of HE on the model's performance, including accuracy and computational overhead, thereby offering a comprehensive analysis of its feasibility and effectiveness in a federated learning setting.

Figure 23 illustrates the performance of both the non-encrypted and HE models in the federated learning system. Both models demonstrate an increase in accuracy over epochs, but the model with homomorphic encryption starts with a slightly lower accuracy, potentially due to the complexity added by encryption.

However, it gradually catches up, indicating that encryption does not significantly compromise the model's learning capability.

This representation emphasizes the high accuracy achievable in sensitive medical applications like breast cancer detection, even when stringent privacy measures are implemented.

Figure 24 shows the training time required for each epoch. The homomorphically encrypted model consistently requires more time per epoch compared to the non-encrypted model. This increased training time is a trade-off for the added security provided by encryption.

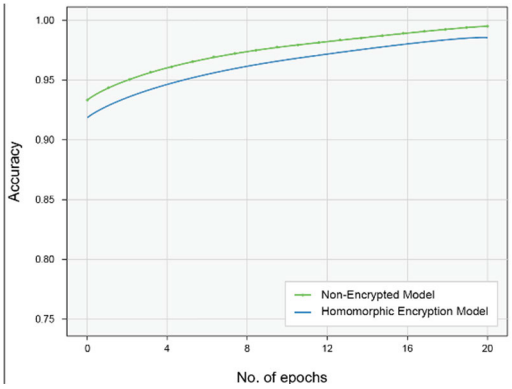

FIGURE 23. Accuracy of Non-Encrypted and HE models over Epochs.

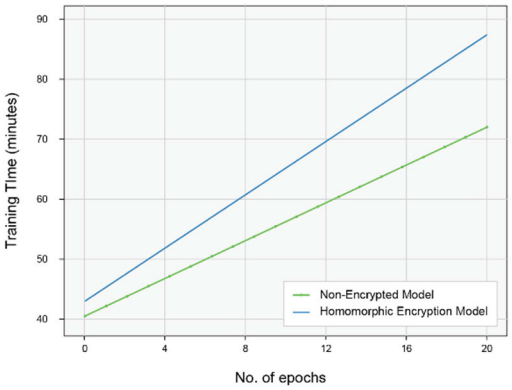

FIGURE 24. Training Time of Non-Encrypted and HE models over Epochs.

Overall, this experiment demonstrates the impact of incorporating HE into our proposed federated learning model, highlighting the balance between enhanced security and computational efficiency. The experiment shows that while encryption adds to the computational cost, it does not substantially hinder the model's overall accuracy, making it a viable option for sensitive applications like breast cancer detection.

VIII. EXPERIMENTS ON COMPUTATIONAL EFFICIENCY

In this section, we conducted an experiment to evaluating the computational efficiency of a federated learning model for breast cancer detection, the energy consumption per node was observed to be 120kWh during the training phase. This value can be attributed to several factors inherent in the federated learning process and the specific demands of training DCNNs, which is inherently a computationally intensive task, requiring substantial energy consumption. This is due to the need for processing large datasets and performing complex mathematical operations, which are central to the training of deep learning models.

In our setup, each node, equipped with a mid-range GPU like the NVIDIA GTX 1660 and a quad-core processor, undergoes these intensive computations. While GPUs significantly enhance the speed of these computations, they also contribute to higher energy usage, especially under the continuous and heavy computational load typical of DCNN training.

The duration of training contributes to the total energy consumption. Although federated learning decentralizes the training process, distributing it across multiple nodes, each node still engages in energy-intensive computations. This distributed nature of federated learning, where each node processes only a part of the dataset, ostensibly offers energy savings compared to a centralized model. However, the aggregate energy consumption across all nodes remains significant.

When considering the application of such a model in real-world medical settings, where resources may be limited, an energy requirement of 120kWh per node is substantial. This highlights the need for optimizing the efficiency of both the AI models and the hardware utilized. In resource-constrained healthcare environments, optimizing for energy efficiency is as crucial as ensuring computational efficiency and accuracy. Balancing these factors is key to the successful deployment of federated learning models in practical healthcare applications.

Thus, while the federated learning approach brings advantages in data privacy and reduced data transfer needs, its energy efficiency is a vital aspect that needs attention. Future improvements in the model's design and hardware optimization could help in reducing the energy footprint, enhancing the feasibility of deploying advanced AI techniques like federated learning in diverse medical settings.

Finally, an essential consideration in deploying federated learning models for breast cancer detection is the latency during both training and inference phases. Real-time or near-real-time diagnostic applications demand quick response times, which can be challenging in federated settings due to the distributed nature of the model training and data aggregation.

