Curr Cancer Rep, 2024, 6(1): 205-224
DOI: 10.25082/CCR.2024.01.002

RESEARCH ARTICLE

Deep learning based capsule networks for breast cancer classification using ultrasound images

Stephen Afrifa[1,2,*] **Vijayakumar Varadarajan**[3,4,*] **Tao Zhang**[1] **Peter Appiahene**[2] **Daniel Gyamfi**[5] **Rose-Mary Owusuaa Mensah Gyening**[6] **Jacob Mensah**[2] **Samuel Opoku Berchie**[2]

[1] Department of Information and Communication Engineering, Tianjin University, Tianjin 300072, China
[2] Department of Information Technology and Decision Sciences, University of Energy and Natural Resources, Sunyani 00233, Ghana
[3] Research Division, Swiss School of Business and Management, Geneva 1213, Switzerland
[4] International Divisions, Ajeenkya D. Y. Patil University, Pune 412105, India
[5] Department of Mathematics and Statistics, Saint Louis University, MO 63103, Missouri, USA
[6] Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi 00233, Ghana

Correspondence to: 1. Stephen Afrifa, Department of Information and Communication Engineering, Tianjin University, Tianjin 300072, China;
Email: afrifastephen@tju.edu.cn
2. Vijayakumar Varadarajan, Research Division, Swiss School of Business and Management, Geneva 1213, Switzerland;
Email: vijayakumar.varadarajan@gmail.com

Received: March 28, 2024;
Revised: July 29, 2024;
Accepted: August 21, 2024;
Published: August 27, 2024.

Citation: Afrifa S, Varadarajan V, Zhang T, et al. Deep learning based capsule networks for breast cancer classification using ultrasound images. *Curr Cancer Rep*, 2024, **6**(1): 205-224.
https://doi.org/10.25082/CCR.2024.01.002

Abstract: Purposes: Breast cancer (BC) is a disease in which the breast cells multiply uncontrolled. Breast cancer is one of the most often diagnosed malignancies in women worldwide. Early identification of breast cancer is critical for limiting the impact on affected people's health conditions. The influence of technology and artificial intelligence approaches (AI) in the health industry is tremendous as technology advances. Deep learning (DL) techniques are used in this study to classify breast lumps. **Materials and Methods**: The study makes use of two distinct breast ultrasound images (BUSI) with binary and multiclass classification. To assist the models in understanding the data, the datasets are exposed to numerous preprocessing and hyperparameter approaches. With data imbalance being a key difficulty in health analysis, due to the likelihood of not having a condition exceeding that of having the disease, this study applies a cutoff stage to impact the decision threshold in the datasets data augmentation procedures. The capsule neural network (CapsNet), Gabor capsule network (GCN), and convolutional neural network (CNN) are the DL models used to train the various datasets. **Results**: The findings showed that the CapsNet earned the maximum accuracy value of 93.62% while training the multiclass data, while the GCN achieved the highest model accuracy of 97.08% when training the binary data. The models were also evaluated using a variety of performance assessment parameters, which yielded consistent results across all datasets. **Conclusion**: The study provides a non-invasive approach to detect breast cancer; and enables stakeholders, medical practitioners, and health research enthusiasts a fresh view into the analysis of breast cancer detection with DL techniques to make educated judgements.

Keywords: breast cancer, capsule network, deep learning, Gabor capsule, medical imaging

1 Introduction

Breast cancer is the world's second greatest cause of mortality among women [1]. In 2019, there are predicted to be 268,600 new instances of invasive breast cancer identified in women in the United States, as well as 62,930 new cases of non-invasive breast cancer [2]. Tumors in the breast are often produced as a result of aberrant development of breast cells, and they may be detected using several imaging modalities. Early detection is the most effective strategy to improve treatment and survival. Currently employed imaging methods include ultrasound (US), magnetic resonance imaging (MRI), computed tomography (CT), and mammography. Breast ultrasound imaging is one of the most efficient screening tools for the diagnosis and prognosis of breast abnormalities among these modalities [3]. It is a popular screening procedure since it is non-invasive, real-time, and inexpensive. Deep learning (DL) has emerged as a popular approach for knowledge discovery, with promising outcomes in cybersecurity [4] and health [5].

Multiple classifier algorithms have recently been used to medical datasets to do predictive analysis on individuals and their medical diagnoses. For instance, employing machine learning (ML) and deep learning (DL) approaches to monitor tumor behavior in breast cancer patients. It should be noted that data imbalance is a key problem in medical image diagnostics utilizing

artificial intelligence (AI) approaches since the likelihood of not having the condition is greater than the chance of having it. The improvement of computer-aided diagnosis (CAD) technologies in recent years has resulted in an effective cancer diagnostic technique. Several preprocessing approaches are used in the CAD system analysis to assist the models grasp the datasets [6]. The convolution neural network (CNN) is the most often used model for image recognition and classification in the domain of deep learning (DL) - CAD (DL-CAD) [7]. This study applies novel deep learning approaches, such as capsule neural network (CapsNet), Gabor capsule network (GCN), and convolutional neural network (CNN), to improve the accuracy of breast cancer classification in women. The goal is to improve the classifier's performance by preparing the dataset by offering an appropriate approach for dealing with the unbalanced dataset and missing values. The study's uniqueness can be observed in the way it compares three advanced deep learning architectures in depth and how it incorporates the Gabor filter into the capsule network framework, something that is uncommon in the field. Furthermore, the study offers lucid descriptions of the image preprocessing procedures, the reasoning behind the use of every approach, and the suitability of performance metrics like accuracy, precision, recall, and ROC-AUC score for the given classification problem. Notable is the creative use of the Gabor filter into a capsule network model (GCN), which may open up new possibilities for feature extraction in medical image analysis.

1.1 Major contributions of the study

The following are the major contributions of the study:

(1) To classify breast cancer masses, the study uses deep learning models of the capsule neural network, Gabor capsule network, and convolutional neural network.

(2) The deep learning models in the present study are trained using two separate breast ultrasound images that reflect both binary and multiclass data.

(3) The study also applies a unique approach to check for class imbalance by establishing a cut-off score that influences the decision threshold in the augmentation procedures.

(4) The deep learning models' efficacy is evaluated using tenfold cross validation and many performance assessment indicators.

(5) The study represents the most recent use of deep learning in the analysis and classification of breast cancer masses.

1.2 Cutting-edge approaches for breast cancer classification

This section describes cutting-edge strategies for tumor classification using breast ultrasound images as summarized in Table 1. CNN models have advanced significantly in breast ultrasound imaging in recent years. Despite several studies been used in the domain of breast cancer categorization, this study provides an advancement with similarities and differences of the used cutting-edge models. This study proposes using deep learning algorithms to classify breast cancer. The models used include the capsule neural network (CapsNet), Gabor capsule network (GCN), and convolutional neural network (CNN) models. Table 1 shows that the bulk of the research evaluated used machine learning techniques, with only a handful considering deep learning utilizing CNN. The researchers used k-Nearest Neighbor (kNN), support vector machine (SVM), logistic regression (LR), random forest (RF), decision tree (DT), and XGBoost (XGB) machine learning models. They also trained their models using a single dataset with binary classifications. However, in the present study, two separate breast ultrasound pictures (BUSI) with both binary and multiclass classification are used. The current study offers a novel method for classifying breast cancer masses in the study domain. This gives the study an advantage over other existing studies in the domain.

Table 1 Cutting-edge analysis with the proposed models

Study	Year	Employed Models	Dataset	Data description
Bokhare and Jha [1]	2023	ML: kNN, SVM, RF, LR	Wisconsin breast cancer data	Binary data
Zhao and Jiang [8]	2023	ML: SVM, kNN, DT, XGB	Male breast cancer from SEER hospital	Binary data
Prodan et al. [9]	2023	DL: CNN	Mammogram	Binary data
Ardakani et al. [10]	2023	ML: RF and SVM	Breast ultrasound data	Binary data
Proposed model		DL: CapsNet, GCN, CNN	US – BUSI and BUSI	Both binary (US – BUSI) -and multiclass (BUSI)

The rest of this paper is structured as follows. Section 2 includes works that are related. Section 3 describes the research technique. Section 4 presents the study's experimental outcomes. Section 5 is an in-depth overview of the study titled "Discussion." Section 6 presents the

limitations of the study and Section 7 ends with the conclusion and future work.

2 Related works

Artificial intelligence (AI) techniques such as deep learning (DL) have been used in the study of biomedicine. This section includes similar works in the subject of applying DL for modeling biological research to supplement our study with previous works. DL algorithms, notably convolutional neural networks (CNNs), dominate cutting-edge techniques in breast mass classification utilizing ultrasound imaging, since they excel at automated feature extraction from pixel data, resulting in better classification accuracy. Transfer learning improves these models, particularly when the data is sparse. Attention mechanisms in CNNs fine-tune the emphasis on important picture areas, improving interpretation. Generative Adversarial Networks (GANs) are used for data augmentation, producing synthetic pictures to improve model resilience. Despite advances, issues in model interpretability, data quality, and generalization remain, necessitating more study for optimal clinical use.

