AUT Journal of Electrical Engineering
10.22060/eej.2024.23377.5609

A Comprehensive Framework for Multi-Class Breast Lesion Classification Using a Gaussian Process-Optimized VGG16-Capsule Hybrid Model on Ultrasound Radio Frequency B-mode Images

Mahsa Arab[1], Ali Fallah[2*], Saeid Rashidi[3], Maryam Mehdizadeh Dastjerdi[4], Nasrin Ahmadinejad[5]

[1]Ph.D. Candidate, Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran
[2]*Ph.D., Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran, afallah@aut.ac.ir
[3]Ph.D., Medical Sciences and Technologies Department, Islamic Azad University, Science and Research Branch, Tehran, Iran
[4]Ph.D., Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran
[5]M.D., Radiology-Medical Imaging Center, Cancer Research Institute, Imam Khomeini Hospital Advanced Diagnostic and Interventional Radiology Research Center, Tehran University of Medical Sciences, Tehran, Iran

Abstract:

Effective breast cancer screening is essential for early detection and treatment. Ultrasound (US) radio frequency (RF) data offers a novel, equipment-independent approach. However, class imbalance and limited interpretability hinder its application in clinical practice. This study proposes a hybrid deep learning model combining a pre-trained convolutional neural network (CNN) based on VGG16 and capsule neural networks (CapsNets) to classify breast lesions. The model was evaluated using an RFTSBU dataset, comprising 220 data points from 118 patients, acquired on the SuperSonic Imagine Aixplorer® system with a linear transducer. To address data imbalance, the synthetic minority over-sampling technique (SMOTE) was employed to generate synthetic samples while preserving data distribution. Furthermore, Gaussian process (GP) was applied to fine-tune CapsNet hyperparameters, improving classification performance. Three experiments were conducted to classify breast lesions into two, three, and four classes: (I) CapsNet with balanced datasets based on class weight, (II) CapsNet with balanced datasets using SMOTE, and (III) CapsNet with hyperparameters optimized using GP on SMOTE-balanced datasets. The proposed model achieved average accuracies of 98.81%, 97.89%, and 95.94% for two-, three-, and four-class classifications, respectively. The hybrid VGG16-CapsNet model effectively addresses class imbalance and captures critical

lesion attributes such as size, perspective, and orientation. Integrating GP optimization achieves superior accuracy in multi-class breast lesion classification. The proposed approach can serve as a valuable aid in breast tumor classification using US RF B-mode images. Its enhanced interpretability and efficiency enable clinicians to move beyond binary classification, facilitating identifying and differentiating a broader spectrum of breast lesions.

Keywords:

ultrasound, radio frequency, breast cancer classification, capsule neural networks, Gaussian process

1. Introduction

Breast cancer is a significant health concern worldwide, with its prevalence particularly notable among Iranian women, where it constitutes a substantial portion of cancer cases [1]. Globally, it is projected that by 2040, the burden of breast cancer will exceed 3 million new cases and 1 million deaths annually [2]. At the national level in Iran, breast cancer was the most prevalent cancer among women in 2020, accounting for 28.1% of all reported cases in the country [3].

Mammography screening, although common, often necessitates additional procedures for women with abnormal results [4, 5]. The breast imaging-reporting and data system (BI-RADS) method is a valuable tool in decision-making, helping to determine the need for further sampling. Minimizing unnecessary procedures is essential for patient well-being, healthcare access, and reducing screening costs [6-9].

Breast sonography presents advantages over mammography, being radiation-free [10, 11] and more convenient for routine screenings, making it a cost-effective option for various regions [12]. It exhibits higher sensitivity, particularly in dense breast tissue, proving valuable for women under 35 [10, 13]. Breast sonography also reduces false positives compared to mammography, minimizing unnecessary biopsies [14]. Breast sonography demonstrates high accuracy, particularly in detecting simple cysts, with success rates ranging from 96% to 100% [15]. However, its effectiveness relies on the operator's skills and

subjective interpretation by experienced radiologists [11]. Consequently, this paper propounded a pressing need to enhance the effectiveness of current diagnostic approaches.

The rest of the paper is organized as follows: Related works are described in Section 2. Preliminary and fundamental information is covered in Section 3, along with details on the data collection, capsule neural networks (CapsNets), suggested method, and process for classifying breast lesions. In Section 4, the performance of the proposed approach is analyzed, and the experimental results are detailed. Section 5 addresses the techniques, corresponding outcomes, limitations, and directions for future works. Finally, Section 6 concludes the paper.

2. Related Works

Jarosik *et al*. conducted a study using convolutional neural networks (CNNs) for classifying radio frequency (RF) signals from the OASBUD dataset [1]. The research explored three approaches—CNN-1D, CNN-2D, and CNN-1D-2D—focusing on processing small 2D patches of RF signals and their amplitude samples. While this method allows the creation of parametric maps to assess breast mass malignancy, it is limited to classifying benign and malignant lesions only. The final CNN-1D-2D model achieved a modest 70% accuracy in two-class classification and required significant pre-processing and network training time.

In 2020, faster region-based CNN (Faster R-CNN) models were implemented to detect and classify breast lesions in ultrasound (US) images, marking a significant step forward in deep learning-based diagnostics [16]. The algorithm successfully identifies lesions by generating bounding boxes and classifying them as malignant or benign, offering an alternative to traditional methods. It is optimized for CPU and GPU platforms, improving efficiency and reducing training and testing times. However, the study lacks detailed reporting of the network's evaluation parameters, only stating an accuracy rate above 95%, and no validation was conducted due to the limited dataset.

In 2021, Behboodi *et al*. investigated deep learning for breast cancer classification using US images, focusing on multi-tasking learning (MTL) techniques [17]. Due to the limited training data and significant

inter-class variations, training the network from scratch was impractical. Therefore, they used two pre-trained networks, ResNet-34 and MobileNet-v2, as backbone feature extractors with task-specific modifications. The method's key advantage was classifying breast lesions into four categories. However, the accuracy for the two-class classification did not surpass 84%, and for the four-class classification, it stayed at 90%. The study also lacked justification for including the background as a class and did not discuss its impact on the results.

Qiao *et al.* developed a deep learning framework for detecting calcifications using multi-channel US RF signals, offering a promising approach for enhancing breast cancer diagnosis [18]. The study introduced a unique integrated framework, where multi-channel RF signals were merged through beamforming and then transformed using short-time Fourier transform (STFT) to extract frequency domain features. They proposed the RF signal spectrogram-calcification-detection-net (SCD-Net), a CNN based on the YOLOv3 model, enhanced with convolutional short-term memory (ConvLSTM) for detecting calcifications from spectrograms. While the framework showed promise in accurately detecting tumor calcifications and assisting radiologists, it involved complex preprocessing and detection processes and lacked straightforward interpretability. Additionally, the study's focus was limited to calcification detection, suggesting that expanding to include mass screening and tumor classification could further benefit radiologists.

Kim *et al.* undertook a comprehensive approach in their study, developing an end-to-end CNN framework to analyze breast US images using multiparametric images derived from RF signals. This thorough research approach provides reassurance about the validity and reliability of the study's findings [19]. The study utilized entropy and phase images, which offer fine structural and anatomical details, along with conventional B-mode images in the temporal domain. Additionally, attenuation images estimated from the frequency domain of RF signals were used to capture spectral features. The proposed ensemble architecture integrates these diverse parametric images from both time and spectral domains, enhancing the representation of tissue echogenicity and reducing the risk of overfitting by creating a richer dataset for

deep neural network training. The architecture's strength lies in combining various image types, significantly improving classification performance compared to single parametric or B-mode images. However, the study's effectiveness is limited by its relatively small dataset. Allowing voting algorithm weights during training could improve classification accuracy if a larger dataset were available.

In 2022, Byra *et al*. used US RF data in deep learning methods, achieving a higher area under the curve (AUC) than B-mode imaging for breast lesion classification and segmentation [20]. Their approach distinguishes between benign and malignant lesions and segments them, using RF data from the mass and surrounding tissue for a more comprehensive analysis. However, limitations include a small training dataset and less robustness compared to quantitative ultrasound (QUS) techniques. Additionally, their algorithm underperformed in classification, leading to the development of an improved Naive Bayes-based algorithm, which enhanced classification performance [21]. Similarly, Gare *et al*. combined RF data with B-mode images in CNNs, achieving a higher AUC than B-mode alone, highlighting the benefits of incorporating RF data in breast lesion classification [22].

