Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comprehensive Evaluation of Transfer-CNN based Models for Breast Cancer Detection

Bassam Elzaghmouri [1] , Osman Elwasila [2] , Said Elaiwat [3] , Asef Al-Khateeb [7], Saad Mamoun AbdelRahman [4] , Ahmed Abdelgader Fadol Osman [4] , Lamia Hassan Rahamatalla Mohamed[2],Motasem AbuDawas [5] , Marwan Abu-Zanona [2] , Farah H. Zawaideh [6] ,Ahmad Bany Doumi [7]

[1]*Department of Computer Science, Al-Ahliyya Amman University, Amman, Jordan*
[2] *Department of Management Information Systems, King Faisal University, Kingdom Saudi Arabia*
[3] *Faculty of Architecture and Design, Al-Zaytoonah University of Jordan, Amman, 11733, Jordan*
[4] *Applied College, King Faisal University, Al-Ahsa, 31982, Saudi Arabia*
[5] *Department of Computer Science, College of Science and Information Technology, Irbid National University, Irbid*
[6] *Department of Business Intelligence and Data Analysis, Faculty of Financial Sciences and Business, Irbid National University, Irbid, 21110, Jordan*
[7] *Department of Computer Science, Faculty of Computer Science and Information Technology,Jerash University, Jerash 26150, Jordan*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 30 Jan 2025

Accepted: 08 Feb 2025

ABSTRACT

Breast cancer is among the top causes of death for women globally. Detecting this type of cancer at its early stages can reduce the number of early fatalities greatly, due to the fact that manual and biopsy-dependent breast cancer diagnosis takes so long, an automated technique is important for early cancer diagnosis, these Computer-Aided Diagnosis (CAD) approaches are utilized in our study by harnessing deep learning in order to detect breast cancer. In this work, we present a deep analysis of transfer learning based-models to detect and classify benign from malignant breast tumors. In our study, four well-known CNN architectures are utilized, including ResNet, MobileNet, VGG-16 and EfficientNet, and evaluated using different key performance metrics: accuracy, precision, recall, and F1 score, along with the average training time.

Keywords: Breast Cancer; Convolutional Neural Networks; ResNet50; ResNet101; MobileNet; VGG-16; EfficientNet B7; ZCA.

I. INTRODUCTION

Depending on the statistics conducted by the Centers for Disease Control and Prevention (CDC), breast cancer is the most frequent malignancy among women. The survival rates for breast cancer vary greatly depending on multiple factors [1] . The type of tumor and the stage at which the cancer is diagnosed and detected are important factors to be considered. Breast cancer is thought to be caused by the uncontrolled multiplication of abnormal cells in the breast, commonly starting in the lobules or ducts of the breast [2]. Moreover, malignant cells may develop in the breast's fatty tissue (adipose tissue) or fibrous connective tissue. Uncontrolled cancer cells can migrate to healthy breast tissue and metastasize to lymph nodes under the arms [3].

According to the updated data given by the American Cancer Society, breast cancer has claimed the lives of 41,760 women and more than 500 men. Breast cancer has four main types: benign tumors, intra capsular or IN situ carcinoma and extra capsular or invasive carcinoma [4] . A benign tumor causes minor alterations in the breast tissue but is not conventionally seen as cancerous. Carcinoma cancer in situ involves malignant cells spreading within the duct lobule system of the mammary gland and you can only get it if treated early, thus a treatable disease [4] . This type of cancer can form in all parts of the body, but in special cases it can develop in its most malignant form and deteriorate to invasive carcinoma, which means that it can spread to other organs. These include ultrasonography, computed tomograpy (CT), positron emission tomography (PET), Magnetic Resonance Imaging (MRI), mammography (X-ray) and measurement of breast temperature. The most widely accepted approach for diagnosing breast cancer at the moment is by using pathological examination where tissue samples are stained with Hematoxylin and Eosin as recommended by Hu et al[5].

This is particularly so owing to lack of equipment that can enable screening for the disease and also that which is able to differentiate between cancer, its type and the stage when it is still incurable. This is because the cancer cells that are not diagnosed early and go into other regions in the body are challenging to manage and come with severe impacts. As a result, early detection of these diseases should be made easier and if possible, early intervention effected so that they do not progress to other parts of the body. This is not a thing that has been discussed amongst medical authorities because it is true that where there are symptoms that may be indicative of a certain disease, it is always better that the disease is diagnosed early enough. For instance, the patients who have cancer are not aware that they are affected by this disease at the early stage the disease only enhances to the chronic stage adopting the death. Therefore, timely diagnosis remains as critical in boosting prognosis and increasing the probability of appreciable treatment [6].

The lack of prognostic methods capable of getting early and accurate diagnosis presents difficulties for doctors when developing treatment plans intended to increase a patient's overall survival. Therefore, time is a crucial factor in choosing the best course of action while reducing errors. The time-consuming nature of current breast cancer screening procedures including mammography, ultrasounds, and biopsies has led to the need for computerized and fast diagnosis systems that use machine-learning techniques [7]. These machine learning-based approaches are employed to speed up tumor categorization, improve the precision of cell location, and reduce diagnosis time.

In recent years, Computer-Aided Diagnosis (CAD) and thermography have been reasonably safe and have the potential to develop through technical innovation in the future; it has become more and more popular, especially in the realm of cancer diagnosis. The main goal of this field's current research and investigation is to arrive at a more conclusive tumor outcome that may be used to propose breast cancer screening and generate broad consensus. In a similar context, initiatives are in motion to overcome recently revealed obstacles by tackling the laborious screening procedure, particularly in cases where photo processing is necessary [8].