To begin with, Bokhare and Jha [1] demonstrated how to classify breast cancer data using several machine learning (ML) models. The effectiveness of models was compared through results using an accuracy standard, which had not before been done. Their classifiers were evaluated, examined, and compared. The top ML method for the breast cancer data set is the classifier, decision tree, which achieves the maximum accuracy of 97.08% among all models. Additionally, Zhao and Jiang [8] developed a technique for comparing the efficacy of multiple machine learning (ML) models and nomograms to predict distant metastases in male breast cancer (MBC) patients and interpreting the optimum ML model using the SHapley Additive exPlanations (SHAP) framework in a research. Their xGBoost (XGB) model was superior predictor of distant metastasis among MBC patients than other ML models and nomogram; additionally, the XGB model was a potent model for predicting bone and lung metastasis. When combined with SHAP values, it may aid clinicians in intuitively understanding the influence of each variable on result. In another related work, Boumaraf et al. [11] contrasted traditional machine learning (CML) versus deep learning (DL)-based approaches. Their results indicated that DL techniques beat CML approaches with accuracy ranging from 94.05% to 98.13% for binary classification and 76.77% to 88.95% for eight-class classification.

Furthermore, Yang et al. [12] used magnetic resonance imaging morphological features (MRI-MF), Radiomics, and deep learning (DL) techniques based on dynamic contrast-enhanced MRI (DCE-MRI) to create an effective model for assessing lymphovascular invasion (LVI) status in patients with BC. Their MRI-MF model was developed with conventional MR features using logistic analyses. Their joint model incorporating MRI-MF, Radiomics, and DL approaches effectively determined the LVI status in patients with BC before surgery. In a study by Sahu et al. [13], they created five new deep hybrid convolutional neural network-based breast cancer detection frameworks. Their suggested hybrid systems outperformed the respective base classifiers while retaining the benefits of both networks. In all of the datasets tested, their experimental findings confirmed the superiority of the suggested ShuffleNet- ResNet system over the present state-of-the-art techniques. Furthermore, their suggested approach demonstrated higher accuracy for abnormality and malignancy diagnosis in the mini-DDSM and BUSI datasets, respectively. Jabeen et al. [3] offered a novel framework for breast cancer classification from ultrasound pictures in their study, which used deep learning and the fusion of the best selected characteristics. When compared to other models in the research domain, their experimental model, the convolutional neural network (CNN), produced substantial results. Their research demonstrated the effectiveness of CNN in the teaching of medical imagery. To diagnose ultrasound breast tumors, Sirjani et al. [14] created a unique deep learning architecture based on the InceptionV3 network. The suggested architecture's primary selling points were turning the InceptionV3 modules to residual inception modules, increasing their number, and changing the hyperparameters. Their research shown that the upgraded InceptionV3 can accurately diagnose breast cancers, possibly eliminating the need for biopsy in many situations.

In a similar work, Yu et al. [15] has created a hybrid framework for testing both feature ranking (FR) stability and cancer diagnostic efficacy. The architecture they developed identified a reliable FR algorithm for accurate breast cancer detection. Stable and effective characteristics might contribute to a better knowledge of breast cancer diagnosis and related decision-making applications. A case-control study was carried out at the Nganda Hospital Center in Kinshasa, Democratic Republic of Congo by Sulu et al. [16]. One hundred and sixty individuals with breast cancer (cases) were compared to 320 women without breast cancer (controls). STATA version 16 was used to analyze data with statistical significance considered at $p < 0.05$ in their

study. Their results indicated the presence of certain conventional risk factors, however, they failed to utilized ML or DL technologies to provide a non-invasive approach in the domain. Finally, Lee et al. [17] revealed that patient-specific cancer driver genes may be utilized to better reliably forecast cancer prognoses. They constructed patient-specific gene networks before utilizing modified PageRank to generate feature vectors that reflected the influence genes had on the patient-specific gene network to find patient-specific cancer driver genes. Their findings revealed that diverse cancer driver information might be linked to cancer prognosis.

3 Materials and methods

This section describes the methodologies employed in this investigation. It includes everything from data gathering to the models used and their performance evaluation measures. The techniques are thoroughly examined in the following subsections.

3.1 Data collection

The study made use of two (2) publicly accessible datasets of breast ultrasound images (BUSI). The datasets include the United States (US) - BUSI, which can be found at https://qamebi.com/breast-ultrasound-images-database [18], and was accessed on October 10, 2023; and the BUSI, which can be found at https://scholar.cu.edu.eg/?q=afahmy/pages/dataset [19], and was accessed on October 12, 2023. The US-BUSI has 123 and 109 ultrasound pictures of malignant and benign breast tumors, respectively, whereas the BUSI sample size includes 133 normal images with no cancerous masses, 437 images with cancerous masses, and 210 photos with benign masses. It should be noted that the US-BUSI dataset is binary, i.e., benign and malignant, whereas the BUSI data is multiclass, i.e., normal, benign, and malignant. The datasets are all exposed to the study's training techniques.

3.2 Data preprocessing

Data preparation is an important aspect of deep learning (DL) analysis since it allows the models to be trained. Data preparation enhances the quality, reliability, and efficacy of modeling methodologies [20], making the data better suitable to improving the comprehension and performance of deep learning models [21]. The following sections detail the various approaches used in the present study.

3.2.1 Data augmentation

Data augmentation is a technique for artificially augmenting the training set by creating modified copies of a dataset from existing data. By producing new and unique examples for training datasets, data augmentation improves the performance and outcomes of machine learning models. Deep learning models perform best when the dataset is large and diverse. In the augmentation procedures, the rotation and rescaling techniques are used. The photos are loaded and enhanced in Keras version 2.13 (Google LLC, Mountain View, California, United States) using the Image Data Generator class. These strategies were used to increase model prediction accuracy by adding more training data into models and avoiding data scarcity for better models. Table 2 illustrates the Image Data Generator class parameters and their related settings. To ensure that there was no data imbalance throughout the augmentation operations, the datasets were modified to a maximum value of 500, which impacted the decision threshold.

Table 2 ImageDataGenerator class parameters

Parameter	Value
Rotation	10^o
Width shift	2 pixels
Height shift	22 pixels
Shear	0.2 radians
Rescale	[0, 255] to [0, 1]
Fill mode	Nearest

3.2.2 Grayscale conversion

Each image is converted to grayscale from the RGB values of red (R), green (G), and blue (B). A grayscale image is one that is made up of different shades of gray (or black and white). Grayscale pictures have the potential to minimize the computing cost of an image processing activity [22]. This is because the number of channels has been reduced from three (RGB) to one (gray). Grayscale conversion was accomplished using the luminosity approach. Equation (1)

expresses the luminosity approach.

$$Y = 0.299R + 0.587G + 0.114B \tag{1}$$

Grayscale conversion helps to simplify algorithms and eliminates difficulties related with processing requirements.

3.2.3 Denoising technique

To reduce noise from photos, bilateral filtering is utilized. Bilateral filtering is a technique for smoothing pictures while keeping their edges. Bilateral filtering is a non-local denoising method. Non-local means (NLM) denoising is a technique for removing noise from a picture while maintaining its edges and features [21]. It compares each pixel in the picture to all the others, determining their similarity and utilizing that information to estimate the value of the noisy pixel. Comparable patches in an image have comparable values for any NLM denoising technique, even if they are not in the same position. By comparing all patches to each other, the algorithm can predict the image's structure and eliminate noise without blurring or distorting the edges or features [23]. Equation (2) depicts bilateral filtering numerically.

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in S} [(G_{\sigma s}(||p-q||)) .(G_{\sigma r}(|I_p - I_q|))].I_q \tag{2}$$

Where $BF[I]_p$ is the output of bilateral filtering at pixel p, I_q is the intensity at pixel q, $\frac{1}{W_p}$ is the normalization factor, W_p is the normalization term, $G_{\sigma s}$ is the Gaussian function for the standard deviation of spatial Gaussian component, $||p-q||$ is the Euclidean distance between p and q, and $G_{\sigma r}$ is the Gaussian function for the standard deviation of the range Gaussian component, additionally, I_p is the absolute intensity value at p, and I_q is the absolute intensity value at q.

3.2.4 Contrast limited adaptive histogram equalization

The Contrast Limited Adaptive Histogram Equalization (CLAHE) approach is used to improve the contrast of each image. CLAHE is a technique for improving visual contrast by spreading the intensity values in the image, particularly in low contrast images [24]. The normalized and denoised grayscale picture is processed using the CLAHE algorithm defined in the createCLAHE and OPENCV routines. After that, the CLAHE-enhanced picture is transformed back to RGB format.

3.2.5 Feature extraction

The labels and features are kept in distinct lists, and the features are concatenated into a single array. The feature array is reconfigured in four dimensions [25]. The feature array is then normalized using Scikit-learn's StandardScaler function. Using the K-Fold cross-validation approach, the normalized features and labels are divided into training, validation, and testing sets. The K-Fold object divides data into ten folds. The training data is then divided into 80% training and 20% validation sets for each fold. K-fold cross-validation is used to prevent network overfitting [20]. By providing an estimate of the model's performance on unseen data, the cross-validation approach helps to reduce overfitting. The validation sets are designed to be used to monitor the performance of a model during training.