To assess the impact of latency, we measured the training and inference processes across a network scaling from 1 to 10 hospital nodes. Figure 25 shows a nonlinear increase in cumulative training time as more nodes participate in the federated learning process. This increase is attributed to the added communication time required to aggregate the distributed models. However, the inference time, critical for real-time diagnosis, remains consistently low across all node configurations.

The experiment reveals that while federated learning introduces latency during model training, it does not significantly affect the inference phase. Therefore, despite the longer training times, the swift inference capabilities of the federated model are well-suited for real-time breast cancer detection applications, allowing for timely and accurate diagnoses.

IX. DISCUSSIONS

A. ANALYSIS OF RESULTS

The results of our research, as outlined above, shed light on the significance of architecture selection, data distribution,

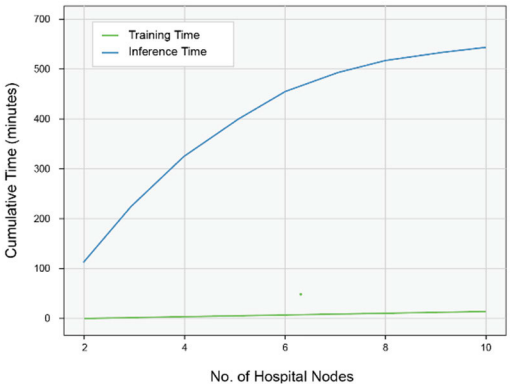

FIGURE 25. Cumulative training and inference latency across ten hospital nodes.

and the federated learning approach in enhancing breast cancer detection based on DCNN. A significant takeaway is the optimization of training through the strategic addition of convolutional layers. The boosted accuracy levels, as depicted in Figure 8, validate our decision to introduce an additional convolutional layer, ultimately enhancing the model's proficiency. The F1-Score's satisfactory performance, particularly for a three-layer model, underscores the balance between precision and recall, which are essential for medical diagnoses where both false negatives and false positives can have dire consequences.

The meticulous assessment of individual datasets, as shown in Figure 8, has provided invaluable insights. Recognizing the constraints of the INBreast dataset due to its limited volume, this individual dataset analysis facilitates a more nuanced understanding of dataset efficiency, allowing us to steer clear of overarching conclusions based solely on aggregate performance metrics.

Another measurement in our study is the ROC curve, illustrated in Figure 9. The favorable ratio of true positives to false positives across all datasets accentuates the efficacy of our chosen discrimination threshold. The commendable performance with the VINDR-MAMMO dataset,in particular, underscores its potential as a robust resource for future breast cancer detection endeavors.

One of the primary objectives of this study was to address the challenges inherent in managing sensitive patient data, especially in the context of multi-institutional collaborations. Our simulation results manifestly demonstrate the merits of a federated learning model over a localized one. By segmenting the research into four distinct experiments, we were able to discern the nuances of data distribution and its ramifications on model performance.

The first experiment reiterates the intuitive assumption that leveraging the entirety of a dataset yields commendable results, as evidenced by Figures 13 and 14. The second experiment brought to light the pitfalls of overfitting, especially when training data is limited. This observation underlines the importance of mindful data allocation, particularly when working with complex architectures. Yet, it was heartening to note that even in the face of overfitting, the model's overall functionality remained uncompromised.

The 3rd and 4th are particularly revealing. They underscore the potential of federated learning as a solution to disparities in data availability across institutions. The near-identical outcomes between the federated architecture and the centralized scheme with complete data availability are indicative of the federated model's prowess. Moreover, the superior outcomes achieved in the 4th experiment, despite some clients having limited data, showcases the efficacy of combining local models within a federated framework.

Finally, we focused on the integration of HE within our federated learning model. The results from this segment were particularly enlightening, revealing that despite the additional computational complexity introduced by HE, the model's accuracy remained impressively high, within the range of 93.5% to 99%. This experiment not only demonstrated the feasibility of applying HE in a practical, federated learning context but also underscored its effectiveness in maintaining data privacy without significantly compromising the model's performance. The successful implementation of HE in our model marks a significant stride in balancing the dual objectives of preserving patient data confidentiality and achieving high diagnostic accuracy in breast cancer detection.

B. POTENTIAL FUTURE IMPLICATIONS

Beyond its immediate impact on breast cancer detection, our federated learning framework holds promising implications and applications across various domains. Some notable applications include:

- Advancing Healthcare and Medical Imaging: Our approach serves as a model for extending federated learning to early detection of other types of cancer (e.g., lung, prostate) and diagnosing a wide range of diseases beyond cancer (e.g., cardiovascular diseases, neurological disorders). This paves the way for more accurate and decentralized healthcare solutions.
- Privacy-Preserving Machine Learning: The principles applied in our framework can inspire data privacy solutions in finance, legal, and education, safeguarding sensitive information while enabling data-driven decision-making.
- Cross-Domain Federated Learning: The scalability and adaptability of our approach extend beyond healthcare, finding relevance in collaborative learning scenarios across industries such as manufacturing, agriculture, and energy. Our method fosters secure, collaborative data analysis while respecting data ownership.
- Global Collaboration: Our federated learning model promotes international collaboration in data-driven research. By complying with diverse data protection regulations, it facilitates global cooperation among institutions, transcending geographical boundaries.

