

# Breast Cancer Detection System using Deep Learning Based on Fusion Features and Statistical Operations

Suleyman A. AlShowarah

Faculty of Information Technology, Mutah University, Karak, Jordan

**Abstract**—Breast cancer is considered as the second cause of death for women. The earlier is diagnosed, the easier the patients can be recovered. The need for studies to detect this kind of cancer easily and accurately came from the growing rate of infected patents by breast cancer exponentially. This study is conducted to investigate the use of deep-learning model for breast cancer detecting using the technique VGG-19 and ultrasound images. Two layers of VGG19 structure were used: (i.e. fc6 and fc7). Based on these two layers (fc6 and fc7), new datasets were created, which are named as statistical operations. These datasets will be employed as input for the following Machine Learning classifiers: K-Nearest Neighbors, Random Forest, Naïve Bayes and Decision Tree. Data augmentation was considered to increase the dataset size for better learning of CNN. Random Forest achieved high accuracy (88.63), precision (0.88), recall (0.88) and F-Measure (0.88). The results of the classification accuracy in the three scenarios are slightly similar; this proves that the breast cancer can be detected even if the size of data in the training dataset was minimal.

**Keywords**—Breast cancer detection; breast cancer classification; deep learning; vgg-19; breast tumor

## I. INTRODUCTION

Cancer is considered as an uncontrolled growth of cells in human body. Breast cancer is one type of cancers, which considered as the second cause of death for women. It is known for patients and doctors, the earlier the cancer is diagnosed and detected, the easier the patients can be recovered. Because of the growing rate of infected by breast cancer exponentially [1]; there is need for studies to detect this kind of cancer easily and accurately. It is considered as a motivation for such study in that the diagnosed people in this kind of cancers is growing day by day. For example, in the US [2], most women are diagnosed with breast cancer compared to any other type of cancer, except for skin cancer. This cancer affects one in three of new female annually. In a year 2023, the estimated diagnosed women in US with invasive breast cancer to be 297,790, and with non-invasive breast cancer will be 55,720 [2]. Although these statistical numbers reflect what is obtainable in most advanced economies, and it was illustrated by studies that about 58% of deaths occur because of this cancer in less advanced countries. The high death rate from breast cancer is because of the lack of early detection, as more than 33% are for population aged 30-49 and 81% for 30-59 years old are diagnosed for this cancer [3], [4], [5].

In order for early detection and saving lives of breast cancer, a mammography was developed by scientists with some limitations of its functions. Despite of that, some of studies showed a reduction in death rate of about 40% after a mammogram screening [6, 7]. About 15 of the 1,000 women seen with mammography are recommended for a biopsy, and about 13 women of these biopsies show a false positive results (not present) [8]. A major limitation of mammographic screening was highlighted by C.K. et al. [9]: breast cancers of prognostic significance are not diagnosed.

Several of strategies were implemented to enhance the performance of screening mammography: including double checking and screening at annual intervals [10], apply two views for each breast [11], and make a comparison with previous mammograms [12]. The serious features can be detected by radiologists for each scan such as architectural distortions, micro-calcifications, and asymmetries as cancer biomarkers or cancer risk. Detecting these serious features manually leads to additional costs and will let the radiologists pay more efforts for mammography [13]. One of systems emerged in the 1990s named as Computer-aided detection (CAD), this is to detect and then classify breast cancers in mammograms automatically. But still the performance of such these traditional systems has not improve screening process significantly, this is mainly due to their lack specificity [14,15]. Specificity relates to how abnormalities can be discriminated by algorithms when screening, which differs from how it is diagnosis; it employs causal inference as to the origin of the abnormality. However, detecting anomalies in screening mammographs is important in the diagnosis.

Recently, the researchers in [16] reported that novel algorithms based on CNN can be used to improve the performance of screening mammography and also to increase the efficiency of mammography professionals. In this matter, some of researchers developed different CNNs-based algorithms for automated mammographic analysis purposes [17].

This study aims to detect and classify the images of breast cancer using deep-learning, which can be employed as system used to help doctors and radiologists in their diagnosis automatically. To achieve this aim, it will be conducted based on CNN using VGG-19. The pre-trained technique VGG-19 is used to achieve high accuracy by finding distinctive details features of image [18] [19]. The two layers of the VGG19 structure were used (i.e. layer 6; which called fc6, and layer 7; which called fc7), and each contains 4096 features. Also, more feature vectors were created from (fc6 and fc7), which named

as statistical operations. Statistical operations are used to generate more datasets using Average (Avg), Minimum (Min), Maximum (Max), and fusion between fc6 and fc7. All aforementioned datasets will be used as input for the processes of classification using different algorithms (i.e. K-Nearest Neighbors (KNN), Random Forest (RF), Naïve Bayes (NB) and Decision Tree (DT)).

The results illustrated that Random Forest algorithm achieved high accuracy (88.63), precision (0.88), recall (0.88) and F-Measure (0.88) for fc7 of second scenario. The results were slightly similar; this approves that these features can provide a better accuracy when used in detection studies. Also, the results of the classification accuracy in the three scenarios are slightly similar; this proves that the breast cancer can be detected even if the number of images in the training dataset were minimal.

The motivation of conducting this study is represented by:

1) The literature need for researches of detecting breast cancer using CNN with new model like VGG-19 based on the two layers: fc6 and fc7.

2) To the best of my knowledge, I could not find any research in the literature performed based on statistical operations for detecting breast cancer using VGG-19.

3) Based on features that were extracted by VGG-19, it can be a contribution for this study by providing the literature with a differentiation between the results of different aforementioned classifiers and with different three scenarios.

This research is designed into five sections. The overview of related studies in literature is introduced in Section II. Section III discusses the methodology for the proposed model and the experiment design. Then, the experimental results and discussion are discussed in Section IV. Finally, Section V presents the conclusion.

