Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

# A Novel Hybrid Attention VGG Method For Benign and Malignant Breast Cancer Classification

**Mustafa Salih BAHAR**[1*]
[1] Yildiz Technical University, Computer Engineering Department, mustafasalihbhr@gmail.com, Orcid No: 0000-0002-1625-9362

ARTICLE INFO

*Article history*:

Received 18 June 2024
Received in revised form 24 October 2024
Accepted 19 December 2024
Available online 26 March 2025

*Keywords*:

*Hybrid Attention VGG, Benign, Malignant, Breast Cancer Classification, Deep Learning Models*

*Doi: 10.24012/dumf.1502403*

* Sorumlu Yazar

ABSTRACT

Worldwide, breast cancer is quite widespread among many types of cancer. Early detection is crucial for effective treatment. While early detection does not cure cancer or prevent its recurrence, it significantly improves treatment outcomes. Regular breast cancer check-ups, including mammograms, play a vital role in early detection. The type of the observed tumour is also crucial. Therefore, our study utilized a range of deep learning methods to accurately classify distinct forms of breast cancer cells, including both benign and malignant varieties. The problem addressed in the study relies on the classification of tumour images as either benign or malignant. We used the augmented MIAS and INBREAST datasets, implementing fourteen deep learning models by adjusting different hyperparameter values. Aside from these, we trained a new model we created, the Hybrid Attention VGG16 model, on the datasets by adjusting the batch size and learning rate values used in other models. Our research has shown that initially models like VGG16, VGG19, ResNet50, ResNet101, EfficientNetV2B0 and EfficientNetV2L performed better at different hyperparameter values, whereas our proposed model, the Hybrid Attention VGG model, achieved one of the highest performance among deep learning models across many hyperparameter values and on both datasets, especially on the Augmented INBREAST dataset. Our newly proposed model, with its unique skip connection and attention mechanism, surpasses the accuracy of models employed in earlier studies, as demonstrated when comparing them in the literature.

## Introduction

Cancer is a highly lethal illness that, if not detected in its early stages, can rapidly spread to other cells, resulting in severe harm to vital organs such as the pancreas, lungs, breasts, and blood. Notably, breast cancer is the most widespread kind of cancer among women worldwide. Breast cancer is a type of cancer that has a high mortality rate when not diagnosed early and accurately. In particular, it is necessary to determine the shade of the detected tumour. Mammography is one of the most helpful imaging tools used by radiologists to correctly diagnose breast lesions by performing many tests under challenging conditions. With the emergence of deep learning models, as opposed to traditional machine learning methods, with fewer data preprocessing steps, more successful results have been achieved through more complex models. Machine learning relies on manual feature extraction, whereas deep learning and artificial neural networks depend upon automatic feature extraction. Convolutional neural networks are a renowned and extensively utilized model in deep learning. A Convolutional Neural Network (CNN) is a form of artificial neural network specifically created to reduce the requirement for lengthy preprocessing. It is a deep, feed-forward network that utilizes a version of multilayer perceptrons.

Since the advent of deep learning, numerous studies have been published that utilize deep architectures [1]. The Convolutional Neural Network (CNN) is the predominant deep learning architecture. Arevalo et al. [2] conducted a study where they tested and compared various Convolutional Neural Networks (CNNs) to detect masses using two manually designed descriptors. Their experimentation was performed using the dataset from the Breast Cancer Digital Repository Film Mammography [3]. Huynh et al. [4] employed the pre-trained AlexNet [5] for mass diagnosis without additional fine-tuning. Jias et al. [6] propose a method that fine-tunes a pre-trained convolutional neural network (CNN) utilizing a part of the digital database specifically designed to screen mammography (DDSM) database. Ting et al. [7] have developed and trained their breast mass classification network. Rampun et al. [8] employed a modified pre-trained and fine-tuned variant of AlexNet on a carefully selected subset of breast images from the DDSM dataset, known as CBIS-DDSM. Benign and malignant tumours are distinctly different. Round or oval shapes characterize benign tumours, while malignant tumours have unpredictable outlines. Furthermore, a comparison is made between support vector machines (SVM) and artificial neural networks [9] to classify healthy, aberrant tissues, benign tumours, and malignant tumours [10]. The breast cancer classification techniques are presented in reference [11]. Several CNN designs are accessible, including CiFarNet [12], AlexNet, GoogLeNet [13], ResNet, VGG16, and VGG19 [14]. Transfer learning refers to utilizing a pre-existing model, which has been trained on a particular problem, to address a different situation. Training an extensive neural network from the beginning requires substantial data and computational resources. Data augmentation is a method that enlarges the training dataset by generating additional samples by implementing random alterations to the existing data [15]. The advantages of this include accelerating the convergence process, avoiding excessive adjustment, and enhancing generalization capabilities [16]. A practical method is to make subtle modifications to limited datasets, such as translation, zooming, flipping, mirroring, rotation, and so on [17].