- Scalability and Efficiency: Scalability is a key attribute of our approach, rendering it suitable for handling vast datasets and diverse data sources. Its potential applications span sectors like autonomous vehicles, IoT, and smart cities, where distributed, privacy-preserving AI is paramount.
- Interdisciplinary Research: Bridging the gap between machine learning and healthcare, our work encourages interdisciplinary collaborations. It fosters knowledge exchange between these domains, propelling innovation at the intersection of healthcare and AI.

C. STRATEGIES FOR DATA IMBALANCE IN FEDERATED LEARNING MODELS

In section (V-B) we described how we used Focal Loss feature to minimize the effects of imbalanced data classification, which is a common issue in medical imaging datasets. In this part, we discuss the challenge of data imbalance in the context of federated learning approach. Data imbalance occurs when certain classes (e.g., types of cancerous lesions) are significantly underrepresented compared to others. This imbalance can skew the model's learning process, leading to poor generalization, especially for the underrepresented classes. There are many techniques where federated learning can address this issue:

In an academic research context, addressing data imbalance within federated learning models, particularly in the domain of medical imaging datasets, can be articulated as follows:

- Implementation of Stratified Sampling Techniques: Federated learning systems can employ stratified sampling to ensure each training batch proportionally represents various classes. This method guarantees consistent exposure of the model to all classes, mitigating the risk of bias towards overrepresented categories.
- Development of Tailored Local Models: The federated learning framework facilitates the customization of local models to address specific data distributions. This approach allows for specialized focus on rare categories within individual clients, enriching the global model with a diverse range of local insights.
- Client-Centric Data Augmentation Strategies: Within the federated learning paradigm, data augmentation is conducted at the client level (such as hospitals or medical centers). This involves the generation of synthetic instances of underrepresented categories utilizing techniques like geometric modifications or advanced methodologies such as Generative Adversarial Networks (GANs), thereby artificially enhancing dataset balance.
- Modification of Loss Functions for Enhanced Sensitivity: Altering the loss function to disproportionately penalize misclassifications of minority classes, an approach known as cost-sensitive learning, ensures greater focus on accurately predicting these underrepresented categories.
- Adoption of Weighted Aggregation in Model Updates: The central server in a federated learning network amalgamates updates from multiple clients to refine the global model. By variably weighting these updates based on the uniqueness or scarcity of the data provided by each client, the model can counteract the effects of data imbalance.
- Rigorous Performance Monitoring and Evaluative Metrics: Ongoing monitoring of model performance, with an emphasis on metrics that offer detailed insights into class-specific performance (such as F1-score, precision, recall, and confusion matrices), is crucial. This approach is instrumental in identifying and addressing biases favoring majority classes.

Generally, federated learning offers a structured approach to managing data imbalance in medical imaging datasets, leveraging decentralized data sources and local model adaptations, combined with global synthesis, to develop models that are more representative of varied medical conditions. Nonetheless, the efficacy of these models in addressing data imbalance necessitates more future experiments, specifically on real datasets.

D. LEGAL AND ETHICAL CONSIDERATIONS OF AI-BASED FEDERATED LEARNING IN HEALTHCARE

The integration of artificial intelligence (AI) in healthcare has brought forth transformative opportunities but has also introduced complex ethical challenges that demand careful consideration. In the context of our research on federated learning for breast cancer detection, we recognize the following ethical considerations, with a particular focus on patient consent and data usage:

- Informed Patient Consent: Patient consent is the cornerstone of ethical healthcare AI. It is imperative to ensure that patients are adequately informed about how their data will be used for AI-based diagnostics. Transparent communication and obtaining informed consent from patients for data collection, sharing, and utilization is paramount. Additionally, the consent process should be ongoing, allowing patients to withdraw their data at any point in time.
- Data Privacy and Security: Safeguarding patient data is an ethical imperative. AI algorithms require access to sensitive medical information, and rigorous data security measures must be in place to protect against breaches or misuse. Beyond mere compliance with regulations such as Health Insurance Portability and Accountability Act (HIPAA) in the United States, it is essential to instill a culture of data stewardship and trustworthiness.
- Bias and Fairness: Bias in AI algorithms can lead to disparities in healthcare outcomes. It is essential to address

bias and fairness concerns in algorithm development, validation, and deployment. Ensuring that AI systems do not perpetuate existing healthcare disparities requires vigilance and ongoing monitoring.