3.3 The deep learning techniques

This section describes each model and its design in depth. The standard capsule network (CapsNet), a capsule network design with a convolutional Gabor layer as its first layer (Gab-CapsNet), and the convolutional neural network (CNN) are among the techniques employed in the study. The subsections that follow provide in-depth study of the models.

3.3.1 The deep capsule network model

The CapsNet architecture is made up of two convolutional layers, a primary capsule block, and a class capsule block [26]. A capsule network is a collection of neurons in which the activity vector reflects the instantiation parameters and the length of the vector denotes the probability of an entity's existence. This property enables capsule networks to learn picture properties like as deformations, location, and texture. There are three main approaches for implementing capsules: auto-encoders, vector capsules based on dynamic routing, and matrix capsules based on expectation-maximization (EM) routing [27]. To train the data in this study, vector capsules based on dynamic routing were used.

Vector capsules use vector routings to represent picture parameters. As activation functions in CNNs, ReLU, Sigmoid, and Tangent functions are employed [28]. The activation function for a vector capsule is known as a squash function, as shown in equation (3).

$$v_j = \frac{\| s_j \|^2}{1 + \| s_j \|^2} \frac{s_j}{\| s_j \|} \tag{3}$$

Where v_j = output of capsule j, and s_j = total input of the capsule. The total input value of capsule s_j is found by the weighted sum of the prediction vectors ($U_{i|j}$) in lower-layered capsules except in the first layer of the capsule network. The prediction vector is produced by multiplying the output u_j of a capsule in the lower layer by a weight matrix presented in equation (4) and equation (5).

$$s_j = \sum_i c_{ij} u_{i|j} \tag{4}$$

$$u_{i|j} = W_{ij} u_i \tag{5}$$

Where c_{ij} = coupling coefficients that are determined by the iterative dynamic routing process (Figure 1). The coupling coefficients are determined by a SoftMax function which is expressed in equation (6).

$$\mathrm{c_{ij}} = \frac{\exp(\mathrm{a_{ij}})}{\sum_{\mathrm{k}} \exp(\mathrm{a_{ik}})} \tag{6}$$

Where a_{ij} = log prior probability. In capsule networks, a margin loss has been proposed to determine the presence of objects of a particular class [28]. This margin loss is calculated as presented in equation (7).

$$\mathrm{L_k} = \mathrm{T_k} \max\left(0,\ \mathrm{m}^+ - \|\mathrm{v_k}\|^2\right)^2 + \lambda\left(1 - \mathrm{T_k}\right) \max\left(0, \|\mathrm{v_k}\| - \mathrm{m}^-\right)^2 \tag{7}$$

3.3.2 Summary of the deep capsule network

The feature maps of the CNN's block 4 dropout layer are sent into the first convolutional layer (Block5_Conv1). To generate 7×7 feature maps, the convolutional procedure is conducted using 128 filters, a kernel size of 7×7, and ReLU activation. To generate 2×2 feature maps, the second convolutional layer employs a convolution of 128 filters, a 6×6 kernel size, and ReLU activation. A convolutional layer plus a reshape layer make up the main capsule block. The convolutional layer creates 32-channeled 1×1 feature maps, which are subsequently molded and compressed into two 16-dimensional capsules. The main capsule layer's output is then transferred into the Class capsule layer. The Class capsule's output is transferred to the last tier, the lambda layer. This layer computes the class probabilities and acts as the output layer. This design uses capsule networks to record spatial hierarchies and feature interactions, which are critical for tasks such as picture categorization. The introduction of capsules enables the model to preserve more comprehensive and spatially aware representations of the input data, perhaps resulting in more accurate classification results than typical CNN designs. The deep capsule network design is depicted in Figure 1.

3.3.3 The deep Gabor capsule network model

The Gabor Capsule Network (Gab-CapsNet) design is similar to that of the capsule network paradigm. A convolutional block, a main capsule block, and a class capsule block comprise the Gab-CapsNet architecture. A Gabor filter is a linear filter in image processing that is created by mixing sinusoid and Gaussian functions. Several factors, including orientation, scale, aspect ratio, frequency, and phase, may be used to customize the Gabor filter. A Gabor filter can be represented mathematically in equation (8).

$$g\lambda,\ \theta,\ \sigma,\ \gamma(x,\ y) = \exp\left(-\frac{x^{'} + \gamma y^{'2}}{2\sigma^2}\right) \cos\left(2\pi \frac{x^{'}}{\lambda} + \phi\right) \tag{8}$$

Where $x^{'} = x\cos\theta + y\sin\theta$, $y^{'} = -x\sin\theta + y\cos\theta$. Gabor filters are powerful tools that may be used to recognize edges, analyze texture, and extract characteristics from pictures. Gabor filters may be used as preprocessing techniques to extract features for usage in CNNs [29]. A collection of global Gabor filters is used in the convolution process to extract attributes from an input picture [30]. This convolution operation (*) applied to the image and the global Gabor filter bank G(x, y: w, θ), generates a set of features ($\mathrm{O}_{m,n}$(x, y)) that can be represented mathematically by the equation (9):

$$Om,\ n(x,\ y) = I(x,\ y) \times G(x,\ y:\ w,\ \theta) \tag{9}$$

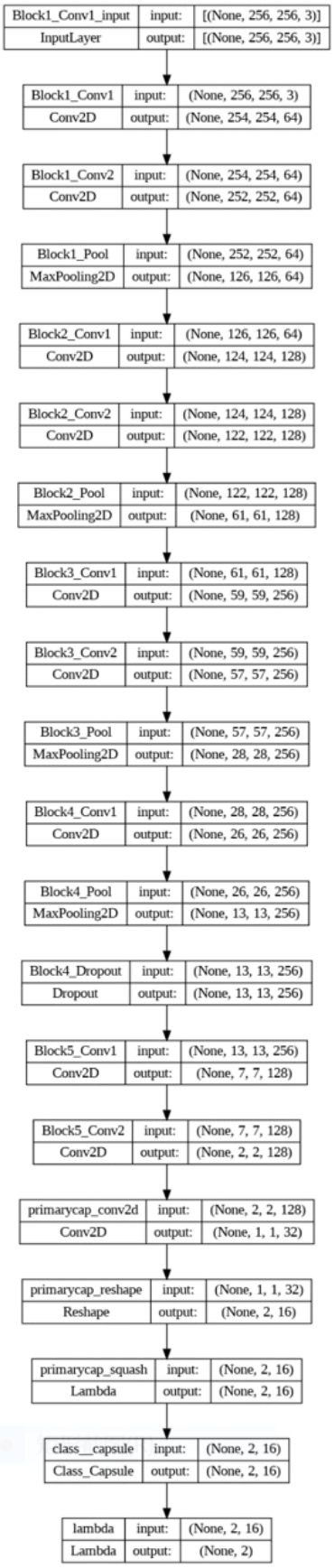

Figure 1 The deep capsule network architecture

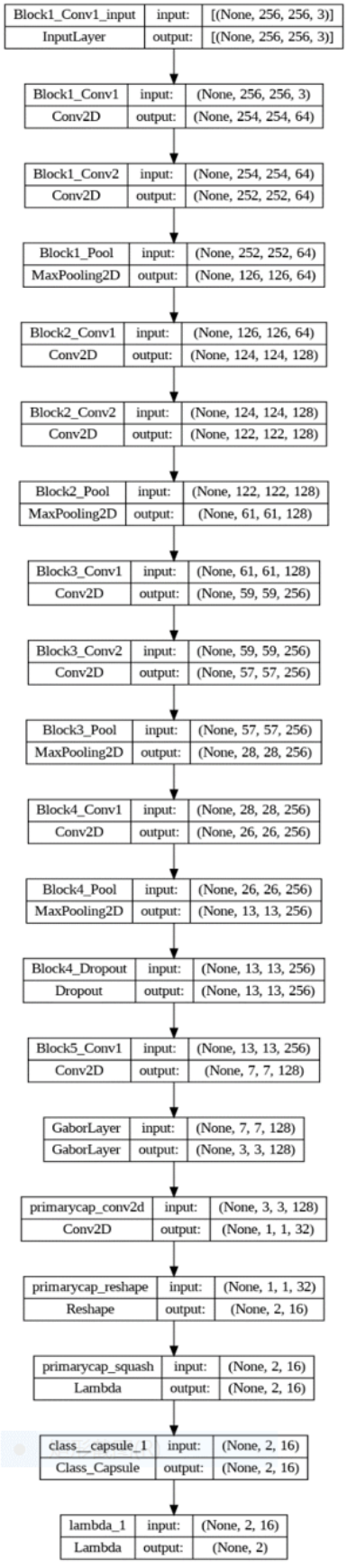

Figure 2 The deep Gabor capsule network architecture

3.3.4 Summary of the Deep Gabor Capsule Network

Figure 2 depicts the architecture of the Gab-CapsNet. The feature maps of the CNN's block 4 dropout layer are sent into the first convolutional layer (Block5_Conv1). To generate 7×7 feature maps, the convolutional procedure is conducted using 128 filters, a kernel size of 7×7, and ReLU activation. The second convolutional layer then uses 256 Gabor filters, a 7x7 kernel size, and ReLU activation to generate 7×7 feature maps [29]. A convolutional layer plus a reshape layer make up the main capsule block. The convolutional layer produces 32-channel 3x3 feature maps, which are subsequently molded and compressed into 18 capsules with 16 filters. The main capsule layer's output is transferred into the Class capsule layer. The output of the Class capsule is flattened and given to the lambda layer. This layer computes the class probabilities and acts as the output layer. The use of capsules ensures that spatial hierarchies and relationships between features are well-preserved, potentially leading to more robust and accurate classification; the lambda layer at the end facilitates the final prediction of class probabilities, completing the network's classification task. The Gab-CapsNet architecture makes use of the power of Gabor filters to enhance the feature extraction process, particularly in capturing edge and texture details, which are critical in distinguishing between different classes in image data.