## II. LITERATURE REVIEW

Several of studies showed several of automated, and computer vision approaches to classify breast cancer- based images [20, 21]. Some of them have focused on segmentation process, and then features were extracted from images [22]. While in some other studies followed the pre-processing steps for better feature extraction, this is to improve the contrast in the images and then to detect the infected part of image [23]. For example, the most important and the first of the pre-processing steps in the mammogram analysis are applying the segmentation for the infected region, which allows focusing on region of interest in the images. The researchers in [24] applied the technique called texture filter in the segmentation of the breast region.

Lastly, in this matter, there was a study conducted by de Vos, et al. [25] who implemented DL for extraction features for region of interest from cancer images. In their study, they used three techniques of convolutional neural network (ConvNet) to detect and to extract features from a 3D image, which are: the presence of the anatomical structure of interest in the following: 1) axial, 2) coronal, 3) sagittal slices. The method of their localization was compared to the manual

method using the distances between the centroids and the walls with an automatically and manually defined reference frame. Many other researches have adapted a pure deep learning based on its layers for extraction features [26–30], and also using one of most interesting methods such a high pass isotropic filter [31].

Image cropping aims to enhance image quality by removing distracting content/also adding aesthetics, which are mainly categorized-based on that. Different methods are available to achieve such this task. Often, these methods can apply techniques like: machine learning, deep learning, segmentation, saliency-based, and sparse coding. For example, the study conducted by Mishra et al. [32] used ML radiology using classification pipeline. They segmented the region of interest, and then extracted the useful features. Their study was performed on the dataset: (BUSI), and the results showed improving in classification accuracy. While the study conducted by Byra [33] used DL for the classifying the cancer parts from images. The transfer learning (TL) was used and then deep representation scaling (DRS) layers were added between the blocks of pre-trained CNNs to enhance the provided information. In order to analyze these parts classification, the enhancement was only for the parameters of the deep representation scaling layers during training, this is to enhance the pre-trained CNNs, which was much better compared to other techniques.

Some of researchers in [34] developed algorithm: Dilated Semantic Segmentation Network (Di-CNN) and then they used it to detect and classify the breast cancer. The pre-trained DenseNet201 deep model was used in their work and then trained using TL that was used for feature extraction. In addition, they applied a 24-layer CNN and fusion features in their work. The results of the fusion process have improved the classification accuracy in the detection process.

Ahmed et al. [35] used patch selection to classify breast tissue based images using TL. The features were extracted using CNN to discriminate patches, which are an input for an Efficient-Net architecture that is considered as an architecture of CNN and employed for scaling technique that scales all dimensions of: width, depth, and resolution based on a compound coefficient; these input were trained on the dataset: ImageNet. The classifier support vector machine was used for classifying features that were extracted from the Efficient-Net. The results showed that the suggested model achieved better results compared to the standard methods.

The use of DL has outperformed most recent methods. A good example of this, it is the study in [35], which built based on the geometric properties of the edge features to extract the abnormal patches structures in the expected regions. For example, the researchers in [36] conducted a study and the features were extracted from the image using CNN (DenseNet); which are then provided in fully connected layers (FC) for classifying the benign and cancerous cells of breast cancer image. However, the researchers in [37] presented deep learning methods for detecting and classifying models. Deep learning can be used perfectly for computer vision problems, especially image optimization and interpretation. This has led to a wave of pioneering applications of medical imaging, and

available databases of image have presented the growth of DL algorithms aimed to detect cancer images.

Other researchers in [38] conducted a study to extract the most important and useful information using DL including convolution layers for breast cancer detection. They showed that the features extracted by DL models are better in term of accuracy than traditional and manual methods. DL methods showed that it has ability to detect pathological forms of cancer that were previously thought to be difficult to diagnose using conventional and manual methods. The researchers in [39] presented DL-based techniques for detection breast cancer images. In their work, a dataset was published containing canine mammary tumor (CMT). Also, (VGG-16) was used to investigate the performance of hybrid frameworks using different algorithms on CMT and other datasets of breast cancer.

The works that have been done in the literature on DL using CNN among others, have provided an insight to the researchers on how an automatic representation method without supervised descriptors in extracting, i.e. independent of any human intervention that could influence these representations [40].

In fact, Deep CNNs usually have too large number of parameters, so that it is not reasonably trained without a very large dataset. Moreover, medical datasets are usually not large enough to adequately train a deep CNN model from scratch. Thus, transfer learning in deep learning was explored to be used in the medical imaging to solve such problem. So, the transfer learning transfers knowledge between large source and small target domains [41], which can be done by using pre-training a CNN model with the source of images, then re-training parts of the model with the target images. In [42], the researchers used CNN AlexNet to detect the images of breast cancer from dataset named: BreakHis [43]. Their results showed that the classification accuracy is 79.85%.

While the researchers Han et al. in [44] have proposed a framework for breast cancer multi-classification using class structure-based deep CNN model on the dataset named: BreakHis. The results showed 93.2% of classification accuracy.

The researchers Nuh et al. in [45] have conducted a study to discriminate between samples infected and non-infected breast cancers images based on CNN using different spatial patches. The results showed for window sizes: 5x5 and 7x7 are 86.91% and 86.17% respectively.

The researchers in [46] have presented a CNN classifier for the visual analysis of area of cancer in images of malignant breast cancer. The results showed a higher performance for their proposed classifier compared to random forest classifier: 84.23% classification accuracy. It was approved by results of Hafemann et al. [47] that used a CNN; It showed better results compared to traditional approach that always needs huge efforts and effective expert in the field of knowledge [48].

### III. METHODOLOGIES

This section presents the dataset design, and the experimental setup model.

#### A. Database Design

The Dataset of Breast Ultrasound Images (Dataset BUSI) is used in this research and can be obtained online [49]. The dataset consists of 780 images with size 500x500 pixels; including the segmentation masks that refers to 600 patients. The dataset consists of three classes: normal, malignant, and benign. The whole dataset was divided into training and test dataset. However, this is not enough as dataset to train data using the model of deep learning; therefore, a data augmentation step is achieved to increase the dataset size for better learning of CNN. These implemented steps are achieved multiple times until the size of dataset of each class has reached 5872.

