QUBAHAN ACADEMIC JOURNAL
VOL. 4, NO. 4, November 2024
https://doi.org/10.48161/qaj.v4n4a686

Categorizing and Determining of BI-RAD Breast Cancer Score by Using Deep Learning Techniques

Azhar Kassem Flayeh [1,*] and Ali Douik [1]

[1] Department National Engineering School of Sousse, University of Sousse/NOCCS Laboratory, Sousse, Tunisia.
* **Corresponding Author:** azharkassem2004@gmail.com.

ABSTRACT: Breast cancer is second only to skin cancer in prevalence. Failure to detect and treat breast cancer early could be fatal. Breast cancer is more common in underdeveloped countries due to low health awareness and limited early detection methods. However, detecting equipment and radiography advances have helped solve this difficulty and suggest a drop in mortality. Along with ultrasounds, mammography, Magnetic Resonance Imaging (MRI), Computed tomography (CT) scan, and Positron emission tomography (PET) scans are used for diagnosis. These devices add data to test findings for greater accuracy. Diagnosticians determine breast cancer stage by examining tumor size, shape, and spread. Artificial intelligence users employ these factors as fixed values to classify the illness using region of interest (RIO), Gray Level Co-Occurrence Matrix (GLCAM), clustering, or mass constraints like diameter and volume. Artificial intelligence algorithms have created many techniques to extract features from massive volumes of data through training, such as deep learning. The study investigated DL algorithms (Attention mechanism and efficintNetB7) to determine the breast imaging-reporting and data system (BI-RADS) score by categorizing mammography images. The dataset Digital Database for Screening Mammography (DDSM) used in this study, acquired from Kaggle, contains 55,890 training samples, of which 86% are negative, and 14% are positive, categorized into classes C1, C2, C3, C4, and C5. This work utilized deep learning techniques and algorithms to accurately categorize breast cancer at 93%, 92%, 90%, and 93% for each category (C2, C3, C4, and C5). We developed the BI-RADS classifier model using the bycharm framework, an integrated development environment (IDE) for Python programming, and using multiple libraries, namely NumPy, Pandas, TensorFlow, Keras, Matplotlib, and Seaborn. This study used a laptop with the following specifications: an Intel Core i7 CPU, 16GB of RAM, and Intel RTX integrated graphics.

Keywords: Breast Cancer (bc), Efficientnetb7, Mobilenetv2, BI-RAD.

I. INTRODUCTION

Presently, breast cancer constitutes 1 out of every eight cancer diagnoses, amounting to a total of 2.3 million new cases in both males and females combined [1]. In 2020, breast cancer was the most often detected cancer among women, representing 25% of all cancer cases in women. The incidence of this phenomenon is consistently rising, particularly in nations undergoing a period of change [2]. Approximately 685,000 women are projected to perish from breast cancer in 2020, accounting for 16% of all cancer-related fatalities in women, or 1 in every six deaths. The World Health Organization (WHO) has recently launched the Global Breast Cancer Initiative in response to insufficient public health actions that have been taken to address this issue [3]. Cancer arises in the human body due to the aberrant proliferation of tumor cells, leading to the infiltration of the neighboring tissue.

Deep learning has substantially advanced research and its application to medical pictures in healthcare for over ten years [4]. The study on breast cancer explores and successfully applies several deep architectures [5], yielding promising results. A significant number of researchers are now studying the use of computer-aided diagnostics for the diagnosis of breast cancer. Arevalo et al. [6] introduced a hybrid Convolutional

Neural Network (CNN) method where image-based features are learned using supervised learning approaches.

The scoring method used in the Breast Imaging Data and Reporting (BI-RADS) is valuable for discerning the specific classification of breast cancer. It is a clinical tool that helps doctors and physicians categorize breast cancer based on clinical pictures [7]. The American College of Radiology developed BI-RADS intending to establish a standardized system for reporting mammography results and evaluating observations, establish a standardized format for imaging reports, reduce ambiguity in the interpretation of clinical images by creating a common language understood by radiologists, facilitate effective communication among healthcare professionals regarding important findings in existing images and data, and assist in the monitoring of outcomes [8]. The BI-RADS classification system has crucial importance in an imaging report. To use the method, it is necessary to begin the report by describing the general structure of the breast composition [9]. This discovery has the potential to enhance breast screening and facilitate early detection of breast masses, hence reducing the need for surgical procedures. Additionally, it may improve the effectiveness of monitoring patients with breast cancer [10]. The BI-RADS classification system consists of seven categories ranging from 0 to 6. As the number increases, the malignancy also increases [11, 9].

Deep learning is a commonly used machine learning technique that involves a computer model directly categorizing data by acquiring knowledge from images [12]. The models have CNN architectures with many layers [13]. Deep learning algorithms autonomously identify cancer cells from medical photos [14].

The purpose of this study is to help the radiologist (save time and diagnosis) to decide about the cancer stage in the patient. The model can also be used as an application on a mobile phone or a web page after being trained on various datasets to facilitate the doctor's task. The BI-RADS score significantly benefits breast cancer categorization by standardizing reporting, guiding clinical decisions, improving communication, and ensuring quality assurance. It also aids in patient education and supports research efforts, contributing to better overall breast cancer management and outcomes. The contributions of this paper are as follows:

- Efficiency for Radiologists: The model helps radiologists save time and improve diagnostic accuracy in determining cancer stages, particularly for breast cancer.
- Accessibility: The model can be deployed as a mobile or web application, making it easily accessible for healthcare providers.
- BI-RADS Integration: Utilizes the BI-RADS scoring system to: Standardize reporting, Enhance clinical decision-making, Improve communication among healthcare providers, and Assure quality in diagnostic processes.
- Patient Education: Supports better patient education through clearer, standardized reporting.
- Research Support: Facilitates research in breast cancer diagnostics and management.
- Improved Healthcare Outcomes: Contributes to better overall breast cancer management and outcomes.