In 2022, researchers explored deep learning methods using QUS multiparameter maps to predict breast treatment response, achieving an 88% accuracy rate [23]. The approach involved extracting intrinsic features from raw US RF data and integrating them with QUS image features in a CNN-based structure. This method effectively classifies individuals as responders or non-responders to therapeutic interventions by capturing spatial heterogeneity across multiple QUS image channels. However, the model faced overfitting issues due to small datasets, leading the researchers to use early stopping during training to improve generalizability. Table 1 provides a concise overview of relevant previous studies.

Table 1. Method comparison of related breast lesion classification experiments

Study	Database	Proposed Model	Data Preprocessing	Evaluation Metrics	Results	Advantages	Disadvantages
Jarosik *et al*. [1]	OASBUD (78 subjects and 100 US RF Data) 2 Class	CNN-1D, CNN-2D, and CNN-1D-2D	1. Extract 2D RF data patches. 2. Develop 1D convolutional layers to process	Accuracy AUC	Best results: Accuracy=70%, AUC=77%	1. Creating parametric maps to assess potential malignancy in breast masses.	1. Limited differentiation between benign and malignant lesions.

			RF signals for CNN-1D. 3. Train CNN-1D based on the envelope of RF signals. 4. Combine CNN-1D with CNN-2D to create the CNN-1D-2D model.			2. Providing an expressive approach for analyzing breast masses.	2. Low accuracy (Only 70% by CNN-1D-2D model) despite using three CNN-based approaches and RF 2D patches. 3. Requiring considerable time for both pre-processing and network training.
Wei *et al.* [16]	OASBUD (78 subjects and 100 US RF Data) 2 Class	Faster R-CNN	-	Accuracy	Accuracy>95%	1. Identifying breast lesions and the region of interest (ROI). 2. Compatible with both CPU and GPU platforms. 3. Reducing the time required for training and testing.	1. Evaluation lacks detailed metrics (Offering only a general accuracy statement above 95%). 2. Lack of complete validation due to limited dataset (Hindering comprehensive assessment of the method's effectiveness).
Behboodi *et al*. [17]	Publicly available US dataset (40 invasive ductal carcinomas (IDC), 65 cysts (CYST), and 39 fibroadenomas (FA)) 4 class	MTL (ResNet34-MobileNet-V2)	1. Crop images and resize them. 2. Normalize image. 3 Augment data (Apply random on-the-fly).	Accuracy, Precision, Recall, F1-Score	For 2-class: Accuracy=84%, Precision=100%, Recall=25%, F1-Score=40% For 4-class: Accuracy=90%, Precision=80%, Recall=50%, F1-Score=62%	1. Classifying breast lesions into four classes. 2. Utilizing background as an additional class.	1. Lack of justification for including the background as a class. 2. Lack of discussion on the impact of including the background class.
Qiao *et al*. [18]	US RF signals of 337 breast tumors with calcifications 2 class	Deep learning architecture based on the YOLOv3 model and integrated features through ConvLSTM	1. Apply beamforming. 2. Utilize STFT to extract frequency domain features.	Accuracy, Precision, Recall, F1-Score, Precision-Recall (PR)	Accuracy=84.13%, Precision=88.47%, Recall=94.5%, F1-Score=91.38%, Precision-Recall (PR)=88.25%	1. Performing strongly in accurately diagnosing tumor calcification.	1. Complex framework steps. 2. Less clear network interpretability. 3. Focusing only on calcification detection.
Kim *et al.* [19]	OASBUD (78 subjects and 100 US RF Data) 2 Class	End-to-end ensemble CNN	1. Extract the entropy and phase images. 2. Construct the B-mode images. 3. Extract the attenuation image.	Accuracy, Precision, Recall, F1-Score, AUC	Accuracy=83%, Precision=79.57%, Recall=91.33%, F1-Score=85.05%, AUC=91.61%	1. End-to-end ensemble architecture for 1D time series classification of RF signals. 2. Utilizing multiple parametric images from time and spectral domains.	1. Relatively small dataset. 2. Enhancing classification with voting algorithm weights.
Byra *et al*. [20]	273 breast masses 2 Class	Y-Net architecture	1. Resize image. 2. Segment manually.	Accuracy, Sensitivity, Specificity, AUC	Accuracy=86.50%, Sensitivity=82%, Specificity=90.60%, AUC=87.40%	1. Classifying benign/malignant lesions and performing segmentation for detailed analysis. 2. Utilizing RF data characteristics from breast mass and surrounding regions for informed decisions.	1. Requiring significant time. 2. Hindering effectiveness due to a small number of data samples. 3. Less robust than traditional QUS techniques.

Taleghamar *et al.* [23]	181 patients 2 class	Deep convolutional neural network (DCNN) and residual attention network (RAN)	1. Resize images. 2. Normalize images. 3. Augment data (Flipping horizontally and shifting horizontally and vertically).	Accuracy, Sensitivity, Specificity, AUC	Accuracy=88%, Sensitivity=70%, Specificity=92.50%, AUC=86%	1. Extracting essential information by analyzing the entire image during feature extraction. 2. Quantifying spatial heterogeneity across multiple QUS image channels for enhanced analysis. 3. Including tumor core and periphery improves accuracy in predicting treatment responses.	1. Network structure overfitting with small datasets, compromising generalizability. 2. Using early stopping to improve generalizability and reduce overfitting.

Unlike previous studies [1, 16, 18-20, 23] that primarily focus on benign and malignant lesions, this research expands its scope to include a broader range of lesions. The classification process in this study is conducted across four distinct classes, with the primary objective of developing a method that not only achieves high accuracy in two-, three-, and four-class classifications with reduced complexity-unlike [1, 20] but also serves as a supportive system for radiologists in diagnosing breast lesions. Unlike [1, 19, 20, 23], the proposed approach effectively avoids overfitting due to the carefully selected strategies and appropriate dataset size for this method. Evaluating data across three proposed scenarios allows for a thorough assessment of the technique, with the best approach for each classification mode being identified and presented.

One of the research's fundamental strengths and innovations is using the Gaussian process (GP) optimization, which enables the optimal performance of the VGG16 and CapsNet networks through fine-tuned hyperparameters. This is noteworthy because the need to optimize hyperparameters is not addressed in [16, 17, 20], with only [19] highlighting this concern. Additionally, this study goes beyond [16, 17] by providing a comprehensive set of evaluation criteria for each approach, demonstrating the superiority of the proposed method. The comparison between the VGG16 and CapsNet methods and the proposed approach is also detailed, with the best scenario for each classification mode identified and discussed. In summary, the main contributions of this paper are as follows:

- The proposed approach extends its performance beyond benign and malignant lesions, offering a more comprehensive classification capability, by utilizing a more comprehensive range of lesions.
- The GP was employed to optimize the parameters, leading to an improvement in classification accuracy.
- The proposed model has been empirically demonstrated to outperform other classification methods across all evaluation metrics. Its high efficiency makes it particularly applicable in real-world scenarios, especially for analyzing breast US RF B-mode images. This capability holds significant potential for various medical applications, particularly in the clinical classification of breast tumors.
- Unlike [18], the proposed method is more understandable and interpretable for radiologists because it utilizes US RF B-mode images, making it easier to comprehend than studies that employed US RF signals. This enhanced clarity contributes to the method's accessibility and usability in clinical practice.

In this study, the motivation behind developing the VGG16-CapsNet model for breast cancer classification is to achieve a more efficient and accurate diagnostic tool by leveraging the strengths of both CNNs and CapsNets. Despite its power in image classification, CNN struggles with capturing spatial relationships and feature orientations, leading to potential misclassifications, particularly with rotated or differently angled objects [24-28]. It also requires large labeled datasets and is prone to overfitting on smaller ones [28-30].

CapsNets address these limitations by preserving spatial hierarchies and orientations, improving object recognition regardless of pose, and reducing overfitting [24-26]. This study developed the VGG16-CapsNet model for breast cancer classification, combining the strengths of CNNs and CapsNets. While CNNs often rely on handcrafted features, CapsNets employ end-to-end training, extracting more intricate features. Integrating VGG16 with CapsNet enhances feature recognition and model accuracy, offering a more efficient and reliable diagnostic tool for breast cancer classification.

3. Theoretical Modeling

10.22060/eej.2024.23377.5609

This study concentrates on the non-invasive classification of breast lesions into benign, probably benign, suspicious, and malignant classes using VGG16-CapsNet and three scenarios. The classification process is demonstrated in Fig. 1.