#### B. Experimental Setup Model

The experiment of this study was designed based on three scenarios: 1) 50% for training and 50% for testing, 2) 70% for training and 30% for testing, and 3) 80% for training and 20% for testing. So, for each scenario - the experimental setup for the proposed model is displayed in Fig. 1 and consists of set of steps, as follows:

Step 1: The MATLAB is used for automatically extracting feature form images based on Pre-trained VGG-19. The outputs are two datasets for fc6 and fc7. Each of fc6 and fc7 contains 4096 features. These datasets will be used in the step 3.

Step 2: Creating a new dataset from step 1 by performing the statistical operations (i.e. Avg, Min, Max, and fusion of fc6 and fc7). These datasets will be used also in step 3.

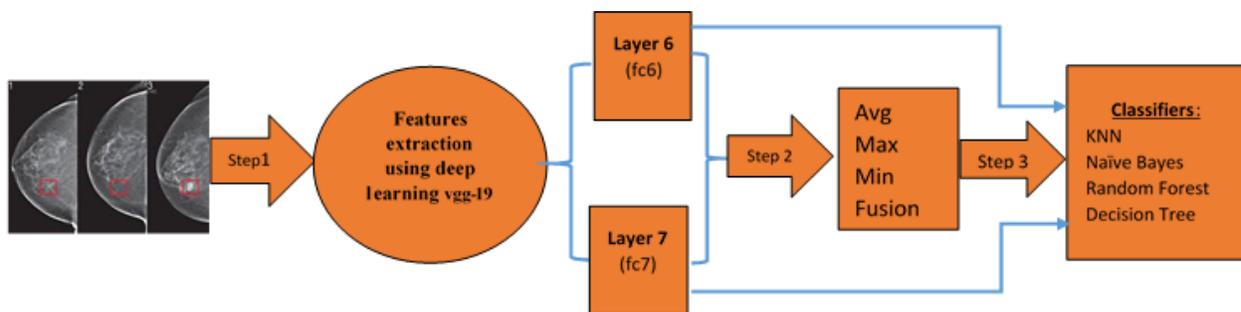


Fig. 1. Proposed experimental setup model.

The explanation for the statistical operations is in the following: 1) Max: It is the largest value of fc6 and fc7. 2) Min: It is minimum value of fc6 and fc7. 3) Avg: It is the average for (fc6 and fc7). 4) Fusion between fc6 and fc7: It is used to combine the first group of fc6: (4096) next to the second group of fc7: (4096), and thus that will create dataset, which contains 8192 features.

Step 3: The aforementioned classifiers will be applied on the datasets that obtained from step 1 and step 2 to provide the results represented by Accuracy, Recall, F-measure, Precision, and duration time.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The study is designed for three scenarios. In each scenario, the results of the evaluation of performance for the breast cancer images is represented by: Accuracy (Acc), Recall, F-measure, Precision (Pre) and duration training time for each classifier. The evaluation of performance is applied on the following four classifiers; KNN, NB, RF, and DT. The results of each scenario are illustrated in the following subsections.

##### A. First Scenario

This scenario was designed based on the percentage of 50% for training and 50% for testing. Its aim is to investigate the influence of 50% of the data size in the training dataset on the classification accuracy. In this scenario, three results are obtained. First, results for original fc6 and original fc7 datasets. Second, results for the statistical operations. Finally, results for fusion feature between (original fc6 and original fc7) dataset.

1) *Results for original fc6 and original fc7 datasets separately:* Table I and Table II show the results of the Fully Connected: fc6 feature vector dataset and fc7 feature vector dataset which were obtained from CNN outputs based on using different classifiers.

The results showed that Random Forest outperforms other algorithms in classifying breast cancer if it is malignant or benign for fc6 and fc7 which are (88.24) and (88.35) of classification accuracy respectively. However, the training time required to conduct the experiment shows that DT required more time (i.e. 16.03) compared to others, but KNN required little time (i.e. 0s). The reason for that, because there is no training model; the comparison occurs directly between the test row with other training rows (examples), and this explains the slow in time for testing, especially if there is large size of data (examples) for the training [52-53]. This results match with the results in [50] in term of that RF outperforms other classifiers used in their study. Their study compared Random Forest with Support Vector Machine, DT, Multilayer Perceptron, and KNN.

To the best of my knowledge, there was no the same study achieved to detect breast cancer based on the same proposed model in using deep learning with these four classifiers together (i.e. KNN, NB, RF, and DT), and also using statistical operations (i.e. Avg, Max, Min, and Fusion of fc6 and fc7), or using the three scenarios.

2) *Results for the statistical operations:* The results of three datasets that created for statistical operations (i.e. Avg, Max, and Min) are presented in this section.

Tables III to V show results of the three statistical operations, whereas Random Forest algorithm outperforms other algorithms in classifying breast cancer if it is malignant or benign for Avg, Max, and Min, which are (88.28), (87.87), and (88.07) respectively. Despite of the Random Forest have showed an acceptable classification accuracy that outperformed other classifiers, it showed also an acceptable training time. While the training time required to conduct the experiment, the classifier Decision Tree required more time (i.e. 14.38s) compared to others, but KNN required little time (i.e. 0s), this is for the same reason mentioned in Section A of First Scenario. This results match with the results in [50] as mentioned in Section A of First scenario in the field of conducting study on RF, but not in using the statistical operation or three scenarios.

3) *Results for fusion feature between (original fc6 and original fc7) dataset:* This dataset is created by fusion of fc6 (4096 feature) and fc7 (4096 feature). The total feature will be 8192. The results in Table VI showed that Random Forest algorithm (88.38) outperformed other algorithms in classifying breast cancer if it is benign or malignant. The second-high accuracy is for KNN (86.78). While the training time for the classifiers, DT required large time (28.79s), but KNNs required less time (0s) when compared to other classifiers.