With image analysis methodologies, digital images are utilized for detection, segmentation, and classification. purposes by employing Deep Learning (DL) methodologies, specifically Convolutional Neural Networks (CNNs). Applications for DL architecture can be found in various fields, including computer vision, speech recognition, object recognition, natural language processing, machine translation, medication design, bioinformatics, medical image analysis, object identification, histopathology diagnosis, and board game software DL offers an improved efficiency and diagnostic accuracy of cancer detection. This study presents a thorough analysis of advanced CNN architectures, including ResNet, MobileNet, VGG-16, and EfficientNet, for the detection of breast cancer. To ensure a comprehensive evaluation, four key performance metrics are considered: accuracy, precision, recall, and F1 score, along with the average training time.

The rest of this paper is organized as follows: Section 1 provides an overview of the related and recent literature interested in detecting Breast Cancer. Section 2 goes over the proposed methodology architecture in detail. The following part presents and compares the experimental results of our model to those of other models. Lastly, section 4 concludes and discusses future work.

II. LITERATURE REVIEW:

This section focuses on the previous research on the use of breast cancer detection as a topic of discussion. Mohammed et al. [7] suggested the following steps toward data preprocessing: Further, a combination of discretization and/or instance resampling and/or missing value removal was applied as the requirement for this situation. The authors considered it appropriate to employ tenfold cross validation in a manner aimed at keeping the validity of the model by splitting the data into ten sets whereby each set undertakes the roles of training and testing of the particular model. They then utilized three machine learning techniques: These classifiers include Naïve Bayes; Decision Tree and we use the J48 algorithm which is an open-source java implementation of the C4. 5 algorithms, and Sequential Minimal Optimization (SMO). Naïve Bayes is based on probabilistic estimation, J48 uses information gain for splitting the attributes and SMO which uses the support vector machine technique where it has a facility to use sequential minimal optimization. They evaluated the performance of the classifiers using five important metrics: It therefore follows that important examiners of the diagnostic tests include the true positive rate, the false positive rate, the ROC curve analysis, and the standard deviation. According to the results, it is observable that when the SRM filter used, it has a positive impact of the classifier because SMO achieved the highest accuracy of 83 percent, which contribute in the highest accuracy of the SRB2 dataset. The NB algorithm achieved relatively good results for the Pima Indians dataset, with an accuracy of 44 % while J48 showed a better performance for the SRB Cancer dataset, with an accuracy of 99. 56 % and 98. 2%, respectively. In a study by Saber et al. [8], a modified DL model was recommended which incorporated the TL technique to enhance the BC site detection and diagnosis abilities. The training and evaluation are done using two approaches: the division of data into a training set and an independent checking set, known as the 80-20 split and cross-validation. They managed to use image analysis-society (MIAS) dataset by employing other pre-trained models that are powered by CNN's including Inception V3, ResNet50, VGG-19, VGG-16 and Inception-V2 ResNet for feature extraction. The experimental results shown in this study can provide promising evidence that can validate the TL using the VGG16 model has a reliable performance in diagnosing BC, with a thorough accuracy of 98, from the proposed 80/20 splitting. As a result, our approach achieved 88%. It also outperformed k-NN with 14%. In addition, achieving the accuracy of 96%, 10-fold cross-validation was 98%. 87% rate.

This section focuses on the previous research on the use of breast cancer detection as a topic of discussion. Mohammed et al. [9] suggested the following steps toward data preprocessing: Further, a combination of discretization and/or instance resampling and/or missing value removal was applied as the requirement for this situation. The authors considered it appropriate to employ tenfold cross validation in a manner aimed at keeping the validity of the model by splitting the data into ten sets whereby each set undertakes the roles of training and testing of the particular model. They then utilized three machine learning techniques: These classifiers include Naïve Bayes; Decision Tree and we use the J48 algorithm which is an open-source java implementation of the C4. 5 algorithms, and Sequential Minimal Optimization (SMO). Naïve Bayes is based on probabilistic estimation, J48 uses information gain for splitting the attributes and SMO which uses the support vector machine technique where it has a facility to use sequential minimal optimization. They evaluated the performance of the classifiers using five important metrics: It therefore follows that important examiners of the diagnostic tests include the true positive rate, the false positive rate, the ROC curve analysis, and the standard deviation. According to the results, it is observable that when the SRM filter used, it has a positive impact of the classifier because SMO achieved the highest accuracy of 83 percent, which contribute in the highest accuracy of the SRB2 dataset. The NB algorithm achieved relatively good results for the Pima Indians dataset, with an accuracy of 44 % while J48 showed a better performance for the SRB Cancer dataset, with an accuracy of 99. 56 % and 98. 2%, respectively. In a study by Saber et al. [10] , a modified DL model was recommended which incorporated the TL technique to enhance the BC site detection and diagnosis abilities. The training and evaluation are done using two approaches: the division of data into a training set and an independent checking set, known as the 80-20 split and cross-validation. They managed to use image analysis-society (MIAS) dataset by employing other pre-trained models that are powered by CNN's including Inception V3, ResNet50, VGG-19, VGG-16 and Inception-V2 ResNet for feature extraction. The experimental results shown in this study can provide promising evidence that can validate the TL using the VGG16 model has a reliable performance in diagnosing BC, with a thorough accuracy of 98, from the proposed 80/20 splitting. As a result, our approach achieved 88%. It also outperformed k-NN with 14%. In addition, achieving the accuracy of 96%, 10-fold cross-validation was 98%. 87% rate.

In their study, Alanazi et al., [11] suggested that the CNN algorithm should be integrated into the existing automated breast cancer detection system to detect invasive ductal carcinoma tissue pattern of whole slide image. (WSI). The developed multiple CNN-based models were then compared with the machine learning (MLs) to identify breast cancer in automated manner. In as much as it records high levels of accuracy than the other conventional ML methods, the proposed model records 9% better accuracy than the 78% recorded by the other conventional ML methods, resulting to 87% accuracy.