The structure of this paper is divided into three sections: the first section contains the introduction and an introduction about deep learning techniques and BI-RADS score, section two focuses on the literature review and previous work, section three the methodology and result, finally, the conclusion, future work and approved references.

II. LITERATURE REVIEW

The research undertaken by Huang et al. [16], titled "Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images," included the development of many models for categorizing BI-RADS in ultrasound images. These models used Convolutional Neural Networks (CNN), ResNet, and Visual Geometry Group (VGG) techniques. During the classification procedure, ResNet demonstrated a classification accuracy of 75.5%, VGG obtained an accuracy of 72.3%, and CNN earned an accuracy of 79.7% for BI-RADS category 3. Based on the accuracy of the findings, it can be inferred that CNN has performed exceptionally well in the classification of pictures. Consequently, we used progressive deep learning algorithms to get more precise outcomes.

An improved deep learning model for classification on the MIAS dataset was presented by Saber, A. et al. [17]. The task is to classify the dataset images into normality, beingness, and malignancy. The researchers aim to diagnose the cancerous area by implementing many preprocessing methods, first by removing noise, then by improving contrast, and removing non-breast regions from breast images. The basic MIAS dataset was preprocessed. The classification accuracy score was improved through freezing and fine-tuning processes while continuously expanding the dataset. According to the researchers, the VGG16 model had the best performance across all metrics (accuracy (98.96%), sensitivity (97.83%), specificity (99.13%), AUC (0.995), and F-score (97.66%)). In their attempt to train a network using the DistBelief distributed machine learning framework, Szegedy, C. [18] focused on employing data parallelism and a few models. While (CPU) architecture was considered, computational results demonstrated that several powerful GPUs could train the Google Neural Network (GAN) to convergence. For the training work, the researchers used an asynchronous stochastic gradient descent algorithm with a momentum of 0.9 and a preset learning rate schedule that decreases by 4% every 8 epochs. Using Polyak averaging, the final model was generated for inference. Alaa Hussein Abdulaal et al. [21] introduce a novel method for enhancing breast cancer classification by modifying the GoogLeNet architecture with a region-based attention mechanism. The updated GoogLeNet model incorporates a spatial transformer network (STN), which enables the model to concentrate on critical areas in breast histopathology images through selective attention. The proposed model achieved an accuracy of 98.08%, significantly higher than the original GoogLeNet's 94.99%. For multi-class classification at 100x magnification, this new model reached an accuracy of 94.63%, compared to the 85.06% accuracy of the unmodified GoogLeNet.

Islam, T. et al. [22], researchers assessed and compared the performance of five machine learning methods using a primary dataset of 500 patients from Dhaka Medical College Hospital. The methods evaluated included decision trees, random forests, logistic regression, naive Bayes, and XGBoost, all of which were applied to optimize results on the dataset. Additionally, the study employed SHAP analysis to interpret the XGBoost model's predictions and understand each feature's influence on its outputs. The findings revealed that XGBoost delivered the highest accuracy, achieving 97%.

III. METHODOLOGY

The study was carried out to investigate the use of deep learning algorithms in the BI-RADS categorization of mammography pictures. The dataset used in this study, acquired from Kaggle, comprises 55,885 mammograms categorized into BI-RADS classes C1, C2, C3, C4, and C5. This dataset comprises information extracted from the publication named "DDSM: Automatic Mammography-Based Age Estimation" by Leakage et al. from the Association for Computing Machinery (ACM), Kyoto, Japan [19-20]. The work was presented at the 3rd International Conference on Digital Medicine and Image Processing (DMIP 2020). Python libraries greatly aid in the creation of machine-learning applications. The deep learning model in this study was developed using multiple libraries, namely NumPy, Pandas, TensorFlow, Keras, Matplotlib, and Seaborn. This study used a laptop equipped with the following specifications: An Intel Core i7 CPU, 16GB of RAM, and Intel RTX integrated graphics. The first phase of this study used an autoencoder to train images and eliminate any unwanted noise. Figure (1) illustrates many steps of this process, which will be further elucidated in the subsequent paragraphs.

The first stage of the proposed model is combining Autoencoder to remove noise from images (preprocessing) while using MobilNet V2 to classify images (infected or uninfected), which provides better accuracy and efficiency compared to other lightweight models like SqueezeNet and ShuffleNet. The second stage is feeding the infected images to the model combining (attention mechanism &efficintNetB7). Combining the EfficientNetB7 architecture with an attention mechanism can significantly improve the classification of breast cancer images by allowing the model to focus on the most relevant parts of the image. Below is an outline of the approach, including the key formulas and algorithms:

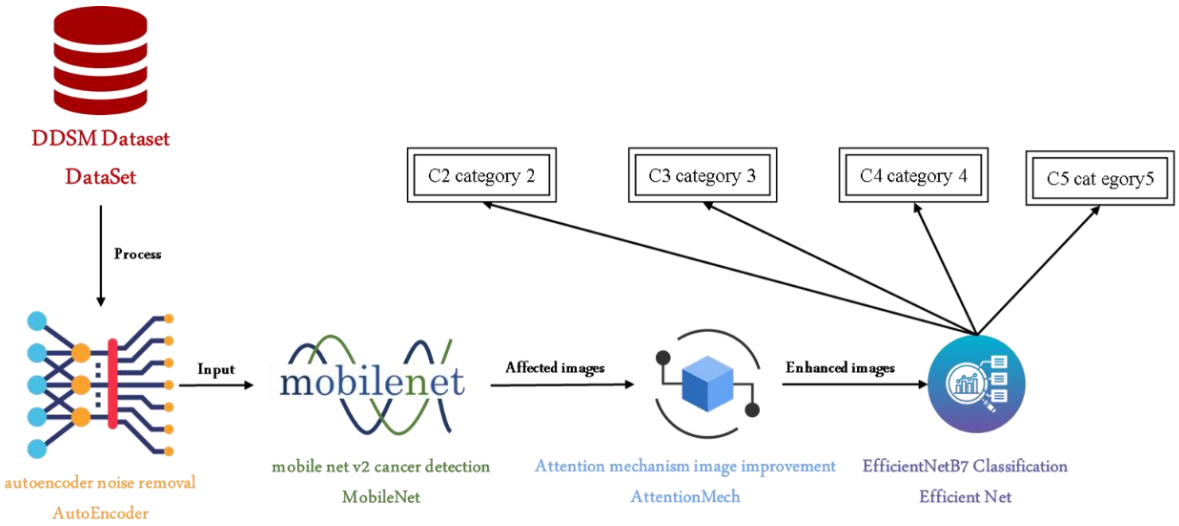

FIGURE 1. Diagram of the proposed method.

Let *(X)* be the input image, $F_{EffB7}(X)$ is the feature map extracted by EfficientNetB7.

$$F = F_{EffB7}(X) \quad (1)$$

Where *F* is the feature map of size (*H, W, C*), where e height, width, and number of channels in the feature map, respectively, to enhance the model's ability to focus on important regions of the image, an attention mechanism is applied to the feature map *F*). Let A_s be the spatial attention map of size (*H, W*, 1):

$$A_s=\sigma(Conv_{1x1}(\text{MaxPool(F)}) + Conv_{1x1}(\text{AvgPool(F)})) \quad (2)$$

Where (sigma) is the sigmoid function, $Conv_{1x1}$is a 1x1 convolution, (MaxPool) and (AvgPool) are max pooling and average pooling operations. Multiply the original feature map (F) with the attention map A_s to get the refined feature map F_s:

$$F_s= A_s \odot F \quad (3)$$

The final refined feature map (F_c)is passed through a global average pooling layer to reduce it to a feature vector, followed by a fully connected layer and a softmax activation for classification.

$$V=GAP(F_c) \quad (4)$$

Where *(V)* is the pooled feature vector of size (*C*).

$$y=Softmax(W \cdot V+b) \quad (5)$$

Where *(W)* and (*b*) are the weights and biases of the fully connected layer, and *(y)* is the output probability distribution over the classes (e.g., C2, C3, C4, C5). The overall model can be described as: Let (A_c) be the channel attention map of size (1, 1, C):

$$A_c=\sigma\ ((F_c\ (MaxPool(F_s))+(F_c\ (AvgPool(F_s)))$$

$$y=Softmax(W\cdot GAP(A_c \odot A_s \odot F_{EffB7}(X)) +b) \quad (6)$$

1. AUTOENCODER

A customized neural network architecture, an image-denoising autoencoder, is specifically developed to eliminate noise from medical pictures. The procedure entails training the autoencoder to acquire the ability to differentiate between the original clean signal and the noise that is added to it, thereby allowing it to learn the identity function in the presence of noise. The architecture has two primary components: the encoder and the decoder. The encoder reduces the dimensionality of the input picture and creates a condensed representation in a latent space, capturing important characteristics and removing unwanted disturbances. This encoder comprises a sequence of convolutional layers that progressively decrease the spatial dimensions of the picture. It is supplemented with pooling layers to down-sample the image and enhance the receptive field.

The latent space representation is a condensed "code" that captures the fundamental information of the input picture. This method seeks to extract the pristine, fundamental structure of the image, devoid of any unwanted noise, within the denoising environment. The decoder component of the Autoencoder utilizes the latent representation to reassemble the initial input. Nevertheless, due to the Autoencoder being trained on pictures with noise, the reconstruction process naturally prioritizes the important characteristics while disregarding the noise. The decoder often replicates the structure of the encoder but in a reversed manner, using up-sampling and convolutional layers to restore the picture to its original dimensions gradually.

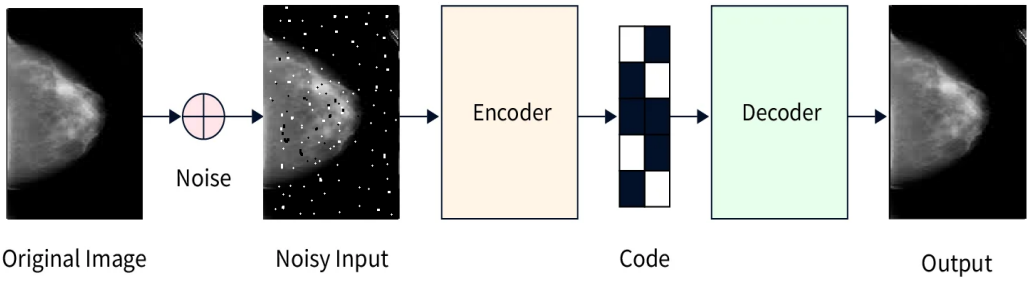

FIGURE 2. The autoencoder architecture.

2. MOBILENETV2

MobileNetV2 is a convolutional neural network architecture optimized for mobile and embedded vision applications. It achieves a good trade-off between computational resource use and model performance. The MobileNetV2 architecture may be customized and optimized to successfully handle the unique characteristics of MRI data when used for classifying "infected" and "normal" images in MRI classification tasks. MobileNetV2's central design is based on the inverted residual structure, which incorporates linear bottlenecks. This architecture begins with a conventional convolutional layer, which is then followed by a sequence of bottleneck layers. The bottleneck layer has three primary components: a 1x1 convolutional layer that increases the number of channels, a depth-wise 3x3 convolutional layer for filtering, and another 1x1 convolutional layer that reduces the dimensionality of the channels. The concept is to execute computationally intensive tasks in a space with many dimensions to achieve improved efficiency in representation. However, the dimensionality is reduced when combining features to retain a concise model.