3-1- Data Collection

This study utilizes the RF time series breast US (RFTSBU) dataset, comprising 220 RF signal frames and corresponding B-mode images, collected at Dezashib Imaging Center in Tehran, Iran, using a SuperLinearTM SL18-5 linear transducer (18.5 MHz) and the US SuperSonic Imagine Aixplorer® medical and research system.

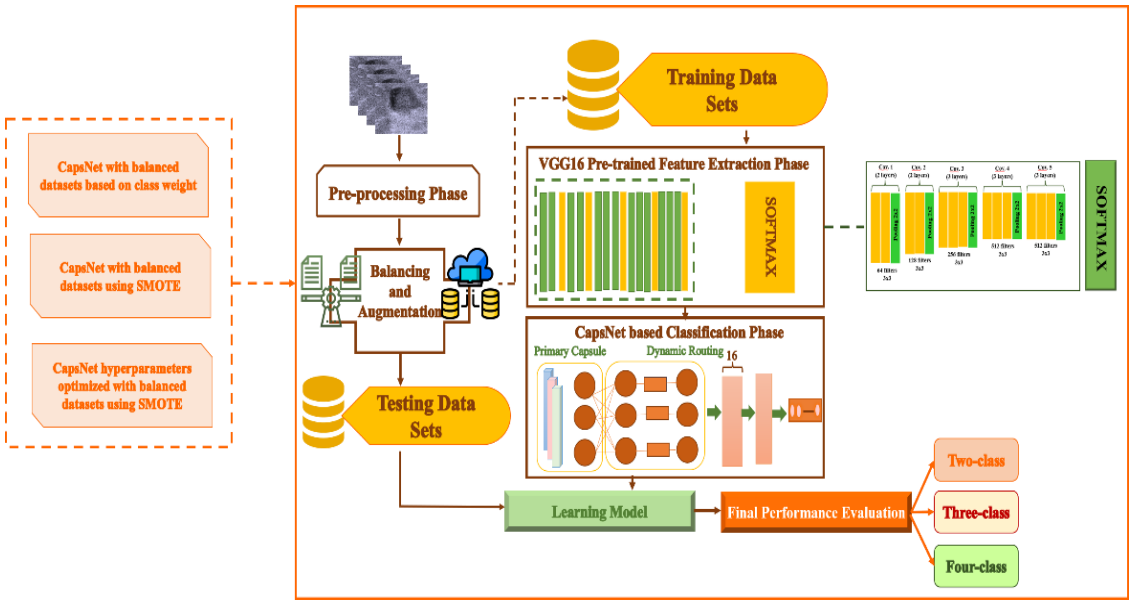

Fig. 1. The proposed VGG16-CapsNet-based breast cancer classification process

With participant consent, ethical approval was obtained from the committee of ethics of the Islamic Azad University, Science and Research Branch, Tehran, Iran. Mammography and biopsies were conducted to confirm malignancy and assess suspicious lesions, with additional data collected for some cases. Patient information, including family history and reproductive details, was recorded. The overall data collection framework and the number of registered data in each group are illustrated in Fig. 2.

3-2- Data Preprocessing: RF Data Parser

10.22060/eej.2024.23377.5609

The RF signals captured by the Supersonic Imagine are saved in a specific format, which requires conversion to the .MAT file for readability in Matlab. A graphical user interface (GUI) named "RF Data Parser" was employed to facilitate this conversion process. This GUI is responsible for converting each recorded RF data into a .MAT file. Additionally, it can generate a B-mode image from each recorded US RF signal and present the user with both the B-mode image and the RF signal. This article utilizes B-mode images derived from US RF data.

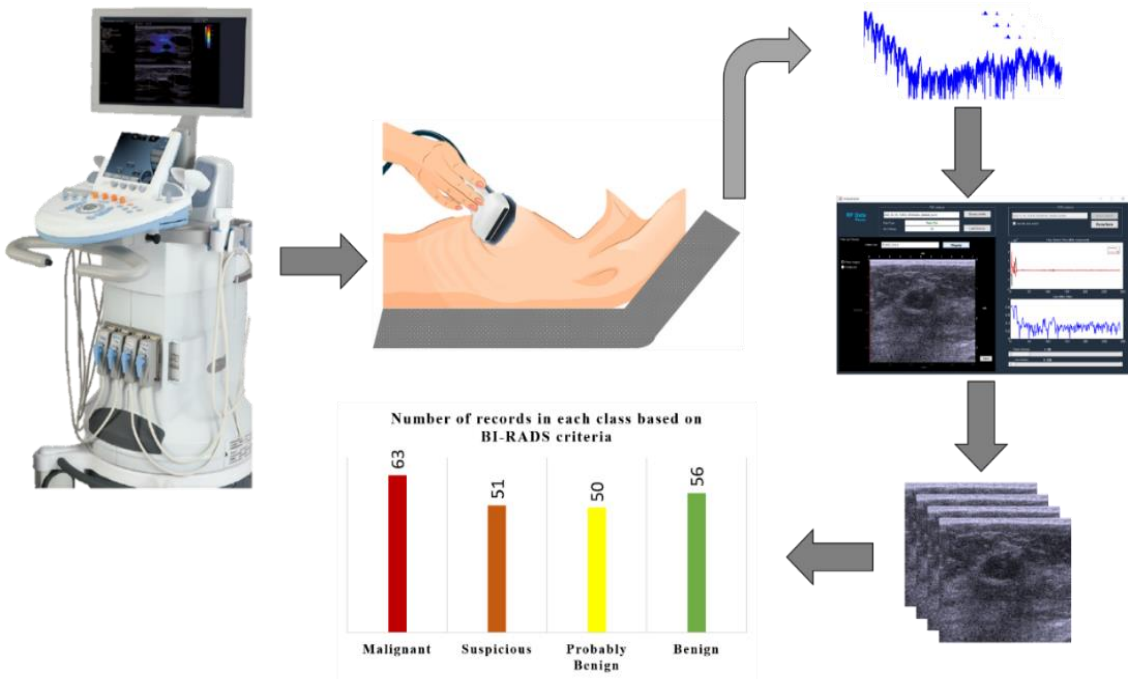

Fig. 2. The data collection framework and the number of records in each class

3-3- Proposed GP Optimized CapsNet based on VGG16

A GP is completely defined by its mean function, $m(x)=E[f(x)]$, and co-covariance, $k(x, x')=E[(f(x)-m(x))(f(x')-m(x'))]$. So, write the GP as [31]:

$$f(x) \sim GP(m(x), k(x, x')) \tag{1}$$

10.22060/eej.2024.23377.5609

The Bayesian model represents a primary instantiation of a GP, the foundation for Bayesian optimization, a widely used approach in general optimization. This iterative method operates by sequentially refining its search for optimal solutions. A probabilistic surrogate model and an acquisition function that directs the choice of the subsequent evaluation point are components of the GP framework. During each iteration, the surrogate model is updated based on all previous evaluations of the target function. The acquisition function then makes use of the predictive distribution derived from the probabilistic model to evaluate the relative usefulness of different candidate points, striking a balance between exploration (i.e., looking for less-explored areas) and exploitation (i.e., concentrating on areas that are most likely to produce high performance). The acquisition function can be thoroughly optimized because it is computationally inexpensive, in contrast to the costly evaluation of the black box function [31, 32]. The acquisition function is the expected improvement (EI), which is derived from Eq. (2) [33].

$$E[I(\lambda)] = E[\max(f_{\min} - Y, 0)] \tag{2}$$

Eq. (3) defines the prediction of the model Y at configuration λ according to a normal distribution, and EI is computed in the closed form.

$$E[I(\lambda)] = (f_{\min} - \mu(\lambda))\phi\left(\frac{f_{\min} - \mu(\lambda)}{\sigma}\right) + \sigma\varphi\left(\frac{f_{\min} - \mu(\lambda)}{\sigma}\right)) \tag{3}$$

where f_{min} is the best-observed value, $\varphi(\cdot)$ is the standard normal density, and $\Phi(\cdot)$ is the standard normal distribution function.

3-4- Breast Cancer Classification Procedure

The proposed breast cancer classification model comprises four main phases: data preparation, pre-training, classification and optimization, and evaluation. The ensuing subsections describe each step in detail.

3-4-1- Data Preparation Step

In order to properly model the training data, a network must learn a high-variance function, which leads to overfitting. Three essential steps are involved in this phase: balancing the dataset, data augmentation, and dividing the data into training and testing sets. Considering its novelty, we found two main problems with the dataset: the small number of US RF B-mode data in each class and the imbalance in the data. There aren't many images in the medical dataset, but deep learning methods need a lot of samples to prevent overfitting. This study used image data augmentation techniques, such as rotating images by 10 degrees to the left and right.