Thus, the study by Suh et al. [12] intended to create and compare the effectiveness of a deep-learning model for breast cancer classification in digital mammography with different Bd or breast density levels. In order to successfully identify the malignant lesions, two cnn models were trained where inputs were the mammogram from 1501 patients and controls. As can be learned from the above Tables, the two models generated averagely an AUC of 0. 952 for the cancer diagnosis for DenseNet-169 while EfficientNet-B5 had an average AUC of 0. 954 for cancer diagnosis.

Another study which was conducted by [13] presented an automated approach which is known as Diverse Features-based Breast Cancer Detection (DFeBCD) that is used to classify the mammogram as normal or abnormal. The system employs four unique sets of features: Taxonomic indexes, statistical measures, and local binary patterns are the features to be included in the study. The fourth set is obtained dynamically from mammography and is stored in the second table in database. The authors then used the two classifiers namely, the Support Vector Machine (SVM) and the Emotional Learning-inspired Ensemble Classifier (ELiEC) with a standard IRMA mammography dataset. The numerical results reveal that the system with an automatically conducted feature set through the highway network-based CNN has a better performance than other approaches suggested.

In the same context, another paper used a DL-based methodology that divided the model into three phases: image scaling, deep learning-based breast area segmentation, and a deep learning model for classification this approach was suggested by [14] The breast area is then extracted from thermal pictures using a U-Net network, which mixes contracting and expansive paths to collect properties. This process of segmentation helps to reduce unnecessary components like the neck and shoulders, which cause noise during CNN training. They then applied, a two-class CNN-based deep learning model in order to differentiate between normal and diseased breast tissue. The results indicated that the proposed model achieved 99.33% accuracy, 100% sensitivity, and 98.67% specificity.

The study proposed by Reshma et al., [15] they used an improved Genetic Algorithm, combined with a CNN architecture utilizing weighted feature selection. The relief approach was used to identify an optimal combination of textural, graph, and morphological traits. This enhanced classifier used these combined features as input, resulting in strong performance. Current breast cancer classification strategies based on histopathological images typically select appropriate samples solely based on SVM classifier parameters, ignoring low-density features and limitations in initial training set selection, increasing the likelihood of selecting incorrect prospective samples and compromising system performance. The results validated that the proposed model outperforms conventional techniques with an accuracy of 92.44% rate.

Another study which was conducted by [28]

provides an in-depth exploration of the application of machine learning in medical diagnosis, emphasizing recent advancements in the field. A comprehensive analysis of over 200 research publications, spanning from January 2000 to December 2022, was conducted to evaluate the strengths, limitations, advantages, and challenges associated with the use of machine learning in addressing various medical issues. Based on the challenges identified in this review, the study also proposes potential research directions and future opportunities for further exploration in this domain.

III. PROPOSED METHODOLOGY:

The proposed approach in this study relies pre-trained CNN models, which are fine-tuned specifically for breast cancer detection. Four pre-trained CNN models are applied and evaluated based on four metrics: accuracy, precision, recall, and F1 score. Figure 1 illustrates the overall strategy applied in this work.

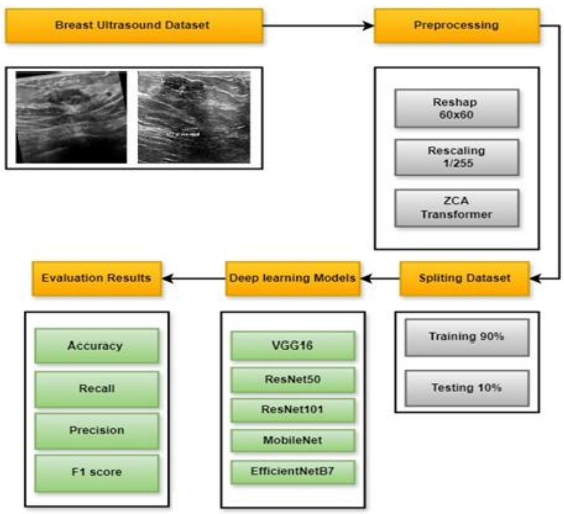

FIGURE 1: [OVERALL RESEARCH DESIGN]

1. *Preprocessing Phase*

Due to the usual variance in resolution and size of the obtained ultrasound breast cancer images, utilizing them directly as training data is impractical. Using high-quality images is time-consuming and may incur substantial computing costs. As a result, all photographs were reduced into 60 × 60-pixel size images. Furthermore, the interrelationship of data inside a picture can inhibit the learning process. To address this issue, we utilized Zero Component Analysis ZCA-based image whitening (Pal & Sudeep, 2016). This technique was harnessed to reduce the correlation between data points while increasing the visibility of important features such as edges and curves. As a result, the Convolutional Neural Network (CNN) can better recognize these traits. In order to get the expected results, the pixel values needed to be altered and normalized to be between 0 and 1, the following equation was implemented:

$$\boldsymbol{X^{\wedge}} = \boldsymbol{X/255} \qquad \textbf{1}$$

Here, X represents normalized image before and normalized image after the normalization process is denoted X^. The image is then normalized by dividing the mean image of the training images to the given image and, subsequently, the mean value of the given image is computed along each feature dimension (pixel position). This mean value is then subtracted from the image, as illustrated below

$$\boldsymbol{X^{-}} = \boldsymbol{X^{\wedge}} - \mu \qquad \textbf{2}$$

where μ is the new reference mean vector of all features for the matrix X^ and X¯ is the mean normalized image. The whitened picture XZCA is determined using the Singular Value Decomposition (SVD) of X¯'s covariance matrix, as follows:

$$\boldsymbol{X_{ZCA}} = \boldsymbol{U.diag(1/\,p\,diag(S)} + \boldsymbol{\varepsilon}\,).\boldsymbol{U^{T}}.\boldsymbol{X^{-}} \qquad \textbf{3}$$

In diagonal method, the entries of the given matrix are replaced with the Diagonal matrix. The points considered are named as U and S for Eigen vector and Eigen value of the covariance matrix's SVD respectively. The other hyperparameter is ε, which determines the whitening coefficient of the input.