3.3.5 The convolutional neural network model

A CNN is a deep learning system that can take an input picture, give priority to various aspects/objects in the image (via learnable weights and biases), and distinguish one from the other [31]. In the current investigation, the CNN was employed to categorize the breast masses. The CNN uses the visual geometry group network (VGGNet) architecture, namely the VGG16. Convolutional and max-pooling layers are included in the CNN design. The softmax activation function was used, with an a = 0.0001 regularization and a maximum iteration of 10. Softmax is a CNN activation function that is frequently used in the output layer, especially for multi-class classification tasks [32]. Because of the data's multi-class categorization, which includes normal, benign, and malignant, the softmax function was used. This converts raw scores (logits) into meaningful probabilities, allowing the model to make educated guesses about which class is the most likely label for a given input instance.

3.3.6 Summary of the convolutional neural network

Convolutional layers and max-pooling layers were employed by the CNN in the sequential model. Following the completion of the convolutional layers, the data was flattened to form three entirely connected layers for output with the softmax activation function. The total number of parameters is 14,789,955, but only 75,267 are trainable, and the rest are used using VGG16 pre-trained values. Two convolutional layers, two max pooling layers, two dropout layers, a flatten layer, and two fully connected layers comprise the CNN used in the study. The two dropouts are key tools for improving the performance and generalization capabilities of CNNs by limiting overfitting, supporting robust feature learning, and swiftly handling vast and complex models [33]. The first dropout is used to minimize noise in the feature maps, while the second is utilized to regulate the whole CNN architecture for improved output. The flattened layer also reshapes the spatially ordered feature maps into a one-dimensional vector that fully linked layers may use to create predictions. The network's architecture requires this flattened representation to connect the convolutional layers to fully connected layers. Finally, the completely linked layers categorize the photos, allowing the breast lumps to be identified. Based on the hierarchical features learnt by the convolutional and pooling layers, the fully connected layer generates final predictions [34]. Figure 3 depicts a graphical representation of the CNN architecture.

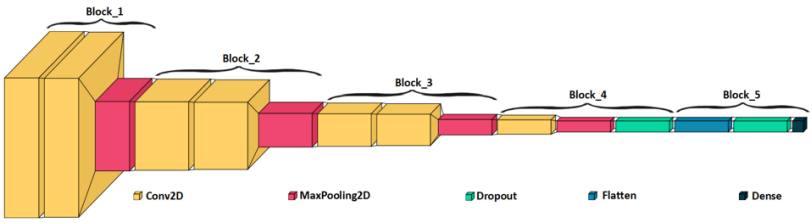

Figure 3 The convolutional neural network architecture

3.4 Model performance evaluation metrics

Model assessment is essential because it assesses a model's performance as a generic model [35]. A performance evaluation is used to assess a model's generalization accuracy

on unseen/out-of-sample data [36]. Accuracy, loss, precision, recall, specificity, and receiver operating characteristic area under the curve (ROCAUC) are the performance assessment measures used to evaluate the models. True positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) are the metrics that are examined.

Accuracy is calculated by dividing the number of properly categorized cases by the total number of occurrences in the dataset [37]. Equation (10) represents accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

In a machine learning model, loss assesses the inaccuracy or difference between anticipated and actual values [38]. Cross-entropy is the most often used loss function in neural networks. Equation (11) represents loss function.

$$loss = \sum_{i=1}^{n} \sum_{j=1}^{m} y_{i,j} \log\left(p_i, j\right) \tag{11}$$

Where n = number of samples, y = true label or ground truth, m = number of classes, j = class iterator, i = samples iterator and p = predicted probability or score.

Precision assesses a model's ability to properly identify positive cases among all favorably predicted instances. It determines the accuracy of a model's positive predictions. The precision is calculated using equation (12).

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

The capacity of a model to accurately identify all positive events in a dataset is measured by recall (Sensitivity). Equation (13) may be used to calculate recall.

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

Specificity measures a model's ability to properly identify negative cases among all negative examples in a dataset. It determines the accuracy of a model's positive predictions. The true negative rate (TNR) is another name for it. Equation (14) provides the definition of specificity.

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

At various thresholds, the ROC curve is used to plot sensitivity versus the false positive rate (1 - specificity). In other words, the ROC curve represents a classification model's performance. A threshold is a value that determines how a model's prediction is classified.

The False Positive Rate (FPR) is the fraction of false positive events that a model mistakenly forecasted as positive out of all real negative instances. Equations (15), (16), and (17) show the formula that makes up the ROC.

$$FPR = 1 - Specificity \tag{15}$$

$$FPR = 1 - \frac{TN}{TN + FP} \tag{16}$$

$$FPR = \frac{FP}{FP + TN} \tag{17}$$

By computing the area under the ROC curve, the Area under the Curve (AUC) score assesses the overall performance of the model. The AUC score is a number between 0 and 1, with a higher value indicating greater model performance [39].

3.5 Summary of the study architectural design

Figure 4 depicts a framework for identifying breast masses used in the present study. Image augmentation, grayscale conversion, denoising method, CLACHE, and feature extraction techniques were performed on the US - BUSI and BUSI data. These data preparation approaches were used to provide clean data that the models could understand during training. The preprocessed data was separated into three sets: training, validation, and testing. The study divided training (80%) and testing (20%), with a portion of the testing set (20%) used to validate the

deep learning models. Deep capsule network, deep Gabor capsule network, and convolutional neural network are the deep learning models used in the study. Deep learning models are used to classify breast masses as benign or malignant in the US - BUSI data, and benign, normal, or malignant in the BUSI data. Several performance assessment criteria are used to evaluate the training results of deep learning models.

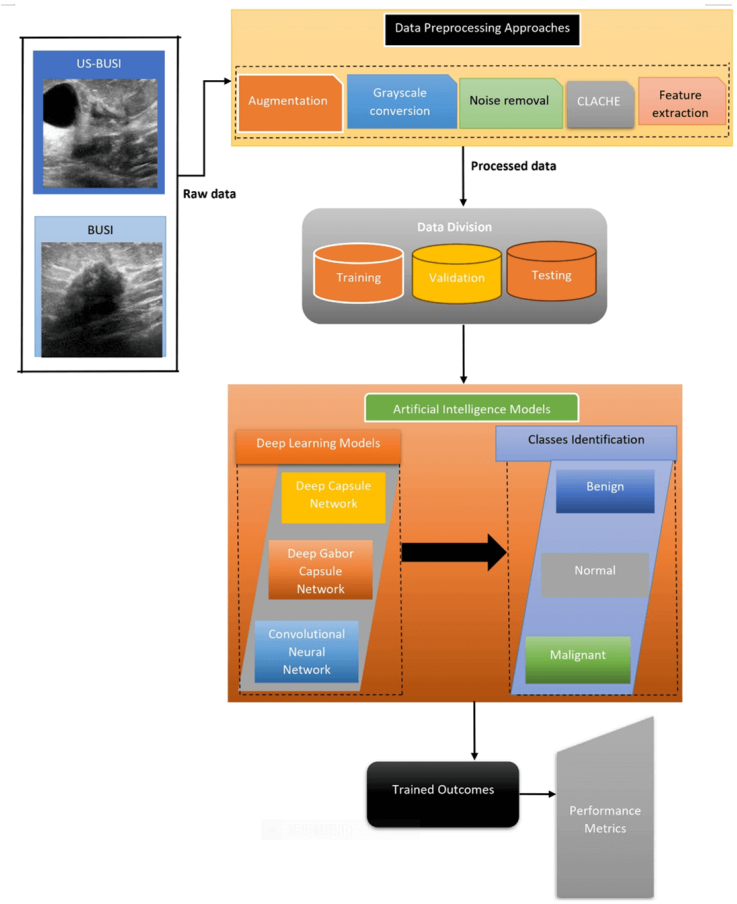

Figure 4 The architectural design of the study

4 Experimental results

This section presents the experimental configuration and results obtained from the models from the two breast ultrasounds datasets.