To address the imbalance in the dataset, we employed two approaches. The first approach is based on class weights, assigning higher costs to misclassifications of the minority class. For manual calculation, the weight for each class can be set as:

$$w_i = \frac{(Total\_Number\_of\_Samples)}{(Number\_of\_Classes)\times(Number\_of\_Samples\_in\_Class\_i)} \quad (4)$$

The second approach utilizes the synthetic minority over-sampling technique (SMOTE), which includes generating synthetic instances of the minority class to balance the dataset. The SMOTE balances the class distribution by introducing synthetic samples to the minority class, which improves the dataset's suitability for machine learning model training [34, 35]. Additionally, we split the RFTSBU into training, testing, and validation sets, with a 70%, 15%, and 15% ratio, to facilitate the proper evaluation of the model.

3-4-2- Pre-training and CapsNet Optimizing Steps

The three steps in the training process are as follows:

Step (1): VGG16 CNN Structure: An essential aspect of CNNs is weight sharing, where similar feature detectors are applied across the entire object. To handle variations in the dataset, CNNs incorporate subsampling layers that address the significance of the precise location of features [36]. Deep feature extraction is further enhanced through transfer learning with pre-trained models like VGG16, which

10.22060/eej.2024.23377.5609

captures activation values from various layers as features [37]. VGG16, with its five-block structure, facilitates immediate feature extraction and is integrated into the CapsNet architecture [38].

Step (2): CapsNet (Breast-caps): In the CapsNet primary capsule (PC) layer, each capsule takes in a tiny subset of the receptive field as input and tries to figure out the pose of a given pattern [39]. A capsule's output is a vector dynamically routed to the layer below the relevant parent capsule. This output vector is used to predict the output of capsules in the next layer (parent capsules) through a learned transformation matrix. The dynamic routing mechanism then calculates the agreement between these predictions and the actual outputs of the parent capsules. Capsules with higher agreement have increased routing weights, ensuring their contributions are emphasized. This process dynamically routes the output vectors to the most relevant parent capsules, enabling accurate information flow through the network. Fig. 3 demonstrates the CapsNet architecture.

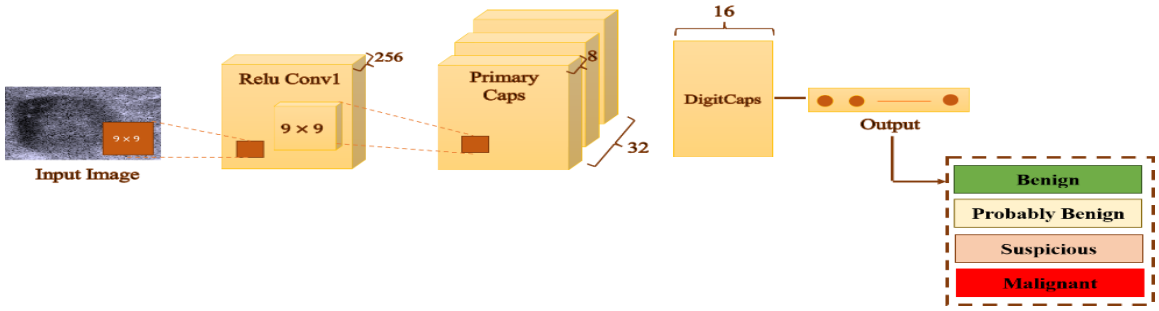

Fig. 3. The CapsNet architecture

Step (3): CapsNet Hyperparameter Optimization: Finding a D-dimensional hyperparameter setting *x* that minimizes the CapsNet's validation loss/error *f* is the goal of hyperparameter optimization. The function *f* maps the validation error of a CapsNet algorithm with learned parameters to a hyperparameter choice *x* of *G* configurable hyperparameters. Eq. (5), which illustrates optimizing *f*, looks for a way to find the ideal hyperparameters automatically:

$$\begin{aligned} &\min_{x \in R^G} f\ (x, \theta; S_{val}) \\ &s.t. \theta = \arg\min_{\theta} f\ (\theta; S_{train}) \end{aligned} \tag{5}$$

10.22060/eej.2024.23377.5609

The enormous complexity of the function f makes it challenging to solve the problem in Eq. (5). In this case, S_{val} stands for the validation dataset and S_{train} for the training dataset. The value of x falls inside a bounded set, and the learning procedure lowers the training loss/error. Among the algorithms for Bayesian optimization is the GP. Bayesian optimization algorithms approximate the costly error function using a low-cost probabilistic surrogate model. As a result, we utilize the GP to optimize the CapsNet's hyperparameters. The specific steps of the breast CapsNet classification model are displayed in Table 2.

Table 2. The breast CapsNet classification model

Input data:
Breast lesions US RF B-mode images (X, Y); where Y={y|y∈{Benign, Probably Benign, Suspicious, and Malignant}}
Output data:
The CapsNet model that classified Breast lesions US RF B-mode images x ∈ X
Begin: // **Pre-processing steps:** {Resize US RF B-mode images to 128×128 dimensions Generate US RF B-mode images using data preparation operations Balance the RFTSBU dataset using the SMOTE method Split into train, valid, and test dataset} // **Deep feature extraction step:** {Utilize pre-trained transfer model VGG16 Extract features using VGG16} // **CapsNet hyperparameter** optimization θ (routing No, capsule No) {Initialize the search space routing No [1, 5]], and capsule No [5, 20]—with random values **while** j∈i+1 …., N Perform search space exploration. Update posterior distribution by incorporating prior information. Select the following sample, the θ minimizing error/loss Train CapsNet (θ) Evaluate y_j=f (x_j, θ)) //objective function Update the GP model of f(x) to refine posterior estimation. **End**
Return optimized hyperparameters
Retrain the CapsNet (θ)

Test the CapsNet
Output the classification

4. Results

The experiments used tensor flow and Keras with a TPU Google COLAB environment. To assess the effectiveness of the proposed approach, its performance was evaluated using several metrics, including accuracy, recall, precision, F-score, and AUC.

Experiment Scenario (I): In this approach, class weights were assigned as follows to address class imbalance: for the four-class scenario (benign, probably benign, suspicious, and malignant), the weights were 0.9545, 1.0758, 1.0892, 1.05, and 0.9242, respectively. The weights in the three-class scenario (benign, suspicious, and malignant) were 1.0697, 1.1358, and 0.8440, respectively. In the two-class scenario (benign and malignant), the weights were 1.0627 and 0.9442, respectively.

A custom function was implemented to calculate network error, allowing for different weights for each class and adjusting the impact of each class's examples on the final error calculation. In CapsNet, "Routing" refers to the process by which capsules in one layer dynamically send their outputs to the next layer based on how well the predictions align [27, 38]. It uses algorithms like dynamic routing by agreement to refine connections iteratively. The "Dimension of Capsules" pertains to the length of the vectors that capsules output, which encodes complex feature attributes such as pose and orientation, allowing for a richer representation of data than traditional scalar outputs. The CapsNet parameters were optimized through 20 experimental repetitions, resulting in a capsule dimension of ten and a routing number of 5. The evaluation results of two-, three-, and four-class classifications are presented in Table 3. Fig. 4 shows the training, testing, and validation accuracy and loss for experiment scenario (I) across two-, three-, and four-class classifications.