Convolutional Neural Networks (CNNs)

In the last few years CNN has emerged as the most attention getting topic because it is used in a number of image processing and Computer Vision applications. CNNs stand for Convolutional Neural Networks which are a type of deep learning models that have been developed for working with data represented in the grid format [16]. The networks below can be used in image processing because they assign a certain relativity learnable weight and bias of an idea in an image to tries to help distinguish between things. At the inception layers in CNN architecture, there are the convolutional layers that were used in creating the spatial representation of the features that are likely to be found in the given input image; therefore, when it comes to detecting and recognizing the features within the image, there is no necessity for prior knowledge of the features that exist in the given image [17]. Since the CNN is designed with shared weights and the local receptive fields, it can effectively identify the spatial and temporal features inherent in the image and represent these features using the image, hence requiring the minimum computations and the minimum parameters. Figure 1 illustrates the overall structure of the CNN with the three main layers: In other words, according to the structural design of the new system, there are three parts of the system: the first part is the input layer, the second is the Feature Learning layer, and the third is the classification layer. The first layer was the preliminary layer and was genuinely a single layer of the internet; it sought for details of the input image grid. The second one is the convolutional layer that introduces features and the other one is the pooling layers. The last layers are mainly computation layers for giving classification results if there is a classification task. It also postulated to be a strength that CNN allows the definition of a series of convolution layers to represent local features for instance horizontal edges while maintaining the structure of images and using comparatively few parameters. Its size and number of filters determine the number of parameters for an input image. Over the last years, enhancements to the CNN design have been made, and the resulting attempts called pre-trained models such as AlexNet [18]., VGG-16 [19]., ResNet [20], and MobileNet [21]. In this work, we exploit the advantages of transfer learning to fine-tune four well-known CNN architectures: MobileNet, VGG16, ResNet50, ResNet101, and EfficientNet

B7, for the task of Breast Cancer Detection. We then perform a comprehensive analysis of their performance across various evaluation metrics.

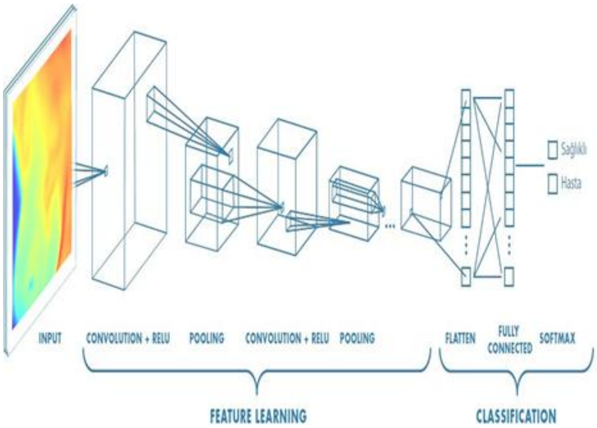

FIGURE 2: [OVERALL ARCHITECTURE OF THE CNN]

2. *MobileNet*

MobileNet as proposed by Howard et al., [21] is a deep learning model developed by the Google Research team based on CNN architecture needed to improve their performance on devices with limited resources while implementing some of the mobile and embedded vision applications. The primary design characteristic of their model is its simplicity; instead of multiple hyperparameters that are present in other CNN models, their model gives less computational time and high accuracy. This was achieved by replacing the ordinary convolution filters with two layers: depthwise convolution and pointwise convolution as provided in the Section 3 above.

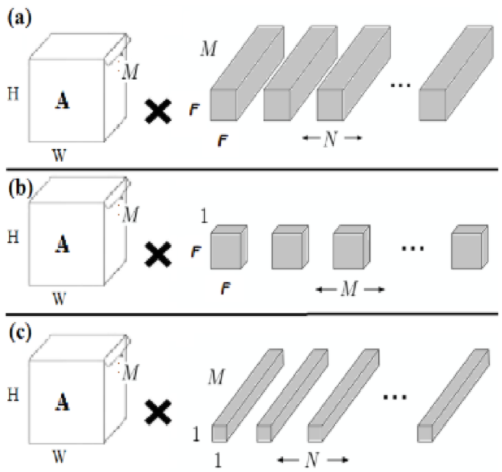

FIGURE 3:[(A): STANDARD CONV FILTERS. (B): DEPTHWISE CONV FILTERS. (C): POINTWISE CONV FILTERS]

3. *VGG16*

The Visual Geometry Group (VGG) at the University of Oxford launched VGG16 in 2014, which is named after its 16-layer architecture [22] This pre-trained CNN was trained using the ImageNet dataset, which included 14 million images [23].giving it a total parameter count of 138 million. As a result, it performs competitively on a variety of picture categorization tasks [24]. VGG16 outperforms AlexNet [25] by using a succession of smaller 3x3 filters instead of larger ones. Our dataset was scaled to 60 by 60 pictures and was fed to VGG16. To ensure compatibility with our binary classification objective of detecting real or fake input images, the last dense layer was activated with a sigmoid function. Figure 4 depicts the workflow of implementing VGG16.