4.1 Experimental set-up and configurations

As a result, the deep learning models were trained using the Google Colab platform with Python software version 3.11.4 (The Python Software Foundation (PSF), 1209 Orange Street, Wilmington, DE, USA). The Random Access Memory (RAM) installed is 12.0 gigabytes (GB), and the operating system is 64-bit. Table 3 shows the system specifications and experimental settings that were employed to get the results.

Table 3 System configuration and specifications

Product	Specification
Processor	11th Gen Intel(R) Core(TM) i5-1155G7 @2.50Ghz 2.50GHz
Random Access Memory (RAM)	12.0 Gigabyte (GB)
System type	64-bit operating system

4.2 Performance of the DL models on the BUSI dataset

The performance of the deep learning models on the BUSI dataset is presented in this section. The subsections that follow provide in-depth examination of the results.

4.2.1 Interpretability of the BUSI dataset

The class activation mapping (CAM) approach was used to identify the elements of an image that influence the judgment or classification of a deep learning model. CAMs help to explain the classifications or predictions of a deep learning model. This ensures that each model predicts using the relevant picture parts in each class. During this work, the models produced are examined and analyzed using Gradient-weighted Class Activation Mapping (Grad-CAM). Using the gradients of the final convolutional layer, Grad-CAM generates a weighted mixture of feature maps. Figure 5, 6, and 7 depict the Grad-CAM of each model on normal, benign, and malignant images, respectively.

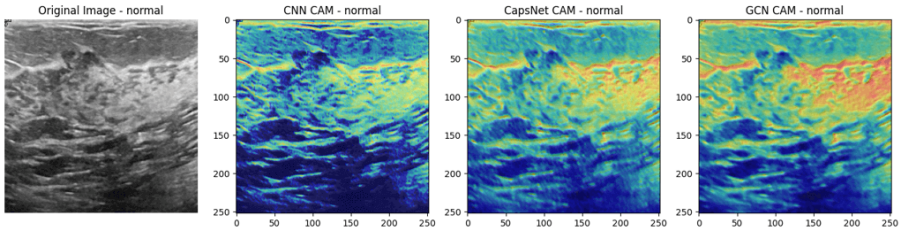

Figure 5 Grad-CAM of the models on the normal image of the BUSI dataset

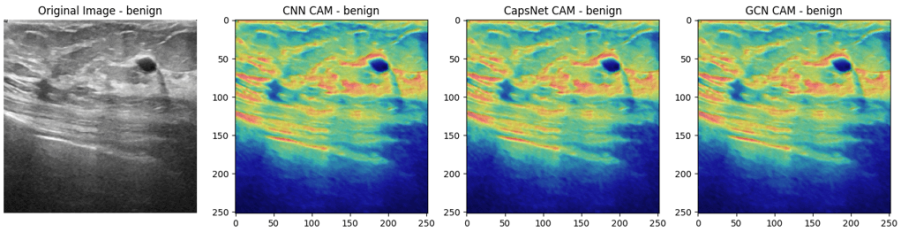

Figure 6 Grad-CAM of the models on the benign image of the BUSI dataset

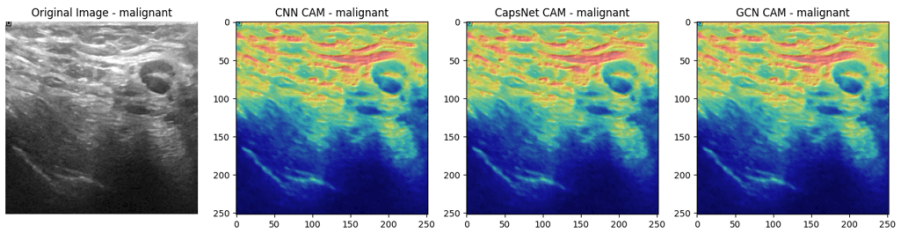

Figure 7 Grad-CAM of the models on the malignant image of the BUSI dataset

4.2.2 The models evaluation on the BUSI dataset

On the BUSI dataset, the capsule neural network, Gabor capsule network, and convolutional neural network were used in this study. It should be noted that the BUSI dataset is multiclassified, with benign, normal, and malignant classifications. Table 4 shows that the deep learning models CapsNet, GCN, and CNN obtained significant results. The accuracy, loss, precision, recall, specificity, and ROC score of each model were calculated. In terms of model accuracy, CapsNet came out on top with a score of 93.62%, followed by CNN with a score of 91.76%. With an accuracy rating of 88.83%, the GCN obtained the lowest value.

Table 4 Performance evaluation of the models on BUSI dataset

Model	Accuracy	Loss	Precision	Recall	Specificity	ROC score
CapsNet	0.9362	0.0083	0.9365	0.9362	0.9191	0.9942
GCN	0.8883	0.0064	0.9182	0.9176	0.9044	0.9538
CNN	0.9176	1.5582	0.9182	0.9176	0.9044	0.9556

Furthermore, the CapsNet outperformed the competition in terms of precision (0.9365), recall (0.9362), and specificity (0.9191). The CNN, the second-best performing model, also produced significant results in terms of performance measures; nevertheless, the CNN fared the worst

in terms of model loss, with 1.5582 as opposed to 0.0064 for GCN and 0.0083 for CapsNet. Despite having the lowest accuracy score, the GCN outperformed all other models in terms of lost value. Surprisingly, the accuracy (0.9182), recall (0.9176), and specificity (0.9044) scores of the GCN and CNN models are the same.

Figure 8, 9, and 10 show the confusion matrices of the CapsNet, GCN, and CNN, respectively, to provide insight into the results of the accuracies. In each of the deep learning models, the algorithms were able to predict equally in terms of the benign, normal, or malignant classifications.

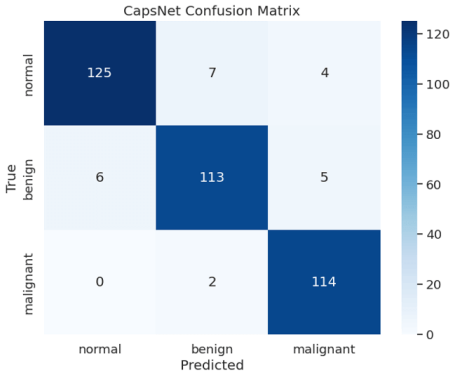

Figure 8 Confusion matrix of the CapsNet model on the BUSI dataset

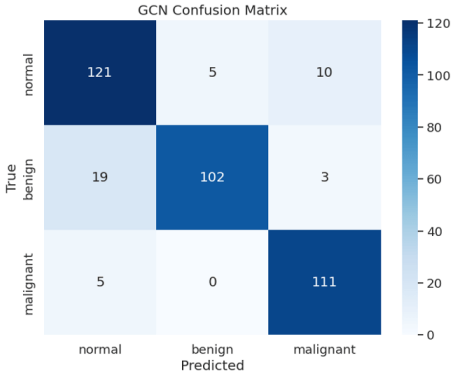

Figure 9 Confusion matrix of the GCN model on the BUSI dataset

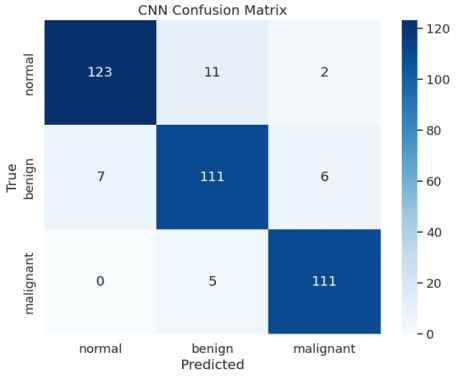

Figure 10 Confusion matrix of the CNN model on the BUSI dataset

Furthermore, Figure 11 shows a graphical depiction of the accuracies attained by the deep learning models used to train the BUSI dataset in this study. As previously stated, CapsNet has the highest accuracy on the BUSI dataset, followed by CNN and GCN in that order. The results of the models used in this work on the BUSI dataset demonstrate the efficacy of dep learning models, particularly the employment of CapsNet and CNN in medical image detection. Having stated that, the GCN also works well in terms of medical image diagnosis, which contributes to the study domain.

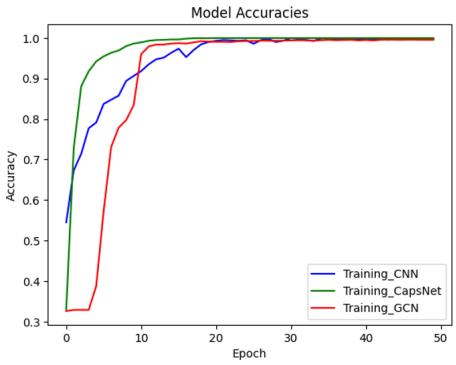

Figure 11 Graphical representation of the models' accuracies on the BUSI dataset

Figure 12 depicts the loss values of the CapsNet (0.0083), GCN (0.0064), and CNN (1.5582). The loss values represent the total of the deep learning models' mistakes. In this study, the significant findings obtained by the loss function on the BUSI dataset measure how effectively (or poorly) the models perform. The results generated by the models used to train the BUSI dataset demonstrate how well the model parameters are designed, giving us a minimum value. It should be noted that the loss function aided in the optimization of the models during training on the BUSI dataset.