- Accountability and Transparency: The responsibility for AI-driven decisions should be clearly defined. Healthcare institutions and AI developers must be accountable for the outcomes of AI systems. Transparency in AI models and decision-making processes is necessary to build trust among patients and healthcare providers.
- Data Ownership and Access: Patients should have control over their health data, including the ability to access, share, or withdraw it as they see fit. Establishing data ownership and access frameworks that empower patients is ethically imperative. Federated learning, as demonstrated in our research, aligns with these principles by preserving data at its source and enabling decentralized control.
- Continuous Evaluation and Improvement: AI models should undergo continuous evaluation for performance, safety, and ethical considerations. This iterative process ensures that the AI system's impact on patient care remains positive and aligned with ethical standards.

With regards to the application of application of federated learning in healthcare, addressing legal and ethical considerations is paramount, especially across diverse geographic regions with varying data protection laws. We list some notable considerations as follows:

In the field of breast cancer detection, addressing legal and ethical considerations is paramount, especially across diverse geographic regions with varying data protection laws such as:

- Cross-Jurisdictional Challenges: Since federated learning can involve collaboration across borders, understanding and complying with international data transfer regulations become critical. The model's development must respect the legal frameworks of all involved regions, particularly when model updates or insights are shared across borders.
- Data Privacy and Regulatory Compliance: Federated learning, by design, keeps patient data localized, which aligns well with privacy concerns and regulations like General Data Protection Regulation (GDPR) and HIPAA. Despite this, the approach must be meticulously aligned with specific local and regional data protection laws. Each region may have distinct requirements regarding data handling, anonymization, and patient consent.
- Governance and Accountability: Clear governance structures are essential to establish accountability for the outcomes of federated learning systems. This involves defining who is responsible for the model's outputs and ensuring robust protocols for auditing and monitoring the system.

TABLE 2. Federated learning vs. traditional DCNNs - a privacy and performance comparison.

Aspect	Federated Learning Approach	Traditional Non-Federated DCNN Models
Data Privacy	Enhanced privacy due to data localization and reduced data movement.	Higher risk of data breaches due to centralized data aggregation.
Data Exposure Risk	Lower risk; only model updates are shared, not raw data.	Higher risk; raw data is often transferred and stored centrally.
Compliance with Data Protection Regulations	Better alignment with GDPR, HIPAA due to minimal data transfer.	Potential challenges in meeting data protection standards due to data centralization.
Model Generalization	Improved through learning from diverse datasets at multiple nodes.	Limited by the homogeneity of centralized datasets.
Scalability and Resource Optimization	High scalability; computational load distributed across nodes.	Scalability limited by central server capacity.
Continuous Learning and Adaptation	Continuously evolves with new data from various sources.	Requires manual updates and retraining with new data.
Privacy-Performance Trade-off	Balance needed between privacy enhancements and coordination complexity.	Direct control over model training and data handling.
Communication Efficiency	Potential challenges with network bandwidth and latency.	Efficient data handling due to centralized model architecture.

- Ethical Considerations: Ethical deployment of federated learning in healthcare necessitates ensuring fairness and avoiding biases in the AI models. This is crucial to prevent systemic biases against certain demographic groups.
- Patient Consent and Autonomy: In healthcare, informed consent is fundamental. Even if federated learning does not entail direct data transfer, patients should be made aware of and consent to how their data is utilized in AI model training. This includes understanding the potential benefits, risks, and the nature of their data's involvement.
- Collaboration with Legal Experts: Regular collaboration with legal advisors and data protection specialists is crucial to navigate the complex landscape of healthcare data regulations and to ensure ongoing compliance as laws evolve.

By addressing these legal and ethical considerations comprehensively, AI and federated learning can be effectively and responsibly integrated into healthcare, particularly in the sensitive field of breast cancer detection.

E. COMPARATIVE ANALYSIS OF FEDERATED LEARNING VS. TRADITIONAL DCNNS

Table 2 provides a comparative analysis between the federated learning approach and traditional non-federated DCNN models, specifically highlighting the differences in privacy and performance aspects. This comparison is essential for understanding the unique benefits and challenges of each approach in the context of breast cancer detection using DCNNs.

Table 2 illustrates the distinct advantages of federated learning in terms of privacy, with its data localization and minimal data transfer, aligning closely with stringent data protection regulations. It also shows how federated learning benefits from continuous learning and improved generalization due to its access to diverse, decentralized data sources. However, it acknowledges the challenges in communication efficiency and the need to balance privacy with performance complexities.