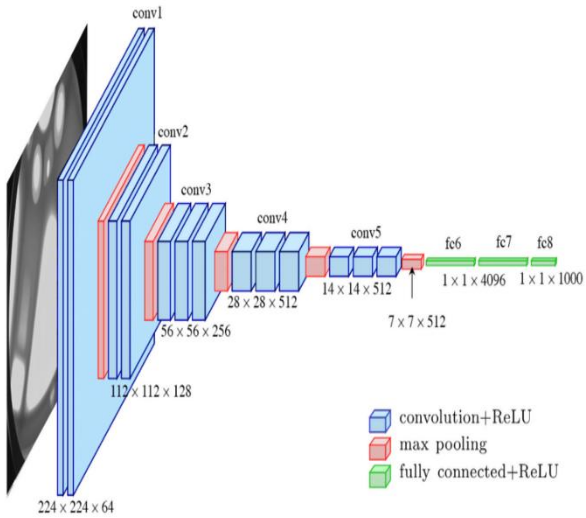

FIGURE 4: [VGG16 DESIGN]

During the training phase, each pixel is preprocessed by subtracting the mean RGB value obtained throughout the training set. The image is then processed using convolutional layers with 3×3 filter sizes. To retain spatial resolution after convolution, both stride and padding are set to one pixel. Following various convolutional layers, max-pooling layers are applied, for a total of five max-pooling layers in the architectural design.

4. ResNet50 and ResNet101

ResNet50, which stands for Residual Network 50, is another type of convolutional neural network. It was first launched in [26] and, as the name implies, has 50 layers: 48 convolutional layers, one Max Pool layer, and one Average Pool layer. ResNet50 uses a bottleneck architecture for its construction components, as seen in Figure 5. This architecture aids in reducing parameter count and matrix multiplications, allowing for faster training inside each layer.

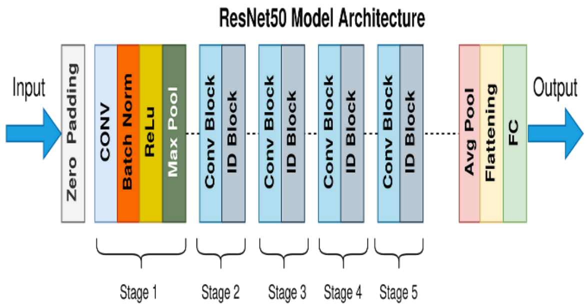

FIGURE 5: [RESNET50 DESIGN]

ResNet101 and ResNet50 have the same architecture; the former, however, has 101 layers of residual connections that are useful in addressing vanishing gradient issue and ease the training of deep networks. ResNet101's structure is similar to ResNet50, using bottleneck building blocks to eliminate the increase in computational load while maintaining its effectiveness. Its signature enables it to detect multiple and various features of images, including how suited it is for image classification or object recognition.

5. EfficientNetB7

The bigger network of the EfficientNet family was developed to strike a workable balance of size and performance. In October last year, Tan and Le introduced compound scaling as a way of optimizing depth, width, and resolution at once. The essence of its design involves layering blocks that use depthwise separable convo- lutions and squeeze-and-excitation networks, which enable efficient feature extraction using smaller parameters. The EfficientNet introduces two rules: The first is that the scaled models shall incorporate the convolutional implementation of the baseline network. Secondly, all layers must be scaled in the same way and ratio the preposition is used to show possession as seen by the following translations: - These two rules are mathematically presented as follows:

$$\boldsymbol{N(d, w, r)} = \sum{}_{1..s} \boldsymbol{F_i^{d,L_i}} \left(\boldsymbol{X(r.H_i, r.W_i, w.C_i)}\right) \quad 4$$

The scaling coefficients w, d, and r alter the network's width, depth, and resolution, respectively. The basic network has fixed values for the parameters, F_i H_i, L_i, W_i, and C_i. The authors offer a simple but effective scaling strategy that uses a compound coefficient to uniformly scale the network's breadth, depth, and resolution systematically. Figure 6 illustrates the overall design of this pre-trained model.

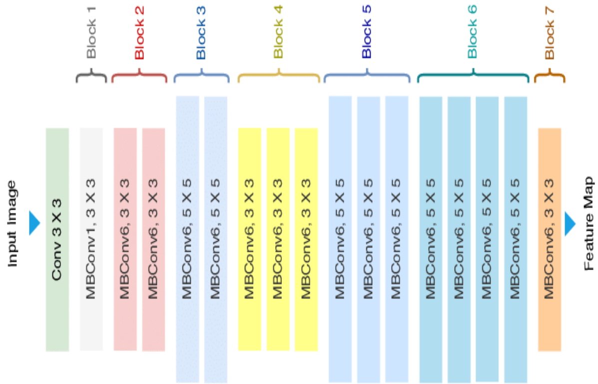

FIGURE 6: [EFFICIENTNETB7 DESIGN]

IV. RESULTS AND EXPERIMENTS:

This section discusses the performance evaluation of fine-tuned CNN structures for breast cancer detection, using a benchmark dataset called Breast Ultrasound Images Dataset [27]. The results are thoroughly analyzed based on four different metrics: accuracy, precision, recall, and F1 score, to assess the performance of each model.

A. Dataset

The benchmark Breast ultrasound data was collected from women aged 25 to 75 years were gathered in 2018 to serve as baseline data. The dataset contains 600 female patients and 780 images, each averaging 500x500 pixels in size. The images are in PNG format, and related ground truth images are included alongside the originals. The photos are categorized into benign, and malignant data [27]. Figure 7 shows a sample of the dataset images.