The summary of first scenario is that the results for all datasets used in first scenario are slightly similar to each other; especially for RF and KNN. This means all the features used in the study can have the same influence on the classification accuracy.

##### B. Second Scenario

This scenario was designed based on the percentage of 70% for training and 30% for testing. Its aim is to investigate the influence of 70% data size in the training dataset on the classification accuracy. In this scenario, three results are obtained. First, results for original fc6 and original fc7 datasets. Second, results for the statistical operations. Finally, results for fusion feature between (original fc6 and original fc7) dataset.

1) *Results for original fc6 and original fc7 datasets separately:* Table I and Table II show the results of the Fully Connected: fc6 feature vector dataset and fc7 feature vector dataset which were obtained from CNN outputs based on using different classifiers.

The results showed that Random Forest outperformed other algorithms in classifying breast cancer if it is benign or malignant for fc6 and fc7, which are (88.24) and (88.63) of classification accuracy respectively. While the training time for the classifier Decision Tree required more time (i.e. 31.09s) for fc7 compared to others. But KNN required little time (i.e. 0s), this is for the same reason mentioned in Section A of First Scenario. This results match with the results in [50] as mentioned in Section A of First scenario in the field of

conducting study on RF, but not in using the statistical operation or three scenarios.

2) *Results for the statistical operations:* The results of three datasets that created for statistical operations (i.e. Avg, Max, and Min) are presented in this section.

Tables from III to V show results of the three datasets (avg, max, and min). The classifier Random Forest outperformed other algorithms in classifying breast cancer for Avg, Max, and Min, which are (88.51), (87.95), and (88.36) respectively.

Despite of the RF have showed an acceptable classification accuracy that outperformed all other classifiers, and the training time was also an acceptable compared with others. While the training time for the classifier Decision Tree required more time (i.e. 81.13s) compared to others. But KNN required little time (i.e. 0s), as this explained earlier in Section A of First Scenario. This results match with the results in [50] as mentioned in Section A of First scenario in the field of conducting study on RF, but not in using the statistical operation or three scenarios.

TABLE I. DETECTION RESULTS OF FC6 FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)
	<i>Fc6 (First Scenario)</i>					<i>Fc6 (Second Scenario)</i>					<i>Fc6 (Third Scenario)</i>				
<b>KNN</b>	86.03	0.86	0.86	0.86	0.01	86.39	0.86	0.86	0.86	0	85.71	0.86	0.85	0.85	0.01
<b>NB</b>	78.54	0.82	0.78	0.79	1.29	77.98	0.82	0.78	0.78	1.76	78.71	0.82	0.78	0.79	3.27
<b>RF</b>	88.24	0.88	0.88	0.87	3.11	88.24	0.88	0.88	0.87	4.38	88.10	0.88	0.88	0.87	9.38
<b>DT</b>	81.19	0.81	0.81	0.812	13.71	81.36	0.81	0.81	0.81	20.18	82.07	0.82	0.82	0.82	57.63

TABLE II. DETECTION RESULTS OF FC7 FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)
	<i>Fc7 (First Scenario)</i>					<i>Fc7 (Second Scenario)</i>					<i>Fc7 (Third Scenario)</i>				
<b>KNN</b>	86.75	0.87	0.86	0.86	0	85.86	0.86	0.85	0.86	0	86.54	0.86	0.86	0.86	0.01
<b>NB</b>	82.93	0.84	0.82	0.83	1.19	83.52	0.84	0.83	0.83	1.98	83.39	0.84	0.83	0.83	2.2
<b>RF</b>	88.35	0.88	0.88	0.87	2.97	88.63	0.88	0.88	0.88	4.52	88.35	0.88	0.88	0.87	5.58
<b>DT</b>	81.77	0.81	0.81	0.81	16.03	82.65	0.82	0.82	0.82	31.09	81.60	0.81	0.81	0.81	31.52

TABLE III. DETECTION RESULTS OF AVERAGE (AVG) FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)
	<i>Avg (First Scenario)</i>					<i>Avg (Second Scenario)</i>					<i>Avg (Third Scenario)</i>				
<b>KNN</b>	86.06	0.86	0.86	0.86	0	86.64	0.87	0.86	0.86	0	85.86	0.86	0.85	0.86	0
<b>NB</b>	79.63	0.83	0.79	0.80	1.14	79.22	0.82	0.79	0.80	1.98	80.54	0.83	0.80	0.81	3.17
<b>RF</b>	88.28	0.88	0.88	0.87	3.01	88.51	0.88	0.88	0.88	8.77	88.35	0.88	0.88	0.87	10.4
<b>DT</b>	81.02	0.81	0.81	0.811	15	82.99	0.82	0.83	0.82	81.13	82.65	0.82	0.82	0.82	43.23

TABLE IV. DETECTION RESULTS OF MAXIMUM (MAX) FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)	Acc	Pre	Recall	F- measure	Time (s)
	<i>Max (First Scenario)</i>					<i>Max (Second Scenario)</i>					<i>Max (Third Scenario)</i>				
<b>KNN</b>	85.35	0.85	0.85	0.85	0	85.69	0.86	0.85	0.85	0	85.48	0.85	0.85	0.85	0
<b>NB</b>	82.45	0.84	0.82	0.83	1.24	82.74	0.84	0.82	0.83	2.8	83.07	0.84	0.83	0.83	2.99
<b>RF</b>	87.87	0.87	0.87	0.87	3.1	87.95	0.87	0.88	0.87	8.14	88.20	0.88	0.88	0.87	8.53
<b>DT</b>	81.36	0.81	0.81	0.81	14.38	81.58	0.81	0.81	0.81	31.7	82.12	0.82	0.82	0.82	35.71