10.22060/eej.2024.23377.5609

Table 3. Result assessment of two-, three, and four-class classification using experiment scenario (I)

Class/Parameters	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	Macro Average AUC (%)
2	98.36	98.41	98.36	98.36	100
3	94.94	94.99	94.94	94.91	99.33
4	94.22	94.23	94.22	94.20	99.50

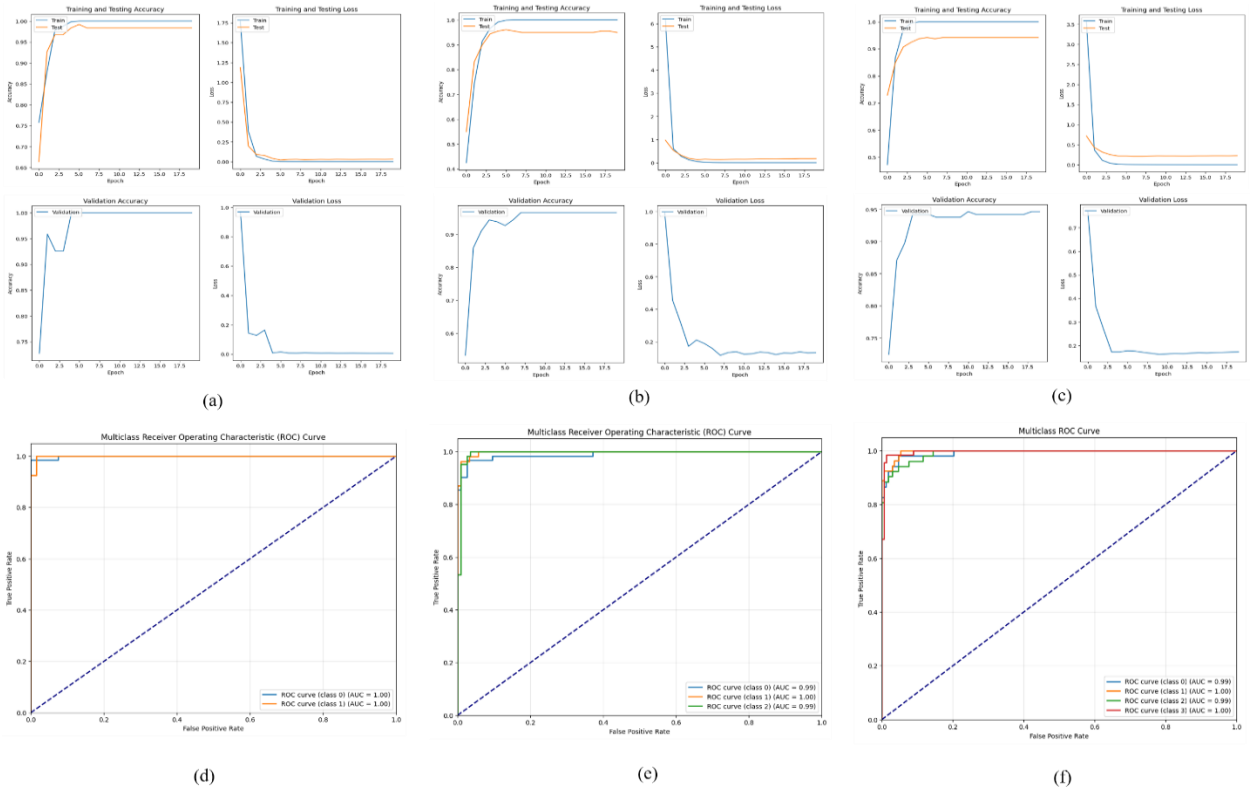

Fig. 4. The training, testing, and validation accuracy and loss for (a) two-, (b) three-, (c) four-class classifications and ROC for (d) two-, (e) three-, (f) four-class classifications in experiment scenario (I)

Experiment Scenario (II): In this approach, the minority class was augmented to match the size of the majority class, yielding 63 images per class, and the SMOTE method was used to balance the dataset. CapsNet parameters were optimized through 20 experimental repetitions, yielding a capsule dimension of ten and a routing number of 5. Table 4 visually represents the experiment scenario (II) evaluation results.

Fig. 5 shows the training, testing, and validation accuracy and loss for experiment scenario (II) across two-, three-, and four-class classifications.

Table 4. Result assessment of two-, three, and four-class classification using experiment scenario (II)

Class/Parameters	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	Macro Average AUC (%)
2	97.87	97.88	97.87	97.87	100
3	97.02	97.16	97.02	97.04	100
4	94.29	94.53	94.29	94.24	98.50

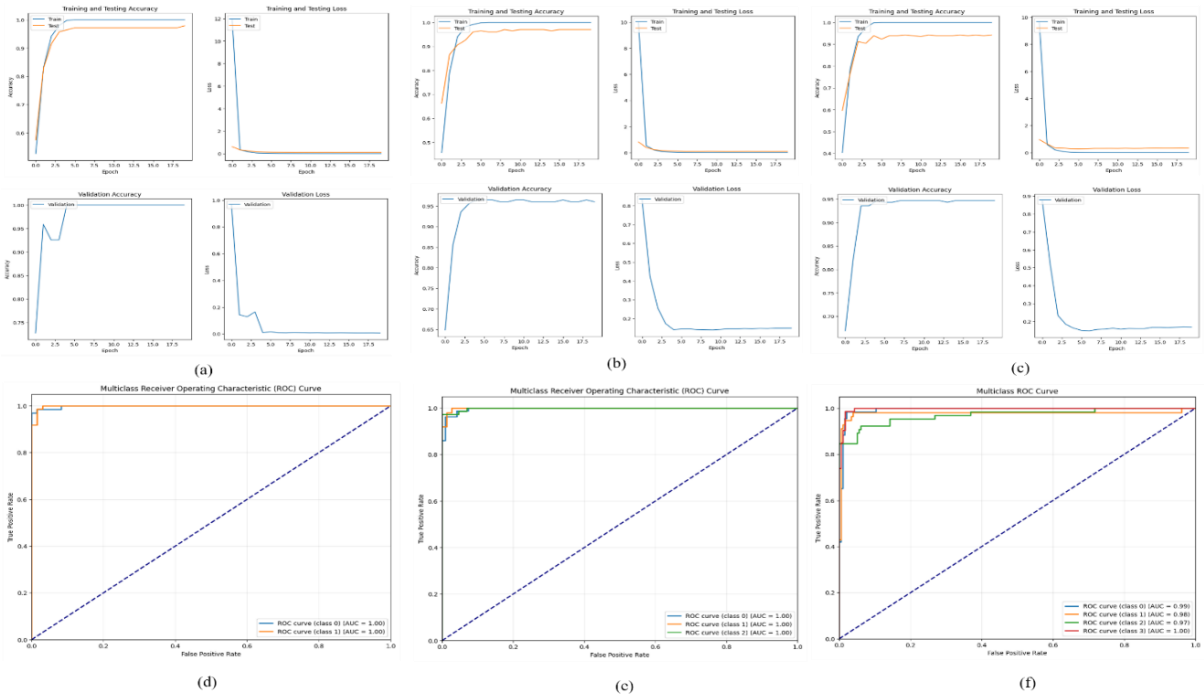

Fig. 5. The training, testing, and validation accuracy and loss for (a) two-, (b) three-, (c) four-class classifications and ROC for (d) two-, (e) three-, (f) four-class classifications in experiment scenario (II)

Experiment Scenario (III): The CapsNet hyperparameters were optimized using the GP method. Optimal performance was achieved with different routing and capsule dimensions for two-, three-, and four-class classifications, followed by a final investigation using 20 epochs. This approach identified the most accurate configurations for each classification task. For four-class classification, the optimal results were achieved

with a routing number of four and a capsule dimension of eight. The best performance was obtained in the three-class classification with a routing number of five and a capsule dimension of 14. For two-class classification, the optimal settings were a routing number of five and a capsule dimension of 20. The evaluation results of two-, three-, and four-class classifications of experiment scenario (III) are depicted in Table 5. Fig. 6 presents the training, testing, and validation accuracy and loss for experiment scenario (III) across two-, three-, and four-class classifications.

Table 5. Result assessment of two-, three, and four-class classification using experiment scenario (III)

Class/Parameters	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	Macro Average AUC (%)
2	98.51	98.56	98.51	98.51	100
3	97.51	97.61	97.51	97.52	99.33
4	95.05	95.08	95.05	95.06	99.50