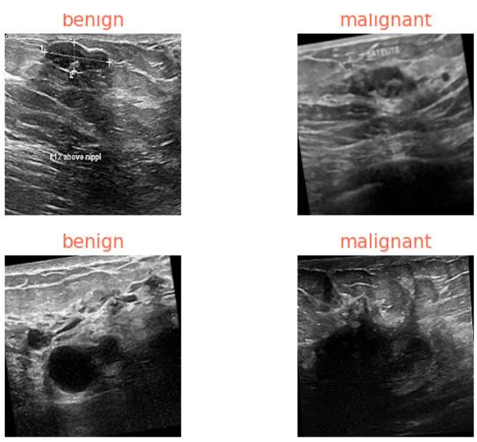

FIGURE 7: DATA SAMPLE

A. *Experiment configurations*

In the model training phase, we used 90% of the data samples to train each CNN model we involved in this work and tested their performance using the remaining 10% of the dataset. In order to normalize the input images, we standardized their sizes to 60 × 60 × 3, where the first two dimensions reflect the width and height, while the third dimension represents the color channels of red, green, and blue (RGB). Throughout the training phase, Batch Gradient Descent was used in every trial. Furthermore, the number of epochs used was set to 10, balancing proper training against the risk of overfitting. The tables ranging from 1 to 6 provide further information about each CNN model's hyperparameters we used in this study.

TABLE 1: [HYPERPARAMETERS OF THE PROPOSED CNN-BASED MODEL]

Hyper Parameters	**Proposed CNN Model**	
Convolution Layers	4	
Number of neurons for each convolution layer	First Layer	32
	Second Layer	64
	Third Layer	128
	Fourth Layer	256
Max Pooling Layers	4	
Fully Connected Layer	1	
Dropout Layer value	NaN	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Batch Normalization	5	
Kernal window size	(3,3)	(3,3)
Pool size	(3,3)	(2,2)
Trainable params	391938	
Non-trainable params	1472	
Total Parameters	39341 0	
Call Backs	1	
epochs	10	
Batch size	10	
Optimizer	Adam	

TABLE 2: [HYPERPARAMETERS OF RESNET50]

Hyper Parameters	Restnet50 Model	
Dropout Layer value	NaN	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Trainable params	16386	
Non-trainable params	23587712	
Total Parameters	23604098	
Call Backs	1	
epochs	10	
Batch size	10	
Optimizer	Adam	

TABLE 3: [HYPERPARAMETERS OF VGG16]

Hyper Parameters	VGG16 Model	
Convolution Layers	13	
Number of neurons for each convolution layer	First Layer	64
	Second Layer	64
	Third Layer	128
	Fourth Layer	128
	Fifth Layer	256
	Sixth Layer	256
	Seventh Layer	256
	Eighth Layer	512
	ninth Layer	512
	Tenth Layer	512
	eleventh Layer	512
	twelfth Layer	512
	Thirteenth Layer	512
Max Pooling Layers	5	
Fully connected Layer	1	
Dropout Layer value	NaN	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Kernal window size	(3,3)	
Pool size	(3,3)	
Trainable params	1026	
Non-trainable params	14714688	
Total Parametes	14715714	

Call Backs	1
epochs	10
Batch size	10
Optimizer	Adam

TABLE 4 : [HYPERPARAMETERS OF RESNET101]

Hyper Parameters	Restnet101 Model	
Dropout Layer value	NaN	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Trainable params	16386	
Non-trainable params	42658176	
Total Parametes	42674562	
Call Backs	1	
epochs	10	
Batch size	10	
Optimizer	Adam	

TABLE 5: HYPERPARAMETERS OF MOBILE NET

Hyper Parameters	Mobile Net Model	
Conv2d Layer	14	
Depthwise Conv2D	13	
Number of neurons for each convolution layer	First Layer	32
	Second Layer	64
	Third Layer	128
	Forth Layer	256
	Fifth Layer	512
	Sixth Layer	1024
Global Average Pooling	1	
Fully connected Layer	2	
Dropout Layer value	1	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Kernal window size	(3,3)	
Pool size	(7,7)	
Trainable params	10242	

Non-trainable params	2257984
Total parameters	2268226
Call Backs	1
Epochs	10
Batch size	10
Optimizer	Adam

TABLE 6: [HYPERPARAMETERS OF EFFICIENTNETB7]

Hyper Parameters	efficientnetB7 Model	
Dropout Layer value	NaN	
Activation function	Relu	Sigmoid
Train Split	0.9	
Test Split	0.1	
Trainable params	20482	
Non-trainable params	64097687	
Total Parameters	64118169	
Call Backs	1	
epochs	10	
Batch size	10	
Optimizer	Adam	

B. Performance Evaluation

Four key metrics, including accuracy, precision, recall, and F1 score, were utilized to evaluate and analyze the performance of fine-tuned models.

Equation 5 represents the accuracy, where TP and TN denote the number of True Positive and True Negative classifications, respectively, while FP and FN indicate the False Positive and False Negative values. Table 7 reports the accuracy results for all models. We can notice that VGG16 model achieved the highest accuracy with a **97%** rate compared to other models. In contrast, efficientnetB7 achieved the lowest accuracy rate of 89.5%. The chart representation of our gained accuracies is illustrated in Figure 8.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}} \qquad \mathbf{5}$$

TABLE7: [TESTING AND TRAINING ACCURACY RATES OF ALL MODELS]

Type Of Structure	Accuracy	Time
VGG16	97%	4s 9ms/step
Resnet50	92.4%	5s 2ms/step
Resnet101	90%	8s 31ms/step
Mobilenet	92%	6s 9ms/step
efficientnetB7	89.5%	12s 51ms/step