In contrast, traditional DCNN models offer direct control and efficient data handling due to their centralized nature but face significant risks in data privacy and compliance with data protection laws. Their performance is often constrained by the limitations of centralized data, impacting model generalization.

This comparison underscores the suitability of the federated learning approach in scenarios where data privacy is paramount and diverse datasets are essential for accurate model training, such as in healthcare applications for breast cancer detection.

X. CONCLUSION

Breast cancer affects millions of women all around the world. What makes the difference in its treatment is detecting it early since this is crucial for successful recovery. However, early detection is a difficult task to accomplish, even for experienced specialists. The medical field is urgently seeking reliable and accurate breast cancer classification techniques. Artificial intelligence systems, especially DCNNs, seem to be the right solution since they have exhibited great potential for effectively detecting and classifying the disease. These technologies have been explored by researchers for the past couple of years, hoping to enhance the important early detection rates and, eventually, save lives.

This study underscores the potential of federated learning architectures, specifically employing Deep Convolutional Neural Networks (DCNNs), in the realm of breast cancer detection. Our approach has shown a notable enhancement in detection efficiency, achieving a 98.9% accuracy rate across three significant datasets: VINDR-MAMMO, CMMD, and INBREAST, while concurrently upholding stringent standards of privacy and security.

Furthermore, the integration of Homomorphic Encryption (HE) within our federated learning framework represents a critical step in balancing heightened computational demands with the precision of the model. This balance is crucial in sensitive fields like breast cancer detection, where data confidentiality is paramount.

However, it is important to approach these findings with a degree of caution. While federated learning demonstrates immense promise as a tool in breast cancer identification and potentially in broader dermatological applications, its position as the future tool of choice is contingent upon further validation. This involves extensive testing in diverse real-world scenarios and gaining acceptance within the medical community. Hence, while our results are encouraging, they should be viewed as a foundational step towards more comprehensive, validated applications in the future.

Our findings in breast cancer detection using federated learning and DCNNs not only address a critical healthcare challenge but also serve as a pivotal contribution to the broader spectrum of medical imaging and AI applications. This research exemplifies how advanced AI techniques can revolutionize diagnostic accuracy and efficiency across various medical fields, potentially leading to earlier and more precise disease identification and better patient outcomes. Furthermore, the successful application of these AI strategies in breast cancer detection sets a precedent for the adaptation and implementation of similar methodologies in other areas of medical imaging, paving the way for more innovative and effective healthcare solutions.

Future research in this field should concentrate on enhancing the model's ability to generalize, which could be carried out by testing on more extensive and more diverse datasets. In our study, we focused on binary classification of breast cancer images into malignant and benign. Future research should explore multi-classification into more specific types, such as Adenosis, Fibroadenoma, Phyllodes Tumor, and Tubular Adenoma for benign, and Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma, and Papillary Carcinoma for malignant. This approach requires data from diverse datasets for comprehensive accuracy. Additionally, examining the feasibility of incorporating the proposed architecture into current practical workflows may result in the model's enhanced usability in the real world. Exploring the possible interpretation of the features pulled out by the suggested architecture could deliver helpful insights into the mechanisms of breast cancer detection, maybe even leading to the appearance of more effective diagnostic tools. However, it may also face challenges like data variability, computational demands, privacy concerns, model generalizability, integration complexities, and regulatory issues. Each of these factors requires careful consideration for successful implementation in clinical settings. Finally, we aim to explore and develop improvements over the standard FedAvg algorithm. These advancements will focus on enhancing the efficiency and accuracy of federated learning in medical imaging analysis, specifically in the challenging domain of breast cancer detection.

REFERENCES

[1] C. Frick, H. Rumgay, J. Vignat, O. Ginsburg, E. Nolte, F. Bray, and I. Soerjomataram, "Quantitative estimates of preventable and treatable deaths from 36 cancers worldwide: A population-based study," *Lancet Global Health*, vol. 11, no. 11, pp. e1700–e1712, Nov. 2023.

[2] N. Razmjooy, V. V. Estrela, and H. J. Loschi, "Entropy-based breast cancer detection in digital mammograms using world cup optimization algorithm," in *Research Anthology on Medical Informatics in Breast and Cervical Cancer*. Hershey, PA, USA: IGI Global, 2023, pp. 645–665.

[3] P. N. Ramkumar, B. C. Luu, H. S. Haeberle, J. M. Karnuta, B. U. Nwachukwu, and R. J. Williams, "Sports medicine and artificial intelligence: A primer," *Amer. J. Sports Med.*, vol. 50, no. 4, pp. 1166–1174, Mar. 2022.

[4] Z. Hameed, B. Garcia-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," *Sci. Rep.*, vol. 12, no. 1, Sep. 2022, Art. no. 15600.