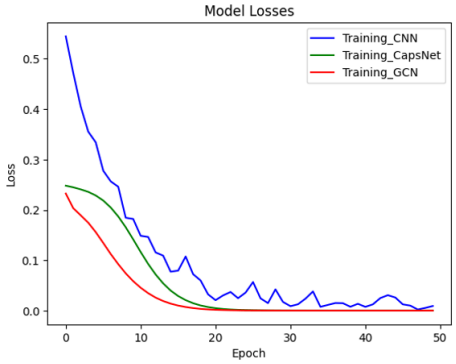

Figure 12 Graphical representation of the loss function of the models on the BUSI dataset

Additionally, the receiver operating characteristics (ROC) of the models were calculated for each of the deep learning models. The Receiver Operating Characteristics (ROC) scores are used to assess the efficacy of the models. The ROC score is a well-known deep learning assessment metric. The capacity of a classification model to differentiate between the breast masses data across numerous classification criteria is measured by the ROC. Figure 13 presents the ROC scores of the models employed in this study.

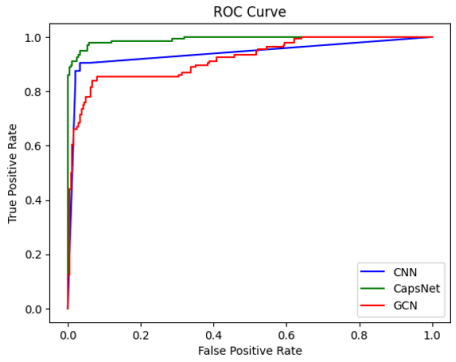

Figure 13 The ROC curves of the models on the BUSI dataset

According to the ROC values shown in Table 4 above and summarized in Figure 13, the CapsNet attained a value of 0.9942, followed by CNN with 0.9556 and the GCN with 0.9538. The ROC plot displays the models' true positive rate (TPR) vs. false positive rate (FPR) at

various categorization levels.

4.3 Performance of the DL models on the US – BUSI dataset

This section also shows how the deep learning models performed on the US - BUSI dataset. The subsections that follow provide in-depth examination of the results.

4.3.1 Interpretability of the US – BUSI dataset

Figure 14 and 15 depict the Grad-CAM of each model used on the US - BUSI dataset. The dataset is classified as benign or malignant. The Grad-CAM is then used to investigate and test the models that were created to train the dataset. Grad-CAM is a key part of model analysis in medical image diagnostics, it should be stressed.

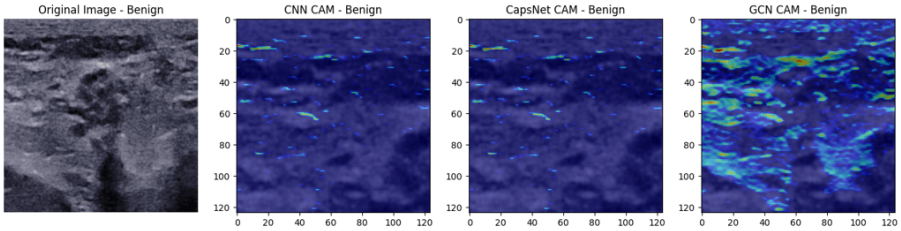

Figure 14 Grad-CAM of the models on the benign image of the US-BUSI dataset

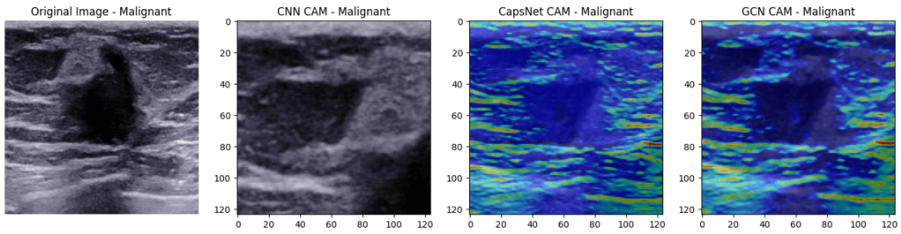

Figure 15 Grad-CAM of the models on the malignant image of the US-BUSI dataset

4.3.2 The models evaluation on the US – BUSI dataset

Table 5 shows the performance of the deep learning models used in this study to train the US - BUSI dataset. The findings show that the models performed well in terms of the assessment criteria of accuracy, loss, precision, recall, specificity, and ROC values. The Gabor capsule network performed the best in training the US - BUSI data, with an accuracy of 97.08%, followed by the capsule neural network with an accuracy of 96.67%, and the convolutional neural network performed the worst, with an accuracy of 95.42%. It should be noted that the dataset used to train the models is binary in nature, with two outcomes, hence benign and malignant.

Table 5 Performance evaluation of the models on US – BUSI dataset

Model	Accuracy	Loss	Precision	Recall	Specificity	ROC score
CapsNet	0.9667	0.0044	0.9440	0.9916	0.9421	0.9964
GCN	0.9708	0.0024	0.9355	0.9748	0.9339	0.9911
CNN	0.9542	0.5601	0.9355	0.9748	0.9339	0.9856

Although the CapsNet did not attain the best accuracy, it outperformed the other models in terms of precision (0.9440), recall (0.9916), and specificity (0.9421). In terms of precision (0.9355), recall (0.9748), and specificity (0.9339), the highest performing model, the GCN, attained the same results as the CNN model. Furthermore, the GCN earned the best error score of 0.0024 on the US - BUSI data, followed by the CapsNet with a score of 0.0044 and the CNN with a value of 0.9542. The loss values (error scores) indicate how successfully the models were trained to anticipate the classification analysis.

Figure 16, 17, and 18 below illustrate the confusion matrices CapsNet, GCN, and CNN, respectively, to help comprehend the accuracies attained by the models used in this investigation. The accuracies of the models, as described by the confusion matrices, reveal an equitable distribution of the in the categorization of the benign and malignant classes of the US - BUSI data.

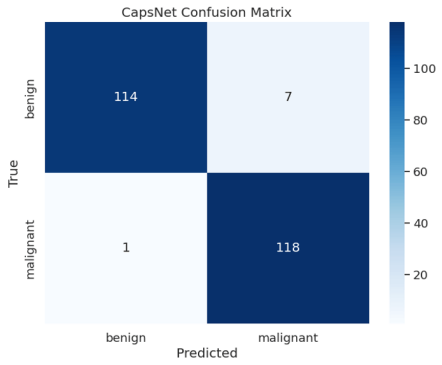

Figure 16 Confusion matrix of the CapsNet model on the US – BUSI dataset

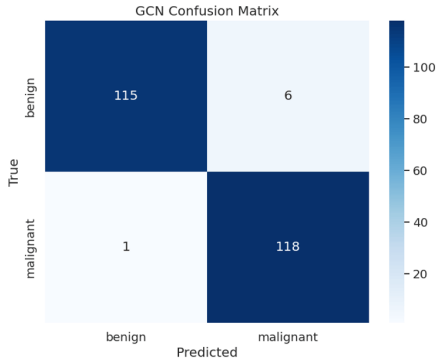

Figure 17 Confusion matrix of the GCN model on the US – BUSI dataset

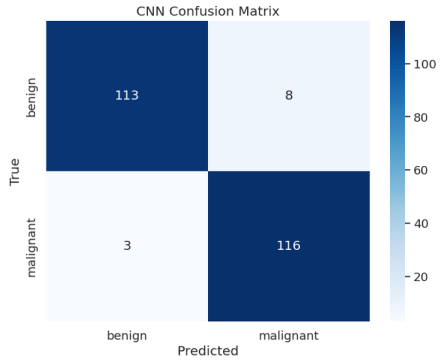

Figure 18 Confusion matrix of the CNN model on the US – BUSI dataset

In addition, Figure 19 shows a graphical depiction of the accuracies attained by the CapsNet, GCN, and CNN models used to train the US - BUSIS dataset. The GCN model got the highest accuracy in training the US - BUSI data, followed by the CapsNet and CNN models. When it comes to binary data analysis, the models produced significant results in terms of accuracies and demonstrated their ability to categorize breast cancer masses.