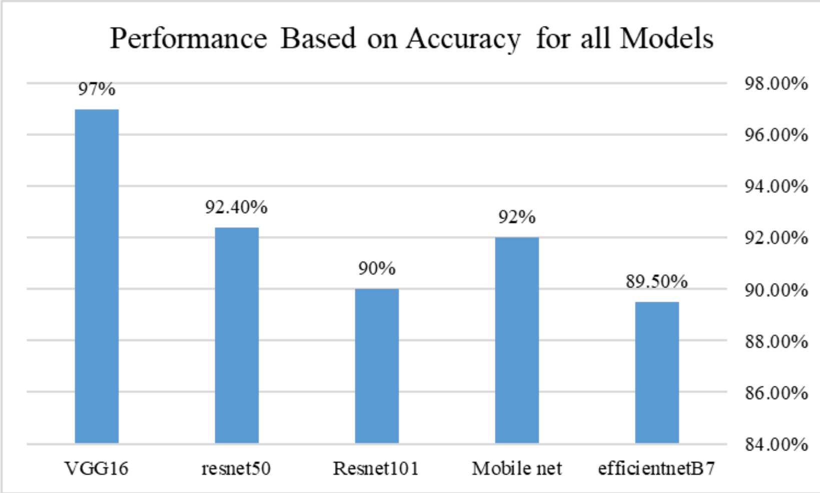

FIGURE 9: [PERFORMANCE BASED ON ACCURACY FOR ALL MODELS]

The Recall is given in Equation 6, T P and F N represent the number of True Positives and False Negatives, respectively. The performance of each model in terms of recall is listed in Figure 9.

$$\textbf{Recall} = \frac{TP}{TP + FN} \qquad (6)$$

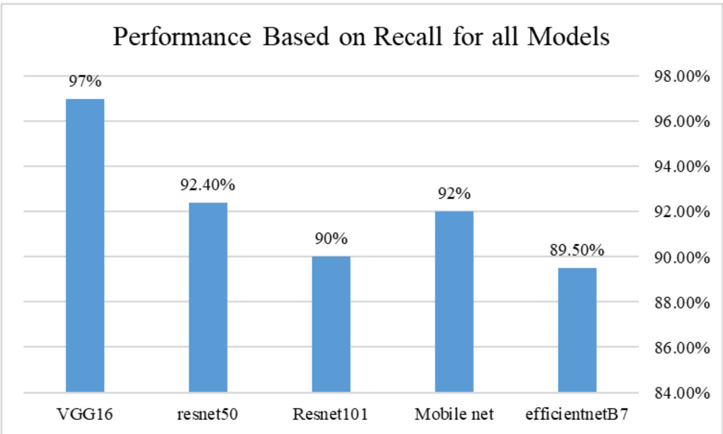

FIGURE 8: [PERFORMANCE BASED ON RECALL FOR ALL MODELS]

VGG16 model achieved the highest Recall rate among other pre-trained models with a 97.0% rate. In contrast, efficientnetB7 achieved the lowest Recall rate.

The Precision and F1 measures were represented in Equations 7 and 8, respectively. where TP and FN represent the number of True Positive and False Negative classifications, respectively. Figures 10 and 11 report the performance of each model based on precision and F1Score, receptively.

$$\textbf{Precision} = \frac{TP}{TP + FP} \qquad (7)$$

$$F1 = 2 \times \frac{TP}{(TP + FP + FN)} \qquad (8)$$

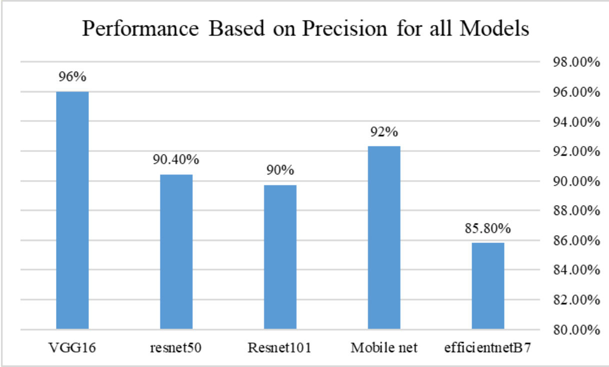

FIGURE 10: [PERFORMANCE BASED ON PRECISION FOR ALL MODELS]

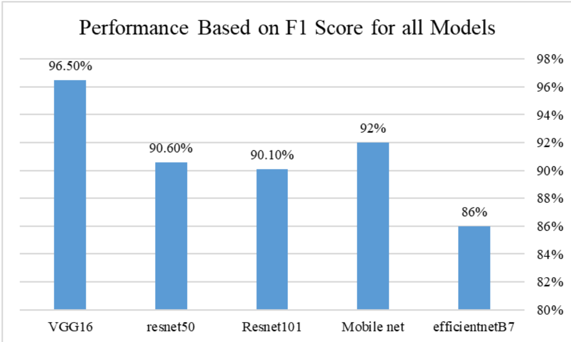

FIGURE 11: [PERFORMANCE BASED ON F1 FOR ALL MODELS]

Overall, we can conclude that VGG 16 model outperforms other pre-trained models in terms Accuracy, Recall, Precision, and F1-Score of 97.0%, 97.0%, 96.0%, and 96.5%, respectively. It is worth noting that efficientnetB7 achieved the lowest performance across all experiments when compared to other pre-trained models.

V. CONCLUSION:

In this work, we presented comprehensive analysis of transfer learning based-models for the task of breast cancer detection. In our analysis, we utilized four well-known CNN architectures including ResNet (50 and 100), MobileNet, VGG-16 and EfficientNet . The evaluation was conducted using a benchmark dataset called Breast Ultrasound Images Dataset. Each model was evaluated using four key metrics: accuracy, precision, recall, and F1 score. The result showed superior performance for VGG-16 model, with accuracy rate 97%, over other models followed by Resnet50. In contrast, reported the lowest performance.