TABLE V. DETECTION RESULTS OF MINIMUM (MIN) FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F-measure	Time (s)	Acc	Pre	Recall	F-measure	Time (s)	Acc	Pre	Recall	F-measure	Time (s)
	<i>Min (First Scenario)</i>					<i>Min (Second Scenario)</i>					<i>Min (Third Scenario)</i>				
<b>KNN</b>	86.23	0.86	0.86	0.86	0	86.32	0.86	0.86	0.86	0	86.67	0.86	0.86	0.86	0.02
<b>NB</b>	76.19	0.81	0.76	0.77	1.23	75.45	0.81	0.75	0.76	2.82	76.64	0.81	0.76	0.77	5.02
<b>RF</b>	88.07	0.88	0.88	0.87	3.13	88.36	0.88	0.88	0.87	7.76	87.88	0.87	0.87	0.87	8.61
<b>DT</b>	83.10	0.83	0.83	0.83	14.45	82.31	0.82	0.82	0.82	38.24	81.03	0.81	0.81	0.81	51.01

TABLE VI. DETECTION RESULTS OF FUSION FEATURE VECTOR

Algorithm	Acc	Pre	Recall	F-measure	Time (s)	Acc	Pre	Recall	F-measure	Time (s)	Acc	Pre	Recall	F-measure	Time (s)
	<i>Fusion (First Scenario)</i>					<i>Fusion (Second Scenario)</i>					<i>Fusion (Third Scenario)</i>				
<b>KNN</b>	86.78	0.87	0.86	0.86	0	86.00	0.86	0.86	0.86	0	86.63	0.86	0.86	0.86	0
<b>NB</b>	81.77	0.83	0.81	0.82	2.42	82.18	0.84	0.82	0.82	3.84	82.75	0.84	0.82	0.83	6.51
<b>RF</b>	88.38	0.88	0.88	0.87	4.31	88.61	0.88	0.88	0.88	10.7	88.250	0.88	0.88	0.87	12.37
<b>DT</b>	81.23	0.81	0.81	0.81	28.79	83.09	0.82	0.83	0.83	73.95	82.48	0.82	0.82	0.82	91.39

TABLE VII. COMPARISON BETWEEN PROPOSED MODEL VS. MOST RELATED WORKS

Authors' References	Description of method/ technique	Dataset Size	Performance metric/s (value/s)
[54]	It consists of two stages: 2D detection and 3D aggregation. In 2D, the magnitude and orientation field map generated using Gabor filters. Then, features extracted using Faster-RCNN to determine the spatial pattern of AD. In 3D, the fusion strategy for regions is used for 2D into 3D. They used Soft classifier.	265 DBT image. They were collected from 68 patients	Mean True positive fraction (MTPF) of 50 +_ 0.04 ACC=90%
[55]	They used VGG-16 with progressive fine-tuning to evaluate its performance on AD detection. Then, the results were compared with a custom CNN architecture that were trained from scratch.	280 image	AUC=0.89
[56]	Breast masses detected system was developed based on texture description, spectral clustering, and support vector machine (SVM). ROIs were segmented using spectral clustering relaying on texture. Then, the optimal features were submitted to SVM. They used SVM classifier.	-	Accuracy 90%
[51]	Different of CNN models are used for feature extraction: VGG16, InceptionV3, and ResNet50. Data were enlarged up to 400X to increase the accuracy. They used Softmax classifier.	7,909 images (Number of patients 82: 24 benign patients and 58 malignant patients).	98.26%.
The proposed model	The following are used in this study: - CNN (VGG-19 technique): features extracted from two layers: (fc6+fc7). - Fusion for two layers: (fc6+fc7) - Apply classifiers on: Fc6, fc7, statistical operations (avg, max, min, and fusion). - Three scenarios in the training dataset. - Used the following classifiers: K-Nearest Neighbors, Random Forest, Naïve Bayes, and Decision Tree.	5872	Accuracy (88.63%)

3) Results for fusion feature between (original fc6 and original fc7) dataset: This dataset is created by fusion of fc6 (4096 feature) and fc7 (4096 feature). The total feature will be 8192. The results in Table VI showed that there was not big difference between them, but Random Forest algorithm achieved higher accuracy compared to other algorithms in classifying breast cancer if it is benign or malignant for (88.61) of accuracy. The second-high accuracy is for KNN (86). While for the training time for classifiers; Decision Tree required more time (73.95s), but KNNs required little time (0s) compared to other.

The summary of second scenario is that results for all datasets used in the second scenario are slightly similar to each other; especially for RF and KNN. This means that all the features used in the study can have the same influence on the classification accuracy

### C. Third Scenario

This scenario was built based on the percentage of 80% for training and 20% for test data set. Its aim is to investigate the influence of 80% data size in the training dataset on the classification accuracy. In this scenario, three results are obtained. First, results for original fc6 and original fc7 datasets.

Second, results for the statistical operations. Finally, results for fusion feature between (original fc6 and original fc7) dataset.

1) *Results for original fc6 and original fc7 datasets separately:* Table I and Table II show the results of the Fully Connected: fc6 feature vector dataset and fc7 feature vector dataset which were obtained from CNN outputs based on using different classifiers.

The results showed that Random Forest outperforms other algorithms in classifying breast cancer if it is benign or malignant for fc6 and fc7, which are (88.10) and (88.35) of classification accuracy. While the training time for the classifier Decision Tree required more time (i.e. 57.63s) for fc7 compared to others, but KNN required little time (i.e. 0s). This results match with the results in [50], as discussed in Section A of First and Second Scenarios.

2) *Results for the statistical operations:* The results of three datasets that created for statistical operations (i.e. Avg, Max, and Min) are presented in this section.

The results of statistical operations are presented in Tables from III to V. The results show the Random Forest algorithm outperformed other algorithms in classifying breast cancer for Avg, Max, and Min, which are (88.35), (88.20), and (87.88) respectively. In addition to these results of having acceptable accuracy for the Random Forest, the confusion matrix; F-measure, recall, and precision are scored high values among all other classifiers.