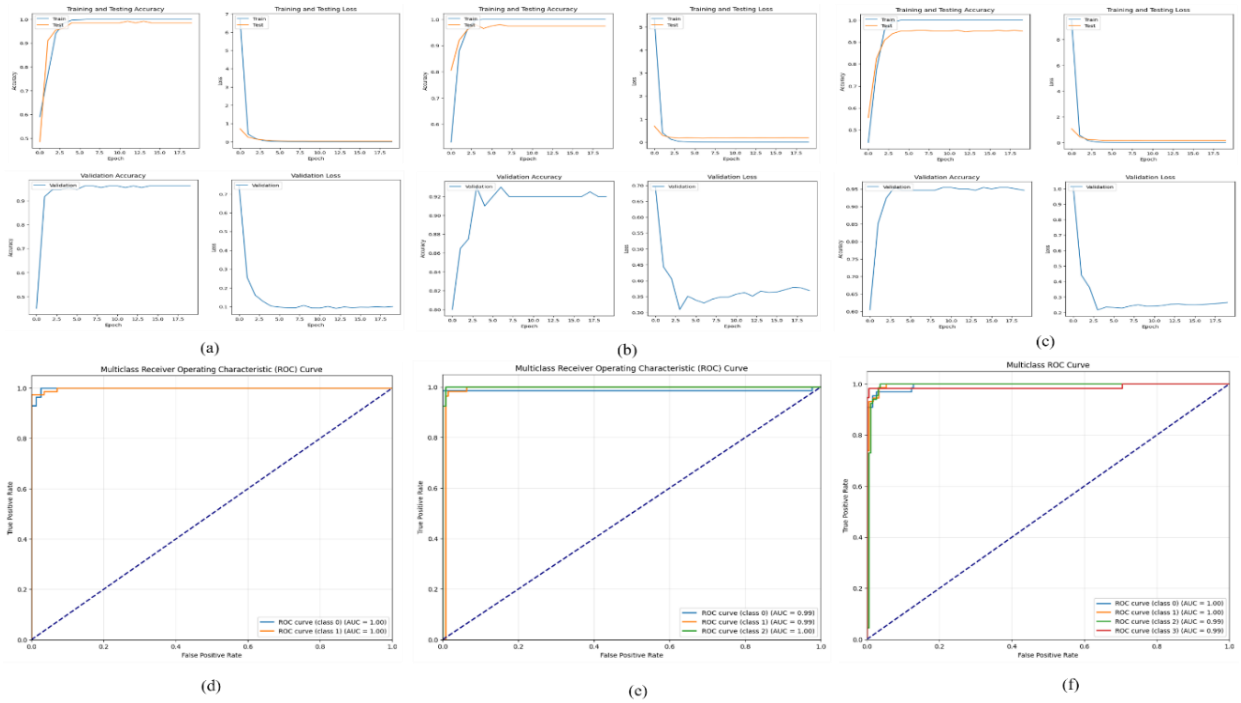

Fig. 6. The training, testing, and validation accuracy and loss for (a) two-, (b) three-, (c) four-class classifications and ROC for (d) two-, (e) three-, (f) four-class classifications in experiment scenario (III)

5. Discussion

The primary objective of this research was to explore three scenarios for classifying breast lesions across two, three, and four classes, which include benign, probably benign, suspicious, and malignant lesions. The proposed method addresses the challenges of data imbalance and the limited number of images in the RFTSBU dataset, which can impact classification results. It combines a pre-trained VGG16 CNN with CapsNet for breast cancer classification. Model accuracy is enhanced by applying GP optimization to fine-tune the CapsNet hyperparameters for two- and three-class classifications. In addition to Tables 3-5, Fig. 4, Fig. 5, Fig. 6, and Fig. 7 demonstrate that scenario (III) yields the best classification results on average, primarily due to the use of the GP method and the optimization of hyperparameters across two-, three-, and four-class classifications. In relative terms, scenario (III) has shown an average improvement of 0.009% and 0.005% over the (I) and (II) scenarios, respectively.

In this study, beyond evaluating the accuracy, precision, recall, and F-score, we employed the ROC curve to assess the model's discriminative capability, with AUC as a crucial performance metric. Scenario (III) achieved exceptional macro average AUC values of 100%, 99.33%, and 99.50% for two-class, three-class, and four-class classifications, respectively. These results underscore the model's discriminative strong power, demonstrating its effectiveness in distinguishing between the target classes. Compared to previous studies [1, 19, 20, 23], our model exhibits a marked improvement, likely due to the innovative approaches we implemented.

The computational complexity of various breast cancer detection models varies significantly due to their different approaches and model architectures. Jarosik *et al.*'s research is computationally intensive due to the direct processing of raw US RF signals [1]. Their CNN-based model requires extensive parameter optimization during training and inference, leading to high computational costs, especially with iterative epochs. Wei *et al.*'s method also demands substantial resources during training and inference because it

uses deep network layers, region proposal processes, and bounding box refinement for lesion detection and classification [16].

For MTL-based models, the computational complexity increases with the number of classification categories, the challenge of managing limited training data, and the need for extensive optimization processes [17]. Qiao *et al*.'s framework adds complexity with signal preprocessing (beamforming and STFT), running deep models like SCD-Net with YOLOv3 and ConvLSTM, and performing temporal tracking with the Kalman filter [18]. This multi-step approach requires high-performance hardware for practical training and deployment.

Kim *et al*. system also demonstrates high computational demands, with image preprocessing steps (entropy, phase, and attenuation image generation), CNN training, and inference, all compounded by multiple input data channels, convolution operations, and data augmentation strategies [19]. Similarly, Byra *et al*.'s method requires significant computational resources due to large RF input sizes, MTL architecture, and intensive convolution operations coupled with data augmentation and interpretability steps [20].

Lastly, Taleghamar *et al*.'s methodology involves deep learning models (ResNet and RAN), attention mechanisms, large input image sizes, and multi-parametric data, increasing computational complexity [23]. Training and evaluation become progressively more demanding with deeper networks, input channels, and larger datasets.

Compared to these approaches, the VGG16-CapsNet with GP optimization method, while computationally expensive due to its deep architecture and optimization techniques, is designed to handle complex tasks like hierarchical feature extraction and multi-class classification more effectively than simpler models. While methods like CNNs are computationally cheaper, they may not achieve the same level of performance in nuanced tasks such as breast lesion classification. Thus, a trade-off between computational cost and model performance must be considered when selecting a model for clinical applications.

Results were obtained independently using VGG16 and CapsNet to compare VGG16, CapsNet, and the VGG16-CapsNet combination comprehensively. Fig. 8 illustrates the average performance of these three approaches across scenarios (I), (II), and (III). On average, the VGG16-CapsNet approach demonstrates a performance improvement of 4.31% compared to the VGG16 approach and 9.06% compared to CapsNet.

Moreover, Fig. 8 highlights the impact of incorporating the GP method (scenario (III)) with the VGG16-CapsNet approach. This optimization method enhances performance across all three approaches—VGG16, CapsNet, and VGG16-CapsNet. Compared to scenarios (I) and (II), scenario (III) incorporates GP optimization, which offers a more comprehensive solution by balancing the dataset and fine-tuning the model's hyperparameters for optimal performance. This approach enhances the model's ability to capture complex data patterns, improves its generalization capabilities, reduces the time required to find optimal hyperparameters, and ultimately results in higher accuracy and robustness in breast lesion classification.

The findings demonstrate that incorporating CapsNet into the VGG16 architecture significantly improves feature detection and classification accuracy. The synergy between VGG16's ability to extract dense, hierarchical features and CapsNet's strength in modelling spatial relationships and preserving feature hierarchies made this combination particularly effective for classifying breast lesions. Other pre-trained networks, while powerful, may not provide the same balance of simplicity, detailed feature extraction, and compatibility with CapsNet's architecture. The dynamic routing between capsules in CapsNet is crucial in enhancing the model's ability to accurately classify breast cancer cases by emphasizing the most relevant features during the classification process. This integration results in a more efficient and accurate diagnostic tool, particularly in scenarios requiring complex classifications.

In this study, we selected the CapsNet architecture for its ability to capture spatial hierarchies of features essential for accurate breast lesion classification. The network's dynamic routing mechanism helps prioritize and apply these features across various classification tasks as follows:

Two-Class Classification (Benign vs. Malignant): CapsNet focuses on key features like lesion size, boundary irregularity, and echogenicity to differentiate benign (smoother, uniform) from malignant (irregular, complex) lesions. Dynamic routing emphasizes lesion shape and structural integrity.

Three-Class Classification (Benign, Suspicious, Malignant): In this task, CapsNet considers additional features like internal texture and lesion orientation to distinguish "suspicious" lesions, which share traits with both benign and malignant categories. Dynamic routing emphasizes subtler features like lesion homogeneity and internal echoes.

Four-Class Classification (Benign, Probably Benign, Suspicious, Malignant): For this detailed classification, CapsNet focuses on finer features like microcalcifications, margin clarity, and small histological changes. Dynamic routing is adjusted to distinguish "probably benign" from "suspicious" lesions by recognizing intricate patterns in lesion structure.

CapsNet adapts its dynamic routing as the classification complexity increases to focus on progressively finer details, highlighting the most distinguishing features. This flexibility makes CapsNet highly effective and allows it to refine its performance across various classification scenarios. Table 6 provides a comparison of the approaches proposed by other researchers.