The final layers of MobileNetV2 were adapted to suit the binary classification challenge for MRI classification. Furthermore, the initial classification head of the network was substituted with a novel fully

connected layer that modifies the input layer to include grayscale medical pictures, representing the "infected" and "normal" categories. Subsequently, the network is adjusted to optimize its performance using a dataset consisting of medical images.

3. ATTENTION MECHANISMS

An attention function can be described as a mathematical operation that takes a query and a set of key-value pairs as input and produces an output. We can represent the query vectors, keys, values, and results on this basis. The result is calculated by taking the weighted sum of the values. The matching function determines the weight of each value by comparing the query with the relevant key. The attention module allocates weights to distinct parts of feature maps, indicating the significance of each region to the given task. The weights are generated using a softmax function, which enables the network to prioritize areas with larger weights. The heat map is a visual representation of the attention weights that have been added to the original picture. The weights highlight the areas the model concentrates on the most, suggesting that these regions contain the most relevant information for the model's objective. For BC images, the attention mechanism identifies and emphasizes regions that exhibit abnormal forms, densities, or textures. This issue is crucial for ensuring precise diagnosis. By using this method, the mesh becomes more responsive to the existence of malignant tumors and less susceptible to being influenced by regular changes in the breast tissue.

4. EFFICIENTNETB7

EfficientNetB7, a sequence of convolutional neural networks (CNNs), plays a crucial role in the second stage of the classification process. Its architecture prioritizes model dimensions to enhance efficiency and accuracy. EfficientNetB7's fundamental structure uses mobile inverted bottleneck convolution blocks (MBConv), which integrate depthwise separable convolutions to minimize the number of parameters and computations. Moreover, EfficientNetB7 has squeeze-and-excitation blocks that dynamically adjust feature responses at the channel level. This issue is achieved by explicitly modeling the interdependencies across channels, improving the capacity to represent information inside the network. Figure 3 shows the architecture of EfficientNetB7.

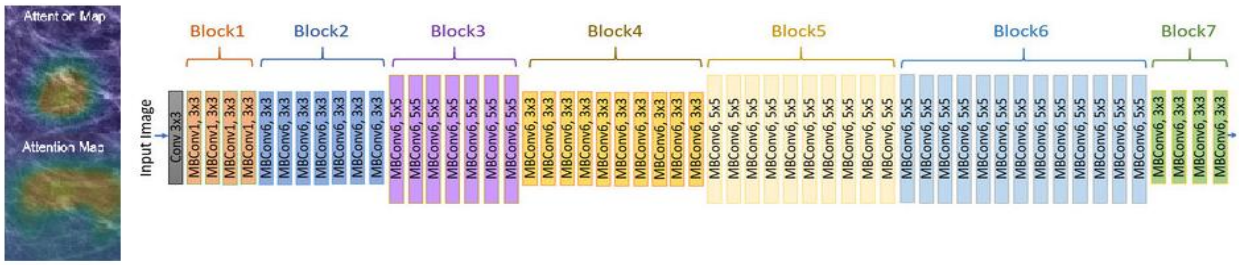

FIGURE 3. The efficintNet7 architecture.

5. DATASET

Digital Database for Screening Mammography DDSM archives provided the images for this collection, which can be downloaded from the following link:

Following preprocessing, the ROIs were removed from the images, and their size was adjusted. We save the data as record files that are unique to TensorFlow. The dataset contains 55,890 training samples, of which 86% are negative and 14% are positive. The five records files include these examples. Across the test files, the data for testing and validation has been distributed equally. Because the test and validation data separation was wrong, the test numpy files only contained mass information, whereas the validation files only contained calcification information. We must combine these files to guarantee a comprehensive and equitable gathering of test data.

The images have two labels applied to them: In the first place, label normal = 0 for the negative and 1 for the positive2. Label: complete multi-class labels (score); 0 indicates no change, 1 indicates benign calcification, 2 indicates benign mass, 3 indicates malignant calcification, and 4 indicates malignant mass. Building prediction models that can reliably identify breast lumps as benign or cancerous is the main goal of this machine learning effort. In order to help healthcare providers make educated decisions about patient treatment and to develop a powerful diagnostic tool that uses sophisticated algorithms. The dataset includes information as the following classification of the photos:

The scores range from 0 (negative) to 4 (malignant and spreading mass), with 1 denoting benign calcification, 2 benign masses, 3 malignant calcifications, and 4 signifying malignant and spreading mass. According to the above, it is necessary to illustrate the meaning of the degree of cancer. There are two ways to find out what stage cancer is in. To start, there's the TNM system. Cancers can be evaluated using three methods:

Tumor size (T), whether or not regional lymph nodes are involved (N), and whether or not distant metastases are present (M). After the T, N, and M categories are established, the patient is assigned a stage from 0 (in situ) to IV (the most advanced stage of the disease). Second, you can use the Breast Imaging-Reporting and Data System (BI-RADS) system to determine the stage of malignancy. As a quality assurance and risk assessment tool, the American College of Radiology has created a reporting framework and standardized vocabulary for breast imaging. Magnetic resonance imaging (MRI), ultrasound, and mammography can all benefit from this. We classify breast imaging tests using one of seven evaluation types:

The BI-RADS scale ranges from 0 (incomplete) to 1 (negative), with 1 indicating symmetry, 0 (no masses, architectural misrepresentation, or worrisome calcifications), and 2 (further imaging evaluations, such as additional mammographic images or ultrasound, are required). The second BI-RADS score indicates a benign condition with no chance of cancer. The third score indicates a 2% chance of cancer, with a recommendation for short intervals of follow-up. The fourth score indicates a 2–94% chance of cancer, with a recommendation to seek a biopsy. BI-RADS 5: strongly indicative of cancer, with a probability of more than 95%; further investigation is warranted if necessary. BI-RADS 6: A biopsy is required to confirm the tumor's malignant nature. The processed medical images from the DDSM Dataset are categorized into two classes. The first group consists of C1, which has 48,646 photos. These images are uninfected. The second group comprises the remaining categories, namely C2, C3, C4, and C5, with a specific numerical value. Cs = C2+ C3+ C4+ C5. The letter C indicates a category from 2-5 because C1 indicates the absence of cancer. The contaminated images refer to a total of 7,239 images across various categories.