The training time for RF was also in an acceptable compared with others. While the training time for the classifier Decision Tree required more time (i.e. 51.01s) compared to others, but KNN required little time (i.e. 0.02s). This results match with the results in [50] in the field of classification accuracy results for RF, as they did not conduct their study in using the statistical operation or three scenarios.

3) *Results for fusion feature between (original fc6 and original fc7) dataset:* This dataset is created by fusion of fc6 (4096 feature) and fc7 (4096 feature). The total feature will be 8192. The results in Table VI showed that there were not big difference between the values of accuracy. Therefore, Random Forest algorithm outperformed other algorithms in classifying breast cancer if it is malignant or benign in fusion dataset that achieved (88.25) of accuracy. The second-high accuracy is for KNN (86.63). While the training time for the classifiers Decision Tree required large time (91.39s), and KNNs required less time (0s) compared to other classifiers.

The summary results for the three scenarios show that the classifier Forest showed better classification accuracy compared to other classifiers. In general, the required time to achieve was high for the case; fusion dataset in the third scenario, this is because the size of data is huge in the training, which was 80% and required a lot of time. In term of investigating the influence of the three scenarios on the classification accuracy, the results have approved that the detection for breast cancer can be achieved with almost similar classification accuracy even if the training dataset was minimal.

Table VII shows the comparison between our proposed model with most similar studies conducted for breast cancer detection. The proposed model can be considered as one of the interesting study, for number of reasons, mentioned below.

Some others of previous studies were performed on small dataset compared to proposal model.

Some others of previous studies were performed on large data size in training dataset (examples) compared to proposal model which contained a few images (examples). Three scenarios were considered in the proposal model. It was approved by our proposal model that using few number of images in the training dataset usually leads to low classification accuracy compared to use large number of images.

Some of previous data enlarge the data up to 400 times categories to increase the data size such as in [51]. This may influence on the training dataset that then would effect on the accuracy. In our proposed model, there were no enlarge in the training datasets.

To the best of my knowledge, we have not come across to any research performed based on statistical operations nor using three scenarios to detect breast cancer for the classification accuracy purposes, which considered a new method in this field.

## V. CONCLUSION

The aim of this study is to investigate the use of deep learning model for breast cancer detecting using VGG-19 that used for extracting features from ultrasound images. Two layers of VGG19 structure were used: (i.e. fc6 and fc7); each of layer contains 4096 features. Also, more feature vectors were created from (fc6 and fc7), which called statistical operations. Different statistical operations are used to generate more datasets such: average, minimum, maximum and fusion of both fc6 and fc7. These datasets will be employed as input for the following ML classifiers: KNN, Random Forest, Naïve Bayes and Decision Tree. Data augmentation was considered to increase the dataset size for better learning of CNN.

Based on the results; Random Forest achieved high accuracy (88.63), precision (0.88), recall (0.88) and F-Measure (0.88) for fc7 of second scenario. The results were slightly similar; this approves that these features can provide a better accuracy when used in detection studies. Also, the results of the classification accuracy in the three scenarios are slightly similar, this approves that the breast cancer can be detected even if the size of data in the training dataset was minimal.

In the future work, it is recommending to conduct more investigation to improve the classification accuracy results and reduce training time using different algorithms.

## REFERENCES

- [1] WebMD, (2022). "Cancer Health Center (CHC): Understanding Cancer—Diagnosis and Treatment". WebMD. Medically Reviewed by Carmelita Swiner, MD on January 31, 2022, Last access: July. 13, 2023. [Online]. Available: <https://www.webmd.com/cancer/understanding-cancer-treatment>.