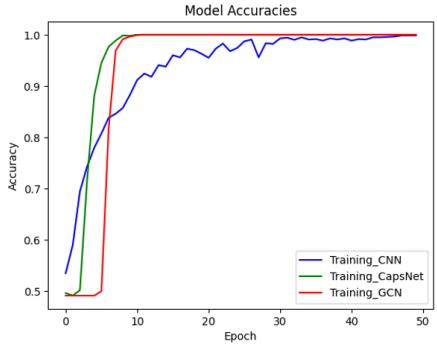

Figure 19 Graphical representation of the models' accuracies on the US – BUSI dataset

Furthermore, loss value scores are required for the summing of mistakes in model training. These error ratings demonstrate how well the models performed on the dataset. The Gabor capsule network earned the best error score of 0.0024, implying that it performed better than the other models. In terms of error score, CapsNet earned a value of 0.0044, whereas CNN achieved a value of 0.5601. Figure 20 shows a summary of the error scores obtained by the deep learning models. The minimum scores show how well the deep learning models performed in the training analysis on US - BUSI data with well-trained configuration parameters.

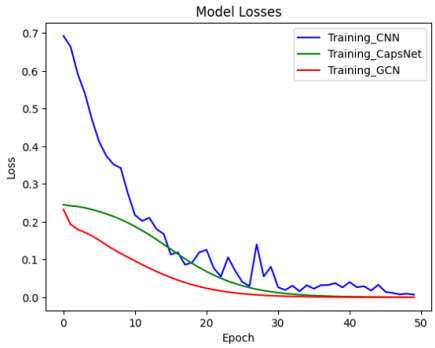

Figure 20 Graphical representation of the loss function of the models on the US – BUSI dataset

In terms of model evaluation, the receiver operating characteristics (ROC) were used to the US - BUSI dataset used to train the deep learning models. These ROC scores are critical for determining the performance of deep learning models, which also serve as a learning assessment tool. Figure 21 shows the ROC scores obtained by the models. CapsNet received the best score of 0.9964, surpassing the GCN, which also received 0.9911. The CNN model, which performed the worst on the US-BUSI data, had a ROC value of 0.9865. The ROC values, which are significant, forecast the true positive rate (TPR) vs. the false positive rate (FPR) of the models.

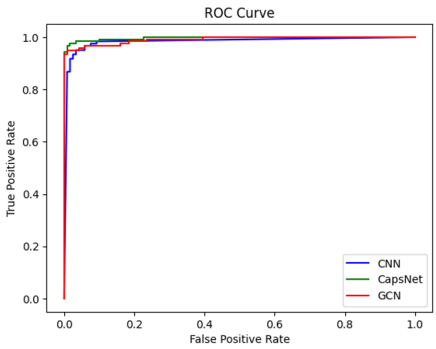

Figure 21 The ROC curves of the models on the US – BUSI dataset

4.4 Summary of the DL models on the BUSI and US – BUSI datasets

The capsule neural network, Gabor capsule network, and convolutional neural network were applied to the BUSI dataset in this work. The BUSI dataset has three classifications: benign, normal, and malignant. This should be acknowledged, thus, a multiclass data. We computed each model's accuracy, loss, precision, recall, specificity, and ROC score. With a model accuracy score of 93.62%, Capsule network led the field, followed by convolutional neural network with a score of 91.76%. The accuracy rating of 88.83% was the lowest for the Gabor capsule network. Additionally, the results demonstrate that the models fared well when evaluated using the binary (benign and malignant) US-BUSI data in terms of accuracy, loss, precision, recall, specificity, and ROC values. With an accuracy of 97.08%, the Gabor capsule network outperformed the other two in training the US - BUSI data: the convolutional neural network did lowest, with an accuracy of 95.42%, and the capsule neural network came in second with a 96.67% accuracy. Given that labeled data might be difficult to collect, medical datasets are frequently small. The annotation of medical images necessitates specialized expertise, which

adds to the time and expense of creating huge datasets. Because of this restriction, the model's capacity to generalize may be hampered, and the likelihood of overfitting may rise. However, in order to train the models and assess their effectiveness, the study used two different datasets of ultrasound images. Furthermore, an unbalanced set of data might result in a model that is biased in favor of the majority class, which lessens the model's ability to identify essential but uncommon situations. The datasets were altered to a maximum value of 500, which had an effect on the decision threshold, in order to guarantee that there was no data imbalance during the augmentation processes. The models in the study use cutting-edge deep learning methods to enhance breast cancer categorization, including CNNs and capsule networks. These models provide higher accuracy by automatically extracting hierarchical features from data, as opposed to more conventional methods like manual feature extraction and standard machine learning. Their performance depends on the amount and variety of the dataset, and for better performance, any overfitting issues are taken into account. The models perform better than certain current approaches, and because they employ two independent datasets, their applicability to other datasets and clinical circumstances is still exceptional. This highlights the need for further validation and improvement of the models in relation to well-established methodologies in the literature.

5 Discussion

Many imaging modalities are being studied for early diagnosis in the ongoing study of breast cancer detection. The BUSI and US-BUSI breast ultrasound datasets were utilized in this study to classify breast cancer using deep learning algorithms. Extensive data preparation and hyperparameter tweaking were performed on these datasets in order to maximize model performance. The study utilized three advanced models: the convolutional neural network (CNN), Gabor capsule network (GCN), and capsule neural network (CapsNet). These models were assessed using metrics including recall, specificity, accuracy, loss, precision, and ROC. The findings highlight how well deep learning algorithms classify breast cancer cases. Notably, the Gabor capsule network performed exceptionally well on the binary dataset with an accuracy of 97.08%, while the capsule neural network earned the maximum accuracy (93.62%) on the multiclass dataset. This result implies that, although the capsule neural network has potential in handling more complicated, multiclass situations, the Gabor capsule network may be especially well-suited for binary classification tasks. These findings support the promise of deep learning models to improve breast cancer diagnosis by demonstrating their versatility in processing various forms of diagnostic medical data. The datasets in the study were trained using state-of-the-art approaches, which is noteworthy. The study highlights these deep learning models' durability and adaptability in medical image analysis by effectively applying them to both binary and multiclass datasets. This dual capacity is important since it shows that these models are not restricted to a particular kind of classification problem, but may be used to efficiently address a variety of diagnostic difficulties. The study's significance and importance within the scientific community are increased by its broad application. This study highlights the advantages of capsule networks and Gabor capsule networks in training medical images for efficient decision-making, despite the widespread usage of CNNs in medical imaging study. Compared to typical CNNs, capsule networks—which are recognized for maintaining spatial hierarchies in the data—and Gabor capsule networks—which use Gabor filters to capture texture and edge information—offer a number of advantages. The models have exhibited exceptional efficacy in this investigation, specifically for the categorization of breast cancer, hence providing compelling evidence for their wider implementation within the domain. The capacity of the study to train on a variety of breast cancer samples from two different datasets adds to its originality. This methodology places the study at the forefront of current research in breast cancer diagnosis while also validating the models' resilience across many data sources. Although deep learning methods have been extensively explored in other research, the combination of CNNs, Gabor capsule networks, and capsule networks in this study offers a novel viewpoint on the categorization of breast masses and offers important new insights for the continuous development of AI in medical diagnostics.

6 Limitations

By utilizing deep learning models, the study generated an improved categorization of breast cancer illness. Nonetheless, it is important to note several limitations. The study's initial set of data were secondary, and further data from different areas will improve the applicability of the model. Moreover, there may be an improvement in the deep learning classifiers' performance.

The study's evaluation of only a few deep learning models was probably constrained by the limitations of the dataset that was made accessible. Due to the restricted investigation, it is possible that models or architectures that may have been more successful were overlooked, which could have improved classification performance. Extensive validation in a variety of clinical contexts is necessary to guarantee the models' applicability in real-time settings. By doing so, it would be possible to close the gap between experimental research and real-world application and guarantee that the models can accurately aid in the detection of breast cancer in a variety of clinical settings. With the current framework, further work may be done to contribute to classifiers that perform better, which will increase detection performance.

7 Conclusions and future outlook

Breast cancer is a prevalent disease where early detection significantly reduces mortality. This study introduced novel deep learning techniques for classifying breast ultrasound images, using binary and multiclass datasets with preprocessing and hyperparameter tuning. The models employed—capsule network, Gabor capsule network, and convolutional neural network—demonstrated strong performance, with the capsule network excelling in binary classification (93.62% accuracy) and the Gabor capsule network leading in multiclass classification (97.08% accuracy). Despite potential improvements from unexplored models, the study offers a unique approach using capsule and Gabor filters, contributing to medical diagnostics. The research aims to expand data availability and deploy the top models in mobile apps for global use.

Conflicts of interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

The authors are grateful to Adwoa Afriyie, popularly called Amanpene, and Malcolm Afrifa for their encouragement and advise throughout the studies.

Data availability

The study utilized publicly available datasets which include the United States (US) - BUSI, which can be found at https://qamebi.com/breast-ultrasound-images-database, and was accessed on October 10, 2023; and the BUSI, which can be found at https://scholar.cu.edu.eg/?q=afahmy/pages/dataset, and was accessed on October 12, 2023.