[5] P. Oza, P. Sharma, S. Patel, and P. Kumar, "Deep convolutional neural networks for computer-aided breast cancer diagnostic: A survey," *Neural Comput. Appl.*, vol. 34, no. 3, pp. 1815–1836, Feb. 2022.

[6] B. Abunasser, M. R. AL-Hiealy, I. Zaqout, and S. Abu-Naser, "Convolution neural network for breast cancer detection and classification using deep learning," *Asian Pacific J. Cancer Prevention*, vol. 24, no. 2, pp. 531–544, Feb. 2023.

[7] L. Akter, Ferdib-Al-Islam, M. M. Islam, M. S. Al-Rakhami, and M. R. Haque, "Prediction of cervical cancer from behavior risk using machine learning techniques," *Social Netw. Comput. Sci.*, vol. 2, no. 3, p. 177, May 2021.

[8] M. Kolhar and S. M. Aldossary, "Privacy-preserving convolutional bi-LSTM network for robust analysis of encrypted time-series medical images," *AI*, vol. 4, no. 3, pp. 706–720, Aug. 2023.

[9] R. Torkzadehmahani, R. Nasirigerdeh, D. B. Blumenthal, T. Kacprowski, M. List, J. Matschinske, J. Spaeth, N. K. Wenke, and J. Baumbach, "Privacy-preserving artificial intelligence techniques in biomedicine," *Methods Inf. Med.*, vol. 61, no. 1, pp. e12–e27, Jun. 2022.

[10] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020.

[11] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," *J. Med. Internet Res.*, vol. 21, no. 7, Jul. 2019, Art. no. e14464.

[12] N. Wu, "Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1184–1194, Apr. 2020.

[13] S. Salman and X. Liu, "Overfitting mechanism and avoidance in deep neural networks," 2019, *arXiv:1901.06566*.

[14] L. Rice, E. Wong, and Z. Kolter, "Overfitting in adversarially robust deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 8093–8104.

[15] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.

[16] W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, and H. J. Aerts, "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA, Cancer J. Clinicians*, vol. 69, no. 2, pp. 127–157, 2019.

[17] R. Walsh and M. Tardy, "A comparison of techniques for class imbalance in deep learning classification of breast cancer," *Diagnostics*, vol. 13, no. 1, p. 67, Dec. 2022.

[18] J. Gründner, *Cross-hospital infrastructure for Research, Statistical Analysis and the Creation and Deployment of Statistical Models Based on Standardized Data of the German Data Integration Centers*. Erlangen, Germany: Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2022.

[19] W. Zhang, T. Zhou, Q. Lu, X. Wang, C. Zhu, H. Sun, Z. Wang, S. K. Lo, and F.-Y. Wang, "Dynamic-fusion-based federated learning for COVID-19 detection," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15884–15891, Nov. 2021.

[20] W. Zhang, Q. Lu, Q. Yu, Z. Li, Y. Liu, S. K. Lo, S. Chen, X. Xu, and L. Zhu, "Blockchain-based federated learning for device failure detection in industrial IoT," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5926–5937, Apr. 2021.

[21] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.

[22] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 2, pp. 254–259, Jan. 2018.

[23] T. Umamaheswari and P. Sumathi, "Enhanced firefly algorithm (EFA) based gene selection and adaptive neuro neutrosophic inference system (ANNIS) prediction model for detection of circulating tumor cells (CTCs) in breast cancer analysis," *Cluster Comput.*, vol. 22, pp. 14035–14047, Nov. 2019.

[24] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," in *Proc. Pacific Symp. Biocomputing*. Singapore: World Scientific, 2017, pp. 1–11.

[25] S. Punitha, A. Amuthan, and K. S. Joseph, "Enhanced monarchy butterfly optimization technique for effective breast cancer diagnosis," *J. Med. Syst.*, vol. 43, no. 7, pp. 1–14, Jul. 2019.

[26] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," *Sustainability*, vol. 15, no. 4, p. 3317, Feb. 2023.

[27] N. N. Basil, S. Ambe, C. Ekhator, and E. Fonkem, "Health records database and inherent security concerns: A review of the literature," *Cureus*, vol. 14, no. 10, Oct. 2022, Art. no. e30168.

[28] Z. Wan, J. W. Hazel, E. W. Clayton, Y. Vorobeychik, M. Kantarcioglu, and B. A. Malin, "Sociotechnical safeguards for genomic data privacy," *Nature Rev. Genet.*, vol. 23, no. 7, pp. 429–445, Jul. 2022.

[29] J. O. du Terrail, "Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer," *Nature Med.*, vol. 29, no. 1, pp. 135–146, 2023.

[30] 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.