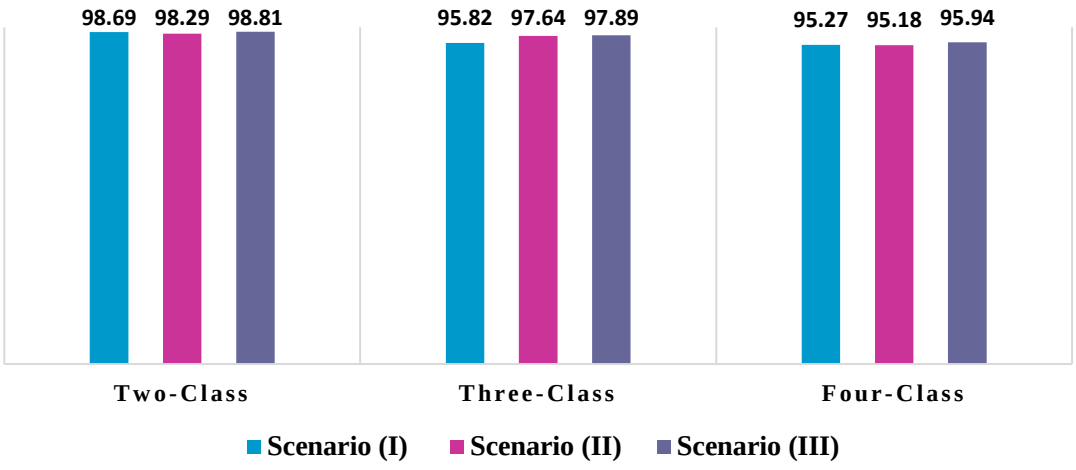

Fig. 7. Average results for scenario (I), scenario (II), and scenario (III) across two-, three-, and four-class classifications

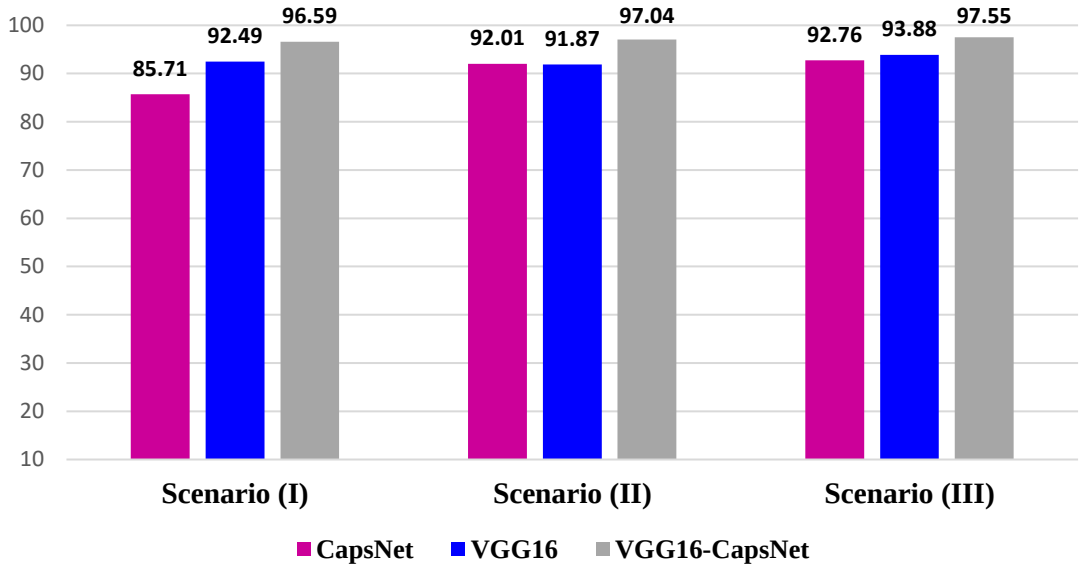

Fig. 8. Comparison of average results for scenario (I), scenario (II), and scenario (III) using CapsNet, VGG16 and VGG16-CapsNet approaches

Table 6. Comparison of relevant RF time series and US RF B-mode methods

Study	Database	Feature Extraction Method	Classifier	No. of classes	Results
Jarosik *et al.* [1]	OASBUD (78 subjects and 100 US RF data)	CNN-1D, CNN-2D, CNN-1D-2D	CNN-1D, CNN-2D, CNN-1D-2D	2	Accuracy= 70.00% with CNN-1D-2D
Kim *et al.* [19]	OASBUD (78 subjects and 100 US RF data)	Entropy and phase images, B-mode images, Attenuation images	End-to-end ensemble CNN	2	Accuracy=83.00% with DenseNet-201
Wei *et al.* [16]	OASBUD (78 subjects and 100 US RF data)	Faster R-CNN	Faster R-CNN	2	Accuracy>95%
Proposed Method	**RFTSBU (118 subjects and 220 RF B-mode data)**	**VGG16**	**CapsNet**	**2**	**Accuracy=98.81%**
				3	**Accuracy=97.89%**
				4	**Accuracy=95.94%**

The following is a summary of this investigation's main benefits:

- Finding solutions for the balancing issues and the small number of US RF B-mode images in the RFTSBU dataset to improve the classification outcomes.
- Using the VGG16 model as input for CapsNet to extract deep features.
- Applying GP optimization to adjust the hyperparameters of CapsNet.
- Creating a classifier for breast cancer using CapsNet.

5-1- Limitations and Future Works

The RFTSBU dataset used in this study is relatively small, comprising 220 data points from 118 patients, which may limit the model's robustness and generalizability to broader populations. This limited sample size poses challenges for training deep learning models. Additionally, the RFTSBU dataset originates from a specific US system (SuperSonic Imagine Aixplorer®), which could restrict the model's applicability to other medical centers, devices, or patient populations with different characteristics such as age or ethnicity.

While GP optimization was employed to fine-tune hyperparameters and enhance performance, it can be computationally intensive and sensitive to kernel choices, potentially leading to suboptimal outcomes in some cases. Although the hybrid model demonstrates strong experimental results, its performance in real-world settings is not guaranteed due to variations in US image quality, device calibration, and noise levels across healthcare environments. Moreover, its reliance on data from a specific US system highlights potential challenges in generalizing the model's effectiveness to data from other devices with differing imaging protocols, resolutions, and performance characteristics.

Future works for improving the breast lesion classification model can focus on the following:

Expanding the dataset is a critical step to improving the model's generalization and clinical applicability. Collecting more extensive and diverse datasets, including samples from various medical centers, US systems, and patient demographics (e.g., age and ethnicity), will help the model adapt to broader real-world

scenarios. This expansion reduces overfitting risks and enhances robustness, ensuring reliable performance across different clinical settings.

Optimization methods also hold significant potential for future work. While this study utilized GP optimization, exploring alternative techniques such as Bayesian optimization, genetic algorithms, or reinforcement learning could yield better hyperparameter tuning while reducing computational overhead. Additionally, addressing real-world clinical challenges, such as robustness to noise and artifacts, is essential. Incorporating techniques like image denoising, adversarial training, and augmented datasets with simulated noise will ensure the model can withstand distortions commonly encountered in clinical practice. Multi-center trials with datasets from diverse hospitals will further enable comprehensive performance evaluation, providing the model's reliability across varying imaging conditions and patient populations.

6. Conclusion

Impressive breast cancer screening relies on early detection and treatment. While US RF B-mode provides a novel, equipment-independent approach, challenges like class imbalance and limited interpretability hinder its clinical application. This study presents a hybrid deep learning model for breast lesion classification that integrates a pre-trained CNN (VGG16) with a CapsNet to address critical challenges in breast cancer screening, including class imbalance and small sample sizes. The model leverages the SMOTE and data augmentation to mitigate data imbalance and enhance training stability. GP optimization was applied to fine-tune the hyperparameters of the CapsNet, ensuring optimal performance across various classification scenarios.

The experimental results demonstrate that the proposed hybrid model achieves superior classification accuracy compared to alternative approaches. Specifically, it achieves average accuracies of 98.81%, 97.89%, and 95.94% for two-, three-, and four-class classifications, respectively. This performance reflects the model’s robustness in handling complex, multi-class breast lesion data and accurately capturing key lesion characteristics, such as size, shape, orientation, and internal texture.

Integrating VGG16 with CapsNet allows for effective feature extraction and hierarchical representation of lesion attributes, while GP optimization fine-tunes the model's hyperparameters to enhance diagnostic precision. This model improves classification accuracy and offers significant interpretability, making it a viable solution for clinical applications, particularly in classifying breast lesions using US RF B-mode images. The results underscore the potential of advanced hybrid deep learning models in real-world medical diagnostics, with promising implications for further research in improving breast cancer detection and classification.