- [2] Cancer.Net Editorial Board, (2023). "Breast Cancer: Statistics". American Society of Clinical Oncology (ASCO). Accessed: <https://www.cancer.net/cancer-types/breast-cancer/statistics>.
- [3] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, (2008). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008" *Int. J. Cancer*, vol. 127, no. 12, pp. 2893–2917, Dec. 2010, doi: 10.1002/ijc.25516.
- [4] D. Adeloye, O. Y. Sowunmi, W. Jacobs, R. A. David, A. A. Adeosun, A. O. Amuta, S. Misra, M. Gadanya, A. Auta, M. O. Harhay, and K. Y. Chan, (2018). "Estimating the incidence of breast cancer in africa: A systematic review and meta-analysis" *J. Global Health*, vol. 8, no. 1, Jun. 2018, Art. no. 010419, doi: 10.7189/jogh.08.010419.
- [5] R. Ben-Ari, A. Akselrod-Ballin, L. Karlinsky, and S. Hashoul, (2017). "Domain specific convolutional neural nets for detection of architectural distortion in mammograms". 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 552–652. DOI:10.1109/ISBI.2017.7950581.
- [6] C. Nickson, K.E. Mason, D.R. English, A.M. Kavanagh, (2012). "Mammographic screening and breast cancer mortality: a case-control study and meta-analysis". *Cancer Epidemiol. Biomark. Prev.* 21 (9) (Sep. 2012) 1479–1488, <https://doi.org/10.1158/1055-9965.epi-12-0468>.
- [7] M. Broeders, et al., (2012). "The impact of mammographic screening on breast cancer mortality in europe: a review of observational studies" *J. Med. Screen* 19 (1) (Sep. 2012) 14–25, <https://doi.org/10.1258/jms.2012.012078>.
- [8] W.A. Berg, R.E.H. PhD, D.B. Kopans, P. Robert, A. Smith, (2020). "Frequently asked questions about mammography and the USPSTF recommendations: a guide for practitioners". *Soc. Breast Imag.* (2020) 1–17.
- [9] C.K. Kuhl, et al., (2015). "The Changing World of Breast Cancer: A Radiologist's Perspective". *Invest. Radiol.* 50 (9) (Sep. 2015) 615–628, DOI: <https://doi.org/10.1097/rli.000000000000166>.
- [10] S. Destounis, A. Santacroce, (2018). "Age to begin and intervals for breast cancer screening: balancing benefits and harms". *AJR Am J Roentgenol.* 210 (2) (2018) 279–284. DOI: <https://doi.org/10.2214/AJR.17.18730>.
- [11] Robert K. Horsley MD a, Juliana M. Kling MD, MPH b, Suneela Vegunta MD b, Roxanne Lorans MD c, H'hamed Tem kit PhD d, Bhavika K. Patel MD, (2019). "Baseline Mammography: What Is It and Why Is It Important? A Cross-Sectional Survey of Women Undergoing Screening Mammography". Volume 16, Issue 2, February 2019, Pages 164-169.
- [12] A.A. Roelofs, et al., (2007). "Importance of comparison of current and prior mammograms in breast cancer screening" *Radiology* 242 (1) (2007) 70–77. DOI: <https://doi.org/10.1148/radiol.2421050684>.
- [13] A. Rimmer, 2017. "Radiologist shortage leaves patient care at risk, warns royal college" *BMJ Br. Med. J. (Clin. Res. Ed.)* 359 (2017). DOI: <https://doi.org/10.1136/bmj.j4683>.
- [14] J.J. Fenton, et al., (2007). "Influence of computer-aided detection on performance of screening mammography" *N. Engl. J. Med.* 356 (14) (Apr. 2007) 1399–1409. DOI: <https://doi.org/10.1056/nejmoa066099>.
- [15] C.D. Lehman, R.D. Wellman, D.S.M. Buist, K. Kerlikowske, A.N.A. Tosteson, D. L. Miglioretti, (2015). "Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Med.* 175 (11) (Nov. 2015) 1828. DOI: <https://doi.org/10.1001/jamainternmed.2015.5231>.
- [16] A.D. Trister, D.S.M. Buist, C.I. Lee, (2017). "Will machine learning tip the balance in breast cancer screening?" *JAMA Oncol.* 3 (11) (Nov. 2017) 1463. DOI: <https://doi.org/10.1001/jamaoncol.2017.0473>.
- [17] American College of Radiology, Data science Institute, (2020). "Empowering machine learning in radiology". FDA Clear. AI algorithm. (2020), 1–1.
- [18] M. P. Sampat, G. J. Whitman, M. K. Markey, and A. C. Bovik, (2005). "Evidence based detection of spiculated masses and architectural distortions". *Proc. Med. Imag. Image Process.*, San Diego, CA, USA, Apr. 2005, pp. 26–37.
- [19] Computer Creations, (2023). "Why is the VGG Network Commonly Used?". NNART, LLC 2023. Accessed: <https://nnart.org/why-is-vgg-commonly-used/>.
- [20] Masud, M.; Rashed, A.E.E.; Hossain, M.S., (2022). "Convolutional neural network-based models for diagnosis of breast cancer". *Neural Computing and Applications*. volume 34, pages11383–11394 (2022). [Google Scholar] [CrossRef].
- [21] Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V., (2020). "Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review". *Appl. Sci.* 2020, 10, 8298. [Google Scholar] [CrossRef].
- [22] Zeebaree, D.Q., (2020). "A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images". *J. Appl. Sci. Technol. Trends* 2020, 1, 78–91. [Google Scholar].
- [23] Huang, K.; Zhang, Y.; Cheng, H.; Xing, P., (2021). "Shape-adaptive convolutional operator for breast ultrasound image segmentation". In *Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6. [Google Scholar].
- [24] P. Kus and I. Karagoz, (2010). "Accurate segmentation of the breast region with texture filter in mammograms for CAD applications", in *Proc. 15th Nat. Biomed. Eng. Meeting*, Apr. 2010, pp. 1–8.
- [25] J. M. Wolterink, P. de Jong, T. Leiner, W. Viergever, and I. Iágm, (2017). "Convnet based localization of anatomical structures in 3-D medical images", *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1470–1481, Jul. 2017.
- [26] M. A. Al-masni, M. A. Al-antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, E. Valarezo, S.-M. Han, and T.-S. Kim, (2017). "Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network", in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Seogwipo, Japan, Jul. 2017, pp. 1230–1233, doi: 10.1109/EMBC.2017.8037053.
- [27] P. Kaur, G. Singh, and P. Kaur, (2019). "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification". *Informat. Med. Unlocked*, vol. 16, Oct. 2019, Art. no. 100151, doi: 10.1016/j.imu.2019.01.001.
- [28] D. Singh and M. Singh, (2016). "Investigation on ROI selection for mammography using texture models and machine learning classifier". *Int. J. Control Theory Appl.*, vol. 9, p. 45, Dec. 2016.
- [29] M. Izadpanahkakhk, S. Razavi, M. Taghipour-Gorjikoalaie, S. Zahiri, and A. Uncini, 2018. "Deep region of interest and feature extraction models for palmprint verification using convolutional neural networks transfer learning". *Appl. Sci.*, vol. 8, no. 7, p. 1210, Jul. 2018.
- [30] Y. Hou, H. Zhang, S. Zhou, and H. Zou, (2017). "Efficient ConvNet feature extraction with multiple RoI pooling for landmark-based visual localization of autonomous vehicles". *Mobile Inf. Syst.*, vol. 2017, pp. 1–14, Apr. 2017, doi: 10.1155/2017/8104386.
- [31] R. Lakshmanan, S. T. P., V. Thomas, S. M. Jacob, and T. Pratab, (2014). "A preprocessing method for reducing search area for architectural distortion in mammographic images", in *Proc. 4th Int. Conf. Adv. Comput. Commun.*, Aug. 2014, pp. 101–104.
- [32] Mishra, A.K.; Roy, P.; Bandyopadhyay, S.; Das, S.K., (2021). "Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach". *Expert Syst.* 2021, 38, e12713. [Google Scholar] [CrossRef].
- [33] Byra, M., 2021. "Breast mass classification with transfer learning based on scaling of deep representations". *Biomed. Signal Process. Control* 2021, 69, 102828. [Google Scholar] [CrossRef].
- [34] Irfan, R.; Almazroi, A.A.; Rauf, H.T.; Damaševićius, R.; Nasr, E.; Abdelgawad, A., (2021). "Dilated Semantic Segmentation for Breast Ultrasonic Lesion Detection Using Parallel Feature Fusion". *Diagnostics* 2021, 11, 1212. [Google Scholar] [CrossRef].
- [35] Ahmad, N., Asghar, S., & Gillani, S.A., (2022). "Transfer learning-assisted multi-resolution breast cancer histopathological images classification". *Vis. Comput.*, 38, 2751–2770.
- [36] Kousalya, K., & Saranya, T., (2021). "Improved the detection and classification of breast cancer using hyper parameter tuning". Volume