[31] L. Li, N. Xie, and S. Yuan, "A federated learning framework for breast cancer histopathological image classification," *Electronics*, vol. 11, no. 22, p. 3767, Nov. 2022.

[32] 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.

[33] R. Wang, J. Xu, Y. Ma, M. Talha, M. S. Al-Rakhami, and A. Ghoneim, "Auxiliary diagnosis of COVID-19 based on 5G-enabled federated learning," *IEEE Netw.*, vol. 35, no. 3, pp. 14–20, May 2021.

[34] A. Archetti, F. Ieva, and M. Matteucci, "Scaling survival analysis in healthcare with federated survival forests: A comparative study on heart failure and breast cancer genomics," *Future Gener. Comput. Syst.*, vol. 149, pp. 343–358, Dec. 2023.

[35] J. Peta and S. Koppu, "Enhancing breast cancer classification in histopathological images through federated learning framework," *IEEE Access*, vol. 11, pp. 61866–61880, 2023.

[36] M. Al-Rakhami, A. Gumaei, M. Alsahli, M. M. Hassan, A. Alamri, A. Guerrieri, and G. Fortino, "A lightweight and cost effective edge intelligence architecture based on containerization technology," *World Wide Web*, vol. 23, no. 2, pp. 1341–1360, Mar. 2020.

[37] B. McMahan, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[38] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 429–450.

[39] S. Khairnar, S. D. Thepade, and S. Gite, "Effect of image binarization thresholds on breast cancer identification in mammography images using OTSU, Niblack, Burnsen, Thepade's SBTC," *Intell. Syst. With Appl.*, vols. 10–11, Jul. 2021, Art. no. 200046.

[40] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, and F. Caselli, "Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 7, pp. 1422–1430, Jul. 2008.

[41] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process.-Syst. Signal, Image, Video Technol.*, vol. 38, no. 1, pp. 35–44, Aug. 2004.

[42] H. Das, "A novel PSO based back propagation learning-MLP (PSO-BP-MLP) for classification," in *Computational Intelligence in Data Mining*, vol. 2. Cham, Switzerland: Springer, Dec. 2014, pp. 20–21.

[43] R. Kavitha, D. K. Jothi, K. Saravanan, M. P. Swain, J. L. A. Gonzáles, R. J. Bhardwaj, and E. Adomako, "Ant colony optimization-enabled CNN deep learning technique for accurate detection of cervical cancer," *BioMed Res. Int.*, vol. 2023, pp. 1–9, Feb. 2023.

[44] M. Sharif, J. Amin, M. Raza, M. Yasmin, and S. C. Satapathy, "An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor," *Pattern Recognit. Lett.*, vol. 129, pp. 150–157, Jan. 2020.

[45] A. Majid, M. A. Khan, M. Yasmin, A. Rehman, A. Yousafzai, and U. Tariq, "Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection," *Microsc. Res. Technique*, vol. 83, no. 5, pp. 562–576, May 2020.

[46] E. O. Abiodun, A. Alabdulatif, O. I. Abiodun, M. Alawida, A. Alabdulatif, and R. S. Alkhawaldeh, "A systematic review of emerging feature selection optimization methods for optimal text classification: The present state and prospective opportunities," *Neural Comput. Appl.*, vol. 33, no. 22, pp. 15091–15118, Nov. 2021.

[47] H. Kaur, D. Koundal, and V. Kadyan, "Image fusion techniques: A survey," *Arch. Comput. Methods Eng.*, vol. 28, pp. 4425–4447, Dec. 2021.

[48] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 64, pp. 71–91, Dec. 2020.

[49] N. Hong, "Cost-effectiveness analysis of whole-mount pathology processing for patients with early breast cancer undergoing breast conservation," *Current Oncol.*, vol. 23, no. s1, pp. 23–31, 2016.

[50] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

[51] M. S. Al-Rakhami, A. Gumaei, M. M. Hassan, A. Alamri, M. Alhussein, M. A. Razzaque, and G. Fortino, "A deep learning-based edge-fog-cloud framework for driving behavior management," *Comput. Electr. Eng.*, vol. 96, Dec. 2021, Art. no. 107573.

[52] M. S. Al-Rakhami, A. Gumaei, M. Altaf, M. M. Hassan, B. F. Alkhamees, K. Muhammad, and G. Fortino, "FallDeF5: A fall detection framework using 5G-based deep gated recurrent unit networks," *IEEE Access*, vol. 9, pp. 94299–94308, 2021.

[53] V. Verdhan, "Image classification using LeNet," in *Computer Vision Using Deep Learning: Neural Network Architectures With Python and Keras*. Berkeley, CA, USA: Apress, 2021, pp. 67–101.

[54] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, Jul. 2016.