IV. RESULTS

Autoencoder was used as preprocessing to remove noise from the medical images of the dataset (DDSM). A model was created that achieves the main function of the Autoencoder algorithm, and the results are shown in Figure (4).

In the first step of our study, we reached a classification accuracy of 96% via two classification phases. This phase allowed us to classify the infected images, which would be further classified in the following stage. Figure (5), the confusion matrix evaluates a model's predictions for an infection detection task. It shows a large number of true negatives (46700) and true positives (6949), with relatively few false negatives (290) and false positives (1946), indicating a higher ability to identify both infected and uninfected cases correctly.

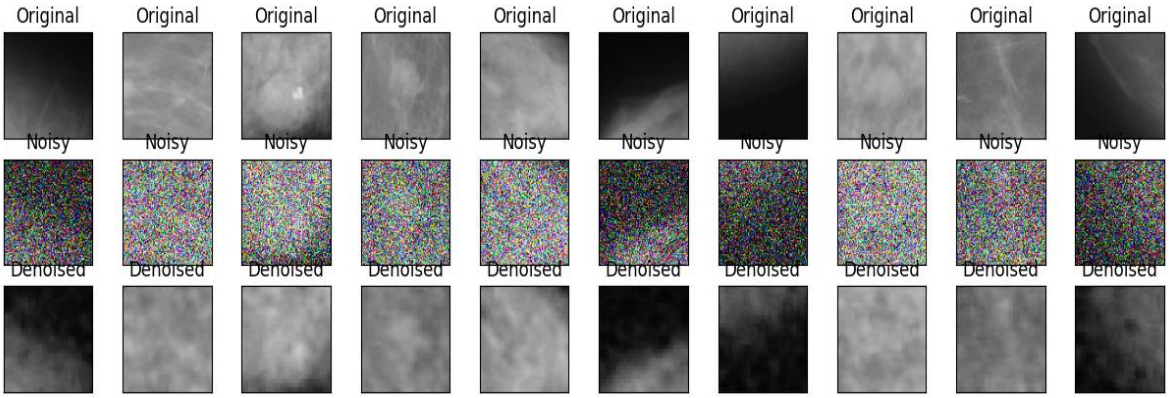

FIGURE 4. Represents the denoising results of the DDSM Dataset.

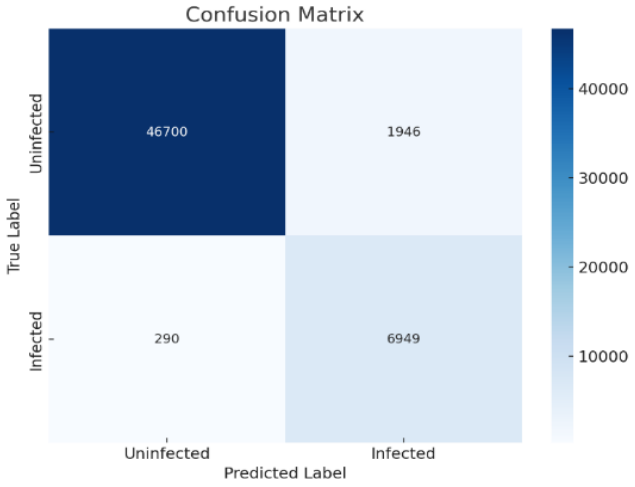

FIGURE 5. Confusion matrix of mobileNetV2.

The confusion matrix for MobileNetV2, as with any model, is a tool that measures its performance on a classification task by comparing the predicted labels with the actual labels. It is a table where each row represents the instances of an actual class, and each column represents the instances of a predicted class. For MobileNetV2, this matrix reveals how well the model identifies different classes, showing each class's number of true positives, false positives, true negatives, and false negatives. This detailed breakdown helps evaluate the model's accuracy, precision, recall, and overall effectiveness and is crucial for understanding its performance in practical applications. The analysis of the confusion matrix is below:

- Precision: 0.781, meaning about 78.1% of the images predicted as "infected" were infected.
- Sensitivity (Recall): 0.960, indicating that about 96.0% of the infected images were correctly identified.
- Specificity: 0.960, indicating that about 96.0% of the actual uninfected images were correctly identified as uninfected.

Figure (6) depicts three-line graphs tracking the training and validation precision, sensitivity (recall), and specificity of a machine learning model over epochs, which are complete cycles through the training dataset.

Each graph shows two lines: one for the training metric and one for the validation metric. The graphs indicate fluctuations in model performance at each epoch, which is common during training. While sensitivity and specificity remain high, precision shows more variability. The goal is to achieve high and stable values for all metrics, with minimal gaps between training and validation, suggesting good model generalization.