In future work, we are planning to use object segmentation approaches like R-CNN, to improve the detection accuracy, and reduce the computational time.

REFERENCES:

[1] Abdollahi, J., Davari, N., Panahi, Y., & Gardaneh, M. (2022). Detection of Metastatic Breast Cancer from Whole-Slide Pathology Images Using an Ensemble Deep-Learning Method: Detection of Breast Cancer using Deep-Learning. Archives of Breast Cancer, 364–376. https://doi.org/10.32768/abc.202293364-376

[2] Adinegoro, A., Sutapa, G., Gunawan, A., Anggarani, N., Suardana, P., & Kasmawan, I. G. (2023). Classification and Segmentation of Brain Tumor Using EfficientNet-B7 and U-Net. Asian Journal of Research in Computer Science, 15, 1–9. https://doi.org/10.9734/ajrcos/2023/v15i3320

[3] Alanazi, S. A., Kamruzzaman, M. M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting Breast Cancer Detection Using Convolutional Neural Network. Journal of Healthcare Engineering, 2021, 1–11. https://doi.org/10.1155/2021/5528622

[4] Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. Data in Brief, 28, 104863. https://doi.org/10.1016/j.dib.2019.104863

[5] Alharbi, O. (2021). A Deep Learning Approach Combining CNN and Bi-LSTM with SVM Classifier for Arabic Sentiment Analysis. International Journal of Advanced Computer Science and Applications, 12, 2021. https://doi.org/10.14569/IJACSA.2021.0120618

[6] Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. Computer Methods and Programs in Biomedicine, 223, 106951. https://doi.org/10.1016/j.cmpb.2022.106951

[7] Allugunti, V. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. 49–56.

[8] Alom, Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., & Nasrin, M. S. (n.d.). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.

[9] Chouhan, N., Khan, A., Shah, J., Hussnain, M., & Khan, M. (2021). Deep Convolutional Neural Network and Emotional Learning based Breast Cancer Detection using Digital Mammography. Computers in Biology and Medicine, 132, 104318. https://doi.org/10.1016/j.compbiomed.2021.104318

[10] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[11] Hamed, G., Marey, M., Amin, S., & Tolba, M. (2020). Deep Learning in Breast Cancer Detection and Classification (pp. 322–333). https://doi.org/10.1007/978-3-030-44289-7_30

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

[13] Hossain, N., Bhuiyan, M., Tumpa, Z., & Hossain, S. (2020). Sentiment Analysis of Restaurant Reviews using Combined CNN-LSTM. https://doi.org/10.1109/ICCCNT49239.2020.9225328

[14] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (arXiv:1704.04861). arXiv. https://doi.org/10.48550/arXiv.1704.04861

[15] Hu, Q., Whitney, H. M., & Giger, M. L. (2020). A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. Scientific Reports, 10(1), 10536. https://doi.org/10.1038/s41598-020-67441-4

[16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[17] Lotter, W., Diab, A. R., Haslam, B., Kim, J. G., Grisot, G., Wu, E., Wu, K., Onieva, J. O., Boyer, Y., Boxerman, J. L., Wang, M., Bandler, M., Vijayaraghavan, G. R., & Sorensen, A. G. (2021). Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nature Medicine, 27(2), 244–249. https://doi.org/10.1038/s41591-020-01174-9

[18] M. Al-Tam, R., & M. Narangale, S. (2021). Breast Cancer Detection and Diagnosis Using Machine Learning: A Survey. JOURNAL OF SCIENTIFIC RESEARCH, 65(05), 265–285. https://doi.org/10.37398/JSR.2021.650532

[19] Mohamed, E. A., Rashed, E. A., Gaber, T., & Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. PLOS ONE, 17(1), e0262349. https://doi.org/10.1371/journal.pone.0262349

[20] Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020). Analysis of Breast Cancer Detection Using Different Machine Learning Techniques. Data Mining and Big Data, 1234, 108–117. https://doi.org/10.1007/978-981-15-7205-0_10

[21] Nassif, A. B., Talib, M. A., Nasir, Q., Afadar, Y., & Elgendy, O. (2022). Breast cancer detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine, 127, 102276. https://doi.org/10.1016/j.artmed.2022.102276

[22] Pal, K. K., & Sudeep, K. S. (2016). Preprocessing for image classification by convolutional neural networks. 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 1778–1781. https://doi.org/10.1109/RTEICT.2016.7808140

[23] Rajinikanth, V., Joseph Raj, A. N., Thanaraj, K. P., & Naik, G. R. (2020). A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection. Applied Sciences, 10(10), Article 10. https://doi.org/10.3390/app10103429

[24] Reshma, V. K., Arya, N., Ahmad, S. S., Wattar, I., Mekala, S., Joshi, S., & Krah, D. (2022). [Retracted] Detection of Breast Cancer Using Histopathological Image Classification Dataset with Deep Learning Techniques. BioMed Research International, 2022, e8363850. https://doi.org/10.1155/2022/8363850

[25] Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. IEEE Access, 9, 71194–71209. https://doi.org/10.1109/ACCESS.2021.3079204

[26] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556

[27] Suh, Y. J., Jung, J., & Cho, B.-J. (2020). Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning. Journal of Personalized Medicine, 10(4), Article 4. https://doi.org/10.3390/jpm10040211

[28] Mohammad Shehab, Laith Abualigah, Qusai Shambour, MuhannadA. Abu-Hashem , Mohd Khaled Yousef Shambour, Ahmed Izzat Alsalibi, Amir H. Gandomi (2024).Machine learning in medical applications: A review of state-of-the-art methods, Computers in Biology and Medicine Volume 145,https://doi.org/10.1016/j.compbiomed.2022.105458