References

[1] P. Jarosik, Z. Klimonda, M. Lewandowski, M. Byra, Breast lesion classification based on ultrasonic radio-frequency signals using convolutional neural networks, Biocybernetics and Biomedical Engineering, 40(3) (2020) 977-986.
[2] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, J.R. Gralow, F. Cardoso, S. Siesling, I. Soerjomataram, Current and future burden of breast cancer: Global statistics for 2020 and 2040, The Breast, 66 (2022) 15-23.
[3] G.C. Observatory, Population Fact Sheets of Iran 2020, in: Population Fact Sheets, International Agency for Research on Cancer, France 2021, pp. 1-2.
[4] M.L. Brown, F. Houn, E.A. Sickles, L.G. Kessler, Screening mammography in community practice: Positive predictive value of abnormal findings and yield of follow-up diagnostic procedures, AJR Am. J. Roentgenol., 165(6) (1995) 1373-1377.
[5] S.P. Poplack, A.N. Tosteson, M.R. Grove, W.A. Wells, P.A. Carney, Mammography in 53,803 women from the New Hampshire mammography network, Radiology, 217(3) (2000) 832-840.
[6] S.P. Poplack, P.A. Carney, J.E. Weiss, L. Titus-Ernstoff, M.E. Goodrich, A.N. Tosteson, Screening mammography: Costs and use of screening-related services, Radiology, 234(1) (2005) 79-85.
[7] J. Haas, C. Kaplan, A. McMillan, L.J. Esserman, Does timely assessment affect the anxiety associated with an abnormal mammogram result?, J. Womens Health Gend. Based Med., 10(6) (2001) 599-605.
[8] N.M. Lindberg, D. Wellisch, Anxiety and compliance among women at high risk for breast cancer, Ann. Behav. Med., 23(4) (2001) 298-303.
[9] R.C. Burack, M.S. Simon, M. Stano, J. George, J. Coombs, Follow-up among women with an abnormal mammogram in an HMO: is it complete, timely, and efficient?, Am. J. Manag. Care, 6(10) (2000) 1102-1113.
[10] K. Drukker, M.L. Giger, K. Horsch, M.A. Kupinski, C.J. Vyborny, E.B. Mendelson, Computerized lesion detection on breast ultrasound, Med. Phys., 29(7) (2002) 1438-1446.
[11] H. Yu-Len, C. Dar-Ren, L. Ya-Kuang, Breast cancer diagnosis using image retrieval for different ultrasonic systems, in: 2004 International Conference on Image Processing, 2004. ICIP '04., 2004, pp. 2957-2960 Vol. 2955.
[12] B.O. Anderson, R. Shyyan, A. Eniu, R.A. Smith, C.H. Yip, N.S. Bese, L.W. Chow, S. Masood, S.D. Ramsey, R.W. Carlson, Breast cancer in limited-resource countries: An overview of the Breast Health Global Initiative 2005 guidelines, The breast journal, 12 (2006) S3-15.
[13] M. Costantini, P. Belli, R. Lombardi, G. Franceschini, A. Mulè, L. Bonomo, Characterization of solid breast masses: Use of the sonographic breast imaging reporting and data system lexicon, J. Ultrasound Med., 25(5) (2006) 649-659;.

[14] C.M. Chen, Y.H. Chou, K.C. Han, G.S. Hung, C.M. Tiu, H.J. Chiou, S.Y. Chiou, Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks, Radiology, 226(2) (2003) 504-514.
[15] B. Sahiner, H.P. Chan, M.A. Roubidoux, L.M. Hadjiiski, M.A. Helvie, C. Paramagul, J. Bailey, A.V. Nees, C. Blane, Malignant and benign breast masses on 3D US volumetric images: effect of computer-aided diagnosis on radiologist accuracy, Radiology, 242(3) (2007) 716-724.
[16] K. Wei, B. Wang, J. Saniie, Faster region convolutional neural networks applied to ultrasonic images for breast lesion detection and classification, in: 2020 IEEE International Conference on Electro Information Technology (EIT), IEEE, Chicago, IL, USA, 2020, pp. 171-174.
[17] B. Behboodi, H. Rasaee, A. Tehrani, H. Rivaz, Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning, in: 2021 Ultrasonic Imaging and Tomography Conference, 2021.
[18] M. Qiao, Z. Fang, Y. Guo, S. Zhou, C. Chang, Y. Wang, Breast calcification detection based on multichannel radiofrequency signals via a unified deep learning framework, Expert Systems with Applications, 168(114218) (2021) 1-11.
[19] S. Kim, J. Park, J. Yi, H. Kim, End-to-end convolutional neural network framework for breast ultrasound analysis using multiple parametric images generated from radiofrequency signals, Applied Sciences, 12(10) (2022) 1-17.
[20] M. Byra, P. Jarosik, K. Dobruch-Sobczak, Z. Klimonda, H. Piotrzkowska-Wroblewska, J. Litniewski, A. Nowicki, Joint segmentation and classification of breast masses based on ultrasound radio-frequency data and convolutional neural networks, Ultrasonics, 121(6) (2022) 1-9.
[21] X. Li, Y. Sang, X. Ma, Y. Cai, Quantitative feature classification for breast ultrasound images using improved naive bayes, IET Image Processing, 17(5) (2023) 1417-1426.
[22] G.R. Gare, J. Li, R. Joshi, R. Magar, M.P. Vaze, M. Yousefpour, R.L. Rodriguez, J.M. Galeotti, W-Net: Dense and diagnostic semantic segmentation of subcutaneous and breast tissue in ultrasound images by incorporating ultrasound RF waveform data, Med. Image Anal., 76 (2022) 1-14.
[23] H. Taleghamar, S.A. Jalalifar, G.J. Czarnota, A. Sadeghi-Naini, Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy, Sci. Rep., 12(1) (2022) 1-13.
[24] L.M. AbouEl-Magd, A. Darwish, V. Snasel, A.E. Hassanien, A pre-trained convolutional neural network with optimized capsule networks for chest X-rays COVID-19 diagnosis, Cluster Computing, 26(2) (2023) 1389-1403.
[25] S. Tiwari, A. Jain, Convolutional capsule network for COVID-19 detection using radiography images, International journal of imaging systems and technology, 31(2) (2021) 525-539.
[26] A.R. Bushara, R.S. Vinod Kumar, S.S. Kumar, An ensemble method for the detection and classification of lung cancer using Computed Tomography images utilizing a capsule network with Visual Geometry Group, Biomedical signal processing and control, 85 (2023) 104930.
[27] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, B.Y. Edward, Capsule networks – A survey, Journal of King Saud University - Computer and Information Sciences, 34(1) (2022) 1295-1310.
[28] M.U. Haq, M.A. Sethi, A.U. Rehman, Capsule Network with Its Limitation, Modification, and Applications—A Survey, in: Machine Learning and Knowledge Extraction, 2023, pp. 891-921.
[29] G.E. Hinton, S. Sabour, N. Frosst, Matrix capsules with EM routing, in: International Conference on Learning Representations (ICLR18), 2018, pp. 1-15.
[30] Y. Wang, D. Ning, S. Feng, A Novel Capsule Network Based on Wide Convolution and Multi-Scale Convolution for Fault Diagnosis, in: Appl. Sci., 2020, pp. 1-16.
[31] C.E. Rasmussen;, C. Williams;, Gaussian processes for machine learning, MIT Press, 2006.
[32] I. Guyon, A. Statnikov, B.B. Batu, Automated machine learning: Methods, systems, challenges, Springer, 2019.
[33] D.R. Jones, M. Schonlau, W.J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, Journal of Global Optimization, 13(4) (1998) 455-492.

[34] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, Journal of Artificial Intelligence Research, 16 (2002) 321-357.
[35] D. Elreedy, A.F. Atiya, A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance, Journal of information science, 505 (2019) 32-64.
[36] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11) (1998) 2278-2324.
[37] A. Kaya, A.S. Keceli, C. Catal, H.Y. Yalic, H. Temucin, B. Tekinerdogan, Analysis of transfer learning for deep neural network based plant classification models, Computers and Electronics in Agriculture, 158 (2019) 20-29.
[38] Sara Sabour, Nicholas Frosst, G.E. Hinton, Dynamic routing between capsules, Adv. Neural Inf. Process. Syst., (2017).
[39] G.E. Hinton, A. Krizhevsky, S.D. Wang, Transforming auto-encoders, in: T. Honkela, W. Duch, M. Girolami, S. Kaski (Eds.) Artificial Neural Networks and Machine Learning – ICANN 2011, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 44-51.