References

[1] Bokhare A, Jha P. Machine learning models applied in analyzing breast cancer classification accuracy. IAES International Journal of Artificial Intelligence (IJ-AI). 2023, 12(3): 1370. https://doi.org/10.11591/ijai.v12.i3.pp1370-1377

[2] WHO. WHO launches new roadmap on breast cancer, 2023. https://www.who.int

[3] Jabeen K, Khan MA, Alhaisoni M, et al. Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. Sensors. 2022, 22(3): 807. https://doi.org/10.3390/s22030807

[4] de Caldas Filho FL, Soares SCM, Oroski E, et al. Botnet Detection and Mitigation Model for IoT Networks Using Federated Learning. Sensors. 2023, 23(14): 6305. https://doi.org/10.3390/s23146305

[5] Hamzeh O, Alkhateeb A, Zheng J, et al. Prediction of tumor location in prostate cancer tissue using a machine learning system on gene expression data. BMC Bioinformatics. 2020, 21(S2). https://doi.org/10.1186/s12859-020-3345-9

[6] Ding IJ, Zheng NW. CNN Deep Learning with Wavelet Image Fusion of CCD RGB-IR and Depth-Grayscale Sensor Data for Hand Gesture Intention Recognition. Sensors. 2022, 22(3): 803. https://doi.org/10.3390/s22030803

[7] Al-Dhabyani W, Gomaa M, Khaled H, et al. Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images. International Journal of Advanced Computer Science and Applications. 2019, 10(5). https://doi.org/10.14569/ijacsa.2019.0100579

[8] Zhao X, Jiang C. The prediction of distant metastasis risk for male breast cancer patients based on an interpretable machine learning model. BMC Medical Informatics and Decision Making. 2023, 23(1). https://doi.org/10.1186/s12911-023-02166-8

[9] Prodan M, Paraschiv E, Stanciu A. Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection. Applied Sciences. 2023, 13(7): 4272. https://doi.org/10.3390/app13074272

[10] Ardakani AA, Mohammadi A, Faeghi F, et al. Performance evaluation of 67 denoising filters in ultrasound images: A systematic comparison analysis. International Journal of Imaging Systems and Technology. 2023, 33(2): 445-464. https://doi.org/10.1002/ima.22843

[11] Boumaraf S, Liu X, Wan Y, et al. Conventional Machine Learning versus Deep Learning for Magnification Dependent Histopathological Breast Cancer Image Classification: A Comparative Study with Visual Explanation. Diagnostics. 2021, 11(3): 528. https://doi.org/10.3390/diagnostics11030528

[12] Yang X, Fan X, Lin S, et al. Assessment of Lymphovascular Invasion in Breast Cancer Using a Combined MRI Morphological Features, Radiomics, and Deep Learning Approach Based on Dynamic Contrast-Enhanced MRI. Journal of Magnetic Resonance Imaging. 2023, 59(6): 2238-2249. https://doi.org/10.1002/jmri.29060

[13] Sahu A, Das PK, Meher S. High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. Biomedical Signal Processing and Control. 2023, 80: 104292. https://doi.org/10.1016/j.bspc.2022.104292

[14] Sirjani N, Ghelich Oghli M, Kazem Tarzamni M, et al. A novel deep learning model for breast lesion classification using ultrasound Images: A multicenter data evaluation. Physica Medica. 2023, 107: 102560. https://doi.org/10.1016/j.ejmp.2023.102560

[15] Yu S, Jin M, Wen T, et al. Accurate breast cancer diagnosis using a stable feature ranking algorithm. BMC Medical Informatics and Decision Making. 2023, 23(1). https://doi.org/10.1186/s12911-023-02142-2

[16] Sulu SMM, Mukuku O, et al. Women's breast cancer risk factors in Kinshasa, Democratic Republic of the Congo. Current Cancer Reports. 2022, 4(1): 139-143. https://doi.org/10.25082/ccr.2022.01.003

[17] Lee S, Jung H, Park J, et al. Accurate Prediction of Cancer Prognosis by Exploiting Patient-Specific Cancer Driver Genes. International Journal of Molecular Sciences. 2023, 24(7): 6445. https://doi.org/10.3390/ijms24076445

[18] QAMEBI. Breast Ultrasound Images Database - QAMEBI, 2023. https://qamebi.com/breast-ultrasound-images-database

[19] Al-Dhabyani W, Gomaa M, Khaled H, et al. Dataset of breast ultrasound images. Data in Brief. 2020, 28: 104863. https://doi.org/10.1016/j.dib.2019.104863

[20] Afrifa S, Zhang T, Zhao X, et al. Climate change impact assessment on groundwater level changes: A study of hybrid model techniques. IET Signal Processing. 2023, 17(6). https://doi.org/10.1049/sil2.12227

[21] Appiahene P, Varadarajan V, Zhang T, et al. Experiences of sexual minorities on social media: A study of sentiment analysis and machine learning approaches. Journal of Autonomous Intelligence. 2023, 6(2): 623. https://doi.org/10.32629/jai.v6i2.623

[22] Ma R, Ma Y, Li C, et al. Potential mechanism exploration of San Wei Tan Xiang capsule for depression treatment by network pharmacology and molecular docking. Medicine in Novel Technology and Devices. 2022, 16: 100160. https://doi.org/10.1016/j.medntd.2022.100160

[23] Afrifa S, Zhang T, Appiahene P, et al. Mathematical and Machine Learning Models for Groundwater Level Changes: A Systematic Review and Bibliographic Analysis. Future Internet. 2022, 14(9): 259. https://doi.org/10.3390/fi14090259

[24] MISHRA A. Contrast Limited Adaptive Histogram Equalization (CLAHE) Approach for Enhancement of the Microstructures of Friction Stir Welded Joints. Published online June 10, 2021. https://doi.org/10.21203/rs.3.rs-607179/v1

[25] Afrifa S, Varadarajan V. Cyberbullying Detection on Twitter Using Natural Language Processing and Machine Learning Techniques. International Journal of Innovative Technology and Interdisciplinary Sciences. 2022, 5(4): 1069-1080. https://doi.org/10.15157/IJITIS.2022.5.4.1069-1080

[26] Sabour S, Nov CV, Hinton GE. Dynamic Routing Between Capsules. no. Nips, 2017.

[27] Hasani M, Saravi AN, Khotanlou H. An Efficient Approach for Using Expectation Maximization Algorithm in Capsule Networks. 2020 International Conference on Machine Vision and Image Processing (MVIP). Published online February 2020. https://doi.org/10.1109/mvip49855.2020.9116870

[28] Toraman S, Alakus TB, Turkoglu I. Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. Chaos, Solitons & Fractals. 2020, 140: 110122. https://doi.org/10.1016/j.chaos.2020.110122

[29] Alekseev A, Bobe A. GaborNet: Gabor filters with learnable parameters in deep convolutional neural network. 2019 International Conference on Engineering and Telecommunication (EnT). Published online November 2019.
https://doi.org/10.1109/ent47717.2019.9030571

[30] He Y, Song F, Wu W, et al. MultiTrans: Multi-scale feature fusion transformer with transfer learning strategy for multiple organs segmentation of head and neck CT images. Medicine in Novel Technology and Devices. 2023, 18: 100235.
https://doi.org/10.1016/j.medntd.2023.100235

[31] Appiahene P, Arthur EJ, Korankye S, et al. Detection of anemia using conjunctiva images: A smartphone application approach. Medicine in Novel Technology and Devices. 2023, 18: 100237.
https://doi.org/10.1016/j.medntd.2023.100237

[32] Appiahene P, Asare JW, Donkoh ET, et al. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Mining. 2023, 16(1).
https://doi.org/10.1186/s13040-023-00319-z

[33] Podda AS, Balia R, Barra S, et al. Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. Journal of Computational Science. 2022, 63: 101816.
https://doi.org/10.1016/j.jocs.2022.101816

[34] Kaur P, Singh A, Chana I. BSense: A parallel Bayesian hyperparameter optimized Stacked ensemble model for breast cancer survival prediction. Journal of Computational Science. 2022, 60: 101570.
https://doi.org/10.1016/j.jocs.2022.101570

[35] Afrifa S, Varadarajan V, Appiahene P, et al. Ensemble Machine Learning Techniques for Accurate and Efficient Detection of Botnet Attacks in Connected Computers. Eng. 2023, 4(1): 650-664.
https://doi.org/10.3390/eng4010039

[36] Adu WK, Appiahene P, Afrifa S. VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends. Journal of Electrical Systems and Information Technology. 2023, 10(1).
https://doi.org/10.1186/s43067-023-00078-1

[37] Zhang YD, Pan C, Chen X, et al. Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. Journal of Computational Science. 2018, 27: 57-68.
https://doi.org/10.1016/j.jocs.2018.05.005

[38] Yansari RT, Mirzarezaee M, Sadeghi M, et al. A new survival analysis model in adjuvant Tamoxifen-treated breast cancer patients using manifold-based semi-supervised learning. Journal of Computational Science. 2022, 61: 101645.
https://doi.org/10.1016/j.jocs.2022.101645

[39] Phillips CM, Lima EABF, Wu C, et al. Assessing the identifiability of model selection frameworks for the prediction of patient outcomes in the clinical breast cancer setting. Journal of Computational Science. 2023, 69: 102006.
https://doi.org/10.1016/j.jocs.2023.102006

(Edited by Snowy Wang)