- 81, Part 2, 2023, Pages 547-552. DOI: <https://doi.org/10.1016/j.matpr.2021.03.707>.
- [37] Duggento, A., Conti, A., Mauriello, A., Guerrisi, M., & Toschi, N., (2020). "Deep computational pathology in breast cancer". *Seminars in cancer biology*, 72, 226–237. doi: 10.1016/j.semcancer.2020.08.006.
- [38] Duc My Vo, Ngoc-Quang Nguyen, Sang-Woong Lee, (2019). "Classification of breast cancer histology images using incremental boosting convolution networks". *Information Sciences*. Volume 482, May 2019, Pages 123-138. <https://doi.org/10.1016/j.ins.2018.12.089>.
- [39] Kumar, A., Singh, S.K., Saxena, S., Lakshmanan, K., Sangaiah, A.K., Chauhan, H., Shrivastava, S., & Singh, R.K., (2020). "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer" *Information Sciences*, Volume 508, pages 405-421.
- [40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, (2012). "Imagenet classification with deep convolutional neural networks". In *Advances in neural information processing systems*, 1097–1105 (2012).
- [41] Z. Weiming, W. Henry, Y. Fung, C. Zhenghao, Z. Seid Miad, L. Zhicheng, and C. Yuk Ying, (2017). "Using Transfer Learning with Convolutional Neural Networks to Diagnose Breast Cancer from Histopathological Images". *ICONIP 2017, Part IV, LNCS 10637*, pp. 669–676, 2017. [https://doi.org/10.1007/978-3-319-70093-9\\_71](https://doi.org/10.1007/978-3-319-70093-9_71).
- [42] F.A. Spanhol, L.S. Oliveira, C. Petitjean, C. and L. Heutte, (2016). "Breast cancer histopathological image classification using convolutional neural networks". In *International Joint Conference on Neural Networks (IJCNN) 2016*, p. 2560-2567.
- [43] F.A. Spanhol, A. Fabio., L.S Oliveira., C. Ptitjean., 2016. "A dataset for breast cancer histopathological image classification". *IEEE Transactions on Biomedical Engineering*, (2016), vol. 63, no 7, p. 1455-1462.
- [44] Z. Han, B. WEI, Y. ZHENG, "Breast cancer multi-classification from histopathological images with structured deep learning model". *Scientific reports*, 2017, vol. 7, no 1, p. 4172.
- [45] N. Hatipoglu and B. Gokhan, (2014). "Classification of histopathological images using convolutional neural network". *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*. IEEE, 2014. p. 1-6.
- [46] R. Cruz , A. Basavanthally, A. Gonzalez et al., (2014). "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks". *Medical Imaging 2014: Digital Pathology*. International Society for Optics and Photonics, 2014. p. 904103.
- [47] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, (2014). "Forest species recognition using deep convolutional neural networks". *International Conference on Pattern Recognition*, 2014, pp. 1103–1107.
- [48] T. H. Vu, H. S. Mousavi, V. Monga, U. A. Rao, and G. Rao, (2015). "Dfdl: Discriminative feature-oriented dictionary learning for histopathological image classification," in *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, Apr. 2015, pp. 990–994.
- [49] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A., (2020). "Dataset of breast ultrasound images. Data in Brief". 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.
- [50] Manas Minnoor and Veeky Baths., (2023). "Diagnosis of Breast Cancer Using Random Forests". *Procedia Computer Science*. International Conference on Machine Learning and Data Engineering. Volume 218, pages 429-437. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2023.01.025>. Accessed: <https://www.sciencedirect.com/science/article/pii/S187705092300025X>.
- [51] Xiaofan Cheng & Liang Tan & Fangpeng Ming, 2021. "Feature Fusion Based on Convolutional Neural Network for Breast Cancer Auxiliary Diagnosis," *Mathematical Problems in Engineering*, Hindawi, vol. 2021, pages 1-10.
- [52] Hassanat, A. (2018). "Furthest-pair-based binary search tree for speeding big data classification using k-nearest neighbors". *Big Data*, 6(3): 225-235.
- [53] S. Al-Showarah et. al. 2020. "The Effect of Age and Screen Sizes on the Usability of Smartphones Based on Handwriting of English Words on the Touchscreen", *Mu'tah Lil-Buhuth wad-Dirasat, Natural and Applied Sciences series*, Vol. 35, No. 1, 2020. ISSN: 1022-6812.
- [54] Y. Li, Z. Xie, Z. He, X. Ma, Y. Guo, W. Chen, and Y. Lu, "Architectural distortion detection approach guided by mammary gland spatial pattern in digital breast tomosynthesis," *Proc. SPIE*, vol. 11314, Mar. 2020, Art. no. 1131417.
- [55] A. C. Costa, H. C. R. Oliveira, L. R. Borges, and M. A. C. Vieira, "Transfer learning in deep convolutional neural networks for detection of architectural distortion in digital mammography," *Proc. SPIE*, vol. 11513, May 2020, Art. no. 115130N.
- [56] Ketabi, H., Ekhlesi, A. & Ahmadi, H. A computer-aided approach for automatic detection of breast masses in digital mammogram via spectral clustering and support vector machine. *Phys Eng Sci Med* 44, 277–290 (2021). <https://doi.org/10.1007/s13246-021-00977-5>.