[55] M. Pandurangan, U. Durga, and C. Ponnuraja, "An integrative machine learning framework for classifying SEER breast cancer," *Sci. Rep.*, vol. 13, no. 1, p. 5362, 2023.

[56] Z. Zhang, "Improved Adam optimizer for deep neural networks," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–2.

[57] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, "A sufficient condition for convergences of Adam and RMSProp," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11119–11127.

[58] S. Bonnabel, "Stochastic gradient descent on Riemannian manifolds," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2217–2229, Sep. 2013.

[59] A.-A. Nahid, M. A. Mehrabi, and Y. Kong, "Histopathological breast cancer image classification by deep neural network techniques guided by local clustering," *BioMed Res. Int.*, vol. 2018, pp. 1–20, Mar. 2018.

[60] C. Laurent, G. Pereyra, P. Brakel, Y. Zhang, and Y. Bengio, "Batch normalized recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2657–2661.

[61] S. Yi, Z. Fan, and D. Wu, "Batch feature standardization network with triplet loss for weakly-supervised video anomaly detection," *Image Vis. Comput.*, vol. 120, Apr. 2022, Art. no. 104397.

[62] I. Rodriguez-Martinez, J. Lafuente, R. H. N. Santiago, G. P. Dimuro, F. Herrera, and H. Bustince, "Replacing pooling functions in convolutional neural networks by linear combinations of increasing functions," *Neural Netw.*, vol. 152, pp. 380–393, Aug. 2022.

[63] S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, Nov. 2016, pp. 189–204.

[64] W. Sun, T.-L. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Computerized Med. Imag. Graph.*, vol. 57, pp. 4–9, Apr. 2017.

[65] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022.

[66] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[67] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020.

[68] H. T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao, and V. Vu, "VinDr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography," *Sci. Data*, vol. 10, no. 1, p. 277, May 2023.

[69] H. Cai et al., "An online mammography database with biopsy confirmed types," *Sci. Data*, vol. 10, no. 123, pp. 1–6, 2023, doi: 10.1038/s41597-023-02025-1.

[70] 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, no. 2, pp. 236–248, 2012.

[71] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[72] C. Gentry, *A Fully Homomorphic Encryption Scheme*. Stanford, CA, USA: Stanford Univ., 2009.

[73] A. Acar, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–35, 2018.

HUSSAIN ALSALMAN received the B.Sc. and M.Sc. degrees in computer science from King Saud University (KSU), Riyadh, Saudi Arabia, and the Ph.D. degree in artificial intelligence, U.K. From 2009 to 2014, he chaired the Department of Computer Science, College of Computer and Information Sciences, KSU. He worked for several years as a Consultant for several companies in the private sector and institutes in the government sector in Saudi Arabia. He is currently a Staff Member with the Department of Computer Science, KSU. His main research interests include medical image processing, machine learning algorithms, neural networks, computational methods for health care monitoring, and ensemble and deep learning models for medical analysis and diagnosis. He was a member of the Review Board of the *Journal of King Saud University Computer and Information Sciences*, from 2004 to 2014.

MABROOK S. AL-RAKHAMI (Member, IEEE) received the Ph.D. degree in information systems from King Saud University (KSU), Riyadh, Saudi Arabia. He is currently an Assistant Professor with the Department of Information Systems, College of Computer and Information Sciences (CCIS), KSU. He has authored and coauthored more than 120 publications, including refereed journals and conferences. He authored two books and translated two books. His research interests include cloud/edge computing, the Internet of Things, artificial intelligence, sensor networks, big data, mobile computing, cyber security, smart computing, 5G/6G networks, and social networks. He was a recipient of a number of awards and listed as one of the top 2% Scientists of the world in the AI field.

TAHA ALFAKIH received the B.S. degree in computer science from the Department of Computer Science, Hadhramout University, Yemen, the M.Sc. degree from the Department of Computer Science, King Saud University (KSU), Riyadh, Saudi Arabia, and the Ph.D. degree from the Department of Information System, KSU. He is currently a Researcher with the College of Computer Science, KSU. His research interests include machine learning, mobile edge computing, and the Internet of Things (IoT).

MOHAMMAD MEHEDI HASSAN (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in February 2011. He is currently a Full Professor with the Department of Information Systems, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. He has authored or coauthored more than 180 publications, including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include edge/cloud computing, the Internet of Things, cyber security, deep learning, artificial intelligence, body sensor networks, 5G networks, and social networks. He was a recipient of a number of awards, including the Best Journal Paper Award from IEEE Systems Journal, in 2018, the Best Paper Award from CloudComp Conference, in 2014, and the Excellence in Research Award from KSU (two times in a row, 2015 and 2016).