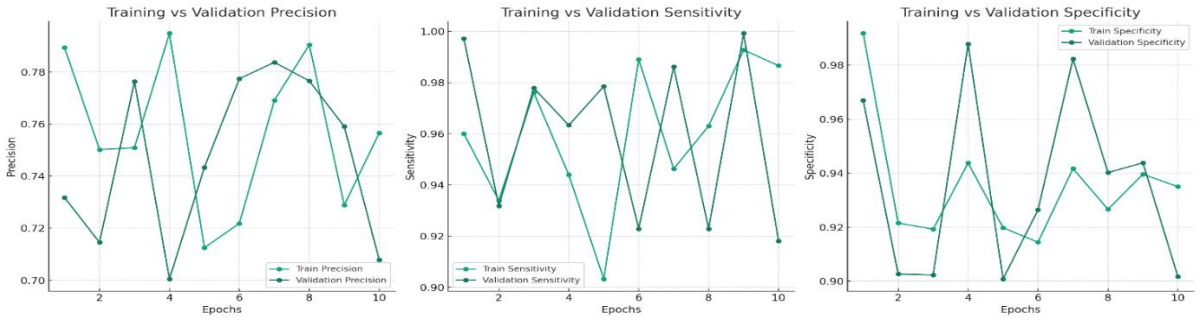

FIGURE 6. Training vs validation (Precision, Sensitivity, Specificity).

The CNN model in Figure (3) utilizes transfer learning of EfficientNetB7 as its initial model. EfficientNetB7 exhibits impressive performance while using only moderate CPU resources. EfficientNetB7 has a substantial parameter count of around 66 million, backed by a commendable level of accuracy. Later, alterations were implemented to convolutional layers (Convolutional2D) and pooling layers (MaxPooling2D), specifically developed to process picture data. The primary role of this layer is to identify and acknowledge. Pictures are formed by applying the Rectifier Linear Unit (ReLU) activation function to the pixels. At this point, 8 filter parameters are imported, indicating that 8 filters with a size of 3x3 (kernel size = (3, 3)) are used in the convolution process. The second layer utilizes 16 filter parameters, the third layer uses 32 filters, the fourth layer incorporates 64 layers, and the fifth layer utilizes 128 filter parameters. The input shape for all layers is defined as (200, 150, 3), indicating that all input pictures must have dimensions of 200x150 pixels and consist of three color channels represented in matrix form. The ReLU activation function is evenly applied to all layers, as it is often used in Convolutional Neural Networks (CNNs). In addition, a MaxPooling2D layer is included to decrease the resolution of the picture while preserving important details. The MaxPooling2D layer functions by picking the maximum value from the processed data, and a pooling size of 2x2 is used in the first layer. The reduced size of the feature maps in this configuration accelerates the analysis time. The predicted value for each categorization is shown in Table 1.

Table 1. Predicted value for each category.

ID	Predicted C2	Predicted C3	Predicted C4	Predicted C5
Actual C2	1921	51	49	42
Actual C3	48	1752	49	42
Actual C4	48	51	1329	42
Actual C5	48	51	49	1667

Figure (7) is the confusion matrix, which represents a table used to describe the performance of a classification model on a set of test data for which the true values are known. Below is a breakdown of its components:

- Labels on the Matrix: The rows represent the actual classes, while the columns represent the predicted classes. In this case, four classes are labeled C2, C3, C4, and C5.

- Matrix Values: Each cell in the matrix shows the count of predictions made by the model. For example, the model predicted class C2 1921.4 times when the actual class was C2, which suggests the data or the count has been processed or normalized in some way since it's a non-integer value.
- Off-Diagonal Values: The not-on-the-diagonal cells show incorrect predictions. For instance, the cell in the row for Actual C3 and the column for Predicted C2 shows 48.2, indicating that there were 48.2 instances where the model incorrectly predicted class C2 when the actual class was C3.
- Color Intensity: The color intensity in each cell usually corresponds to the cell's value, with darker colors representing higher numbers. This visual aid helps to quickly identify which classes are being predicted correctly or incorrectly.
- Side Bar (Color Scale): The color scale on the right side indicates the range of values, with the lighter shade at the bottom representing lower values and the darker shade at the top representing higher values. It helps in understanding the color intensity in the context of the matrix values.

In category C2, a precision of 0.93 indicates that 93% of the model's positive predictions are accurate. The sensitivity of 0.931 and specificity of 0.972 demonstrate excellent accuracy in detecting positive and negative instances, respectively. A 93% total accuracy indicates high efficacy in all forecasts, as shown in Table 2. Other categories may be similarly interpreted, demonstrating the model's proficiency in various prediction characteristics. Figure (7) illustrates the confusion matrix for (7239 infected images) after categories.

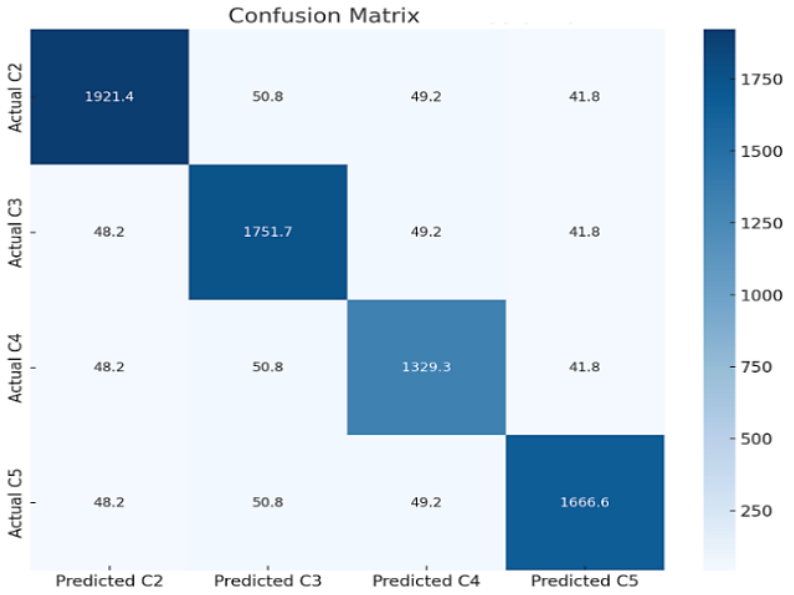

FIGURE 7. Confusion Matrix of EfficientNetB7.

Figure (8) displays mammogram images with predicted BI-RADS categories ranging from C2 to C5. The images likely correspond to different levels of abnormality, with C2 being benign findings and C5 highly suggestive of malignancy. This Figure demonstrates the model's capability to categorize breast imaging.

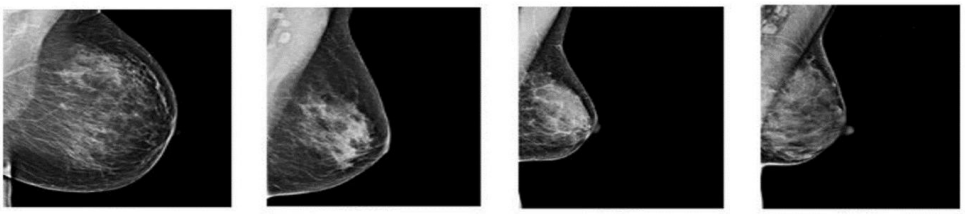

FIGURE 8. Figure description.

Table (2) summarizes performance metrics for a classification system, showing precision, sensitivity (also known as recall), specificity, and accuracy for four distinct categories labeled C2, C3, C4, and C5. Precision measures the proportion of true positives among all positive predictions. Sensitivity gauges the model's ability to identify actual positives correctly. Specificity assesses the model's accuracy in identifying true negatives. Accuracy reflects the overall proportion of true predictions. The table indicates that all categories have high-performance metrics, with precision and sensitivity slightly higher in C2 and C5, specificity highest in C5, and accuracy consistently above 90% for all categories, suggesting a reliable classification system.

Table 2. Displays Precision Sensitivity Specificity Accuracy results for the categories that have been classified.

ID	Precision	Sensitivity	Specificity	Accuracy
C2	0.93	0.931	0.972	93
C3	0.92	0.926	0.972	92
C4	0.90	0.904	0.974	90
C5	0.93	0.918	0.977	93

V. DISCUSSION

Analyzing the deep learning model that utilizes a combination of an autoencoder and MobileNetV2 in the first stage for classifying breast cancer images into infected and non-infected categories, then in the second stage where images of infected are fed into a framework consisting of an attention mechanism and EfficientNetB7.

The Autoencoder serves as a feature extraction tool. Its role is to compress the input images into a lower-dimensional latent space and reconstruct them. This issue helps learn essential features from the images while reducing noise and irrelevant details. In the context of breast cancer images, the Autoencoder is likely used to extract high-level features crucial for distinguishing between infected and non-infected tissues.

MobileNet, known for its efficiency and compact architecture, acts as a classifier on the features extracted by the Autoencoder. MobileNet is designed to be lightweight and fast, making it suitable for real-time image classification tasks. When fed with the processed features from the Autoencoder, MobileNet classifies the images into the desired categories—infected or non-infected.

The attention mechanism helps the model focus on the most relevant parts of an image, highlighting regions that are more important for classification. This issue can be crucial for medical imaging, where specific features or anomalies may indicate cancer. By weighting different parts of the image differently, the attention mechanism can improve the model's ability to distinguish between affected and unaffected areas.

EfficientNetB7 is a highly efficient and accurate convolutional neural network (CNN) that balances model performance and computational efficiency. It uses compound scaling to optimize depth, width, and resolution. When combined with the attention mechanism, EfficientNetB7 can leverage its high capacity and

efficiency to process and classify the important features highlighted by the attention mechanism. We can discuss the model's work, which mainly focuses on categorizing infected images according to the BI-RAD scale. During training, the attention mechanism and EfficientNetB7 will jointly learn to categorize infected images. The attention mechanism helps guide EfficientNetB7's focus, potentially leading to better feature extraction and classification. The training process will involve optimizing the attention mechanism and EfficientNetB7 to minimize categorization loss.

The constraints on generalizability and interpretability highlight the necessity for ongoing research in algorithm advancement and AI integration approaches. To tackle these issues, it is crucial to focus on acquiring diverse and high-quality datasets, creating more transparent and understandable models, employing multimodal analysis methods, and improving robustness and scalability. Addressing these areas will optimize AI's effectiveness in clinical settings. Additionally, data availability, model explainability, costs, and privacy concerns emphasize the importance of developing public datasets and open training and inference platforms to support global developers and healthcare practitioners.

VI. CONCLUSION

In exploring breast cancer detection and categorization within medical imaging, particularly through the analysis of the Digital Database for Screening Mammography (DDSM), a sophisticated multi-stage AI-driven methodology was employed, demonstrating remarkable efficacy in disease identification and subsequent classification. Initially, the Autoencoder algorithm was leveraged to cleanse the DDSM dataset meticulously, an imperative step to ensure the integrity and reliability of the data by mitigating noise and enhancing feature prominence. This preprocessing phase was crucial for the success of subsequent analyses, laying a solid foundation for accurate disease detection. Following the data refinement, the MobileNetV2 algorithm was applied to detect breast cancer within the processed images. MobileNetV2, known for its efficiency and effectiveness in handling image data, particularly in mobile and resource-constrained environments, achieved an impressive detection rate of 96%. This high level of accuracy underscores MobileNetV2's capability to discern intricate patterns and anomalies associated with breast cancer, making it a potent tool in the early and reliable diagnosis of the disease. An attention mechanism was introduced to enhance image quality further and highlight critical features relevant to cancer detection and classification. This innovative approach significantly augmented the clarity and focus of the images, leading to an appreciable improvement in the results by 1.5 to 2.5 percentage points. The attention mechanism's role in amplifying relevant features while suppressing extraneous information was instrumental in elevating the overall diagnostic accuracy. The final phase of the study involved the classification of the enhanced images into four distinct categories (C1, C2, C3, C4) utilizing the EfficientNetB7 algorithm. EfficientNetB7, renowned for its scalability and higher efficiency, demonstrated commendable performance, with classification accuracies reaching 93%, 92%, 90%, and 93% for categories C1, C2, C3, and C4, respectively. These results attest to the algorithm's robustness and precision in categorizing varying stages of breast cancer and highlight the potential for nuanced patient-specific treatment planning. Overall, this multi-faceted approach integrating Autoencoder for noise reduction, MobileNetV2 for disease detection, attention mechanisms for image enhancement, and EfficientNetB7 for detailed classification presents a compelling case for the advancement of AI in medical diagnostics. The high accuracy rates achieved in both detection and classification phases signify a substantial stride towards the automation of breast cancer screening, promising a future where early detection and tailored treatment strategies could significantly improve patient outcomes and survival rates. In our forthcoming study, we will compare the outcomes we achieved in classifying the severity of breast cancer using deep learning with an alternative approach, such as a statistical method like Hu Moment. Additionally, we will endeavor to collect a specialized dataset regarding the results of the Oncotype DX test. This genomic assay offers a Recurrence Score (RS) that predicts the likelihood of recurrence over ten years and the response to adjuvant chemotherapy. We utilize deep learning to employ the Oncotype DX recurrence score to forecast the probability of a distant breast cancer recurrence accurately.

REFERENCES

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (**2021**). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209-249.
2. Zebari, D. A., Haron, H., Sulaiman, D. M., Yusoff, Y., & Othman, M. N. M. (**2022**). CNN-based Deep Transfer Learning Approach for Detecting Breast Cancer in Mammogram Images. In *2022 IEEE 10th Conference on Systems, Process & Control (ICSPC)* (pp. 256-261). IEEE.
3. Anderson, B. O., Ilbawi, A. M., Fidarova, E., Weiderpass, E., Stevens, L., Abdel-Wahab, M., & Mikkelsen, B. (**2021**). The Global Breast Cancer Initiative: A strategic collaboration to strengthen health care for non-communicable diseases. *The Lancet Oncology*, 22(5), 578-581.
4. Das, H. S., Das, A., Mallik, S., Bora, K., & Zhao, Z. (**2023**). Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. *Frontiers in Genetics*, 13, 1097207.
5. Hamidinekoo, A., Garzón-Martínez, G. A., Ghahremani, M., Corke, F. M., Zwiggelaar, R., Doonan, J. H., & Lu, C. (**2020**). DeepPod: A convolutional neural network based quantification of fruit number in Arabidopsis. *GigaScience*, 9(3), giaa012.
6. Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M. A. G. (**2016**). Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 127, 248-257.
7. American College of Radiology BI-RADS Committee. (**2013**). *Acr bi-rads atlas: Breast imaging reporting and data system* (5th ed.). American College of Radiology.
8. Lee, J. (**2017**). Practical and illustrated summary of updated BI-RADS for ultrasonography. *Ultrasonography (Seoul, Korea)*, 36(1), 71–81.
9. Spak, D. A., Plaxco, J. S., Santiago, L., Dryden, M. J., & Dogan, B. E. (**2017**). BI-RADS® fifth edition: A summary of changes. *Diagnostic and Interventional Imaging*, 98(3), 179-190.
10. Ghaemian, N., Tehrani, N. H. G., & Nabahati, M. (**2021**). Accuracy of mammography and ultrasonography and their BI-RADS in detection of breast malignancy. *Caspian Journal of Internal Medicine*, 12(4), 573.
11. Bartolotta, T. V., Orlando, A. A. M., Di Vittorio, M. L., Amato, F., Dimarco, M., Matranga, D., & Ienzi, R. (**2021**). S-Detect characterization of focal solid breast lesions: A prospective analysis of inter-reader agreement for US BI-RADS descriptors. *Journal of Ultrasound*, 24, 143-150.
12. Mostafa, S. A., Ravi, S., Zebari, D. A., Zebari, N. A., Mohammed, M. A., Nedoma, J., ... & Ding, W. (**2024**). A YOLO-based deep learning model for Real-Time face mask detection via drone surveillance in public spaces. *Information Sciences*, 120865.
13. Devare, M. H. (**2019**). Challenges and opportunities in high performance cloud computing. In *Handbook of Research on the IoT, Cloud Computing, and Wireless Network Optimization* (pp. 85-114).
14. Charan, S., Khan, M. J., & Khurshid, K. (**2018**). Breast cancer detection in mammograms using convolutional neural network. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-5). IEEE.
15. Tobías, L., Ducournau, A., Rousseau, F., Mercier, G., & Fablet, R. (**2016**, December). Convolutional Neural Networks for object recognition on mobile devices: A case study. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 3530-3535). IEEE.
16. Huang, Y., Han, L., Dou, H., Luo, H., Yuan, Z., Liu, Q., ... & Yin, G. (**2019**). Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. *Biomedical Engineering Online*, 18, 1-18.
17. 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.
18. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (**2015**). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).
19. Purnama, M. M. R., Sandi, I. N., Gunawan, A. A. N., Sutapa, G. N., Widagda, I. G. A., & Wendri, N. (**2024**). Classification of BI-RADS using convolutional neural network and EfficientNet-B7. *International Journal of Science and Research Archive*, 11(1), 1022-1028.
20. Lekamlage, C. D., Afzal, F., Westerberg, E., & Cheddad, A. (**2020**, November). Mini-DDSM: Mammography-based automatic age estimation. In *2020 3rd International Conference on Digital Medicine and Image Processing* (pp. 1-6).
21. Abdulaal, A. H., Valizadeh, M., Albaker, B. M., Yassin, R. A., Chehel Amirani, M., & Shah, A. F. M. (**2024**). Enhancing breast cancer classification using a modified GoogLeNet architecture with attention mechanism.
22. Islam, T., Sheakh, M. A., Tahosin, M. S., et al. (**2024**). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Scientific Reports*, 14, 8487.