The primary focus of this research is to highlight the significance of the Hybrid Attention VGG model, which is a more efficient and innovative model for classifying benign and malignant breast cancer. This study stands out from others in its emphasis on this particular model. Our objective is to showcase the effectiveness of this model by conducting a comparative analysis with fourteen deep-learning models on two distinct datasets, applying a range of hyperparameter values. The models in examining are VGG16, VGG19, DenseNet121, DenseNet169, ResNet50, ResNet101, MobileNet, MobileNetV2, InceptionV3, InceptionResNetV2, Xception, NasNetMobile, EfficientNetV2B0 and EfficientNetV2L. The investigation used mammography images from the Mammographic Image Analysis Society (MIAS) and INBREAST databases. We acquired both datasets, which employed diverse augmentation strategies, in PNG format. We categorised the dataset into separate folders based on the benign and malignant classes; then, we retrieved the class labels from these folders.

The following sections of the article are arranged in the following manner: Section 2 provides a concise overview of the datasets utilised for mammography imaging and the augmentation strategies employed. Chapter 3 focuses on the deep learning models used in this classification assignment and comprehensively explains the newly presented model. Section 4 contains the performance criteria utilised to assess and compare the achievements of these models. Section 5 presents the performance metrics of the deep learning models employed and our proposed model across various hyperparameter values and datasets. It also includes graphical representations of accuracy and loss values during the training and validation stages for the most successful models and our proposed model. Additionally, the results of the confusion matrix are provided. Chapter 6 evaluates the results, examining the effectiveness of our model concerning prior studies and mentioning possible avenues for further research. Chapter 7 is a conclusion section that explains the reasons behind the success of the suggested model and presents information regarding future study.

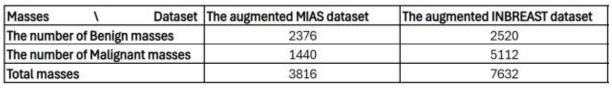
## Materials and Methods

### Imbalanced Datasets and Data Augmentation Techniques

The problem of imbalanced datasets frequently arises in numerous classification tasks. An imbalanced dataset arises when one or more classes possess markedly fewer samples than others, resulting in disproportionate representation among the classes.

In breast cancer medical statistics, malignant tumour cells are typically significantly less than benign tumour cells. Data augmentation approaches represent a practical approach. This technique enhances the minority class or several classes using diverse augmentation methods.

In the dataset utilised, 106 images of breast masses were picked from the 410 mammograms in the INbreast database for this investigation. This study boosted the amount of breast mammography images to 7632 through data augmentation, comprising 2520 benign and 5112 malignant tumours. Primarily, horizontal or vertical flips and rotations between 30 and 330 degrees have been utilized. The augmented MIAS dataset includes 2376 benign and 1440 malignant masses, comprising 3816. Figure 1 provides information about the augmented datasets. Augmentation techniques have been used for both classes in the datasets [18].



| Masses \ Dataset | The augmented MIAS dataset | The augmented INBREAST dataset |
|---|---|---|
| The number of Benign masses | 2376 | 2520 |
| The number of Malignant masses | 1440 | 5112 |
| Total masses | 3816 | 7632 |

Figure 1. The augmented datasets and the counts of benign and malignant masses

**Rotation** involves adjusting the position of an image by a specified angle, either in a clockwise or anticlockwise direction. This method enhances the model's ability to identify things from various perspectives and augments data diversity. In medical imaging, rotating mammography images of masses enables the model to learn to recognize the same mass from several viewpoints, hence enhancing its generalization capability. Nevertheless, rotation can occasionally damage the inherent structure of the data, necessitating cautious use.

**Flipping** denotes the reflection of an image across a horizontal or vertical axis. Horizontal flipping reflects the image laterally, but vertical flipping reflects it vertically, so augmenting the dataset and enhancing its diversity. This enables the model to identify objects from various orientations. In medical imaging, such as mammograms, horizontal flipping can assist the model in accurately analyzing both breasts. Nonetheless, as flipping may modify the anatomical structure in certain instances, particularly in medical circumstances, it should be executed with prudence.

### Dataset

The collection comprises mammography images depicting both benign and malignant tumours. The collection consists of 106 mass photographs obtained from the INBREAST dataset, 53 mass images from the MIAS dataset, and 2188 from the DDSM dataset. Subsequently, they utilize data augmentation and contrast-limited adaptive histogram equalization approaches to preprocess the images. The INBREAST dataset has 7632 photos after data augmentation, the MIAS dataset contains 3816 images, and the DDSM dataset contains 13128 images. Furthermore, they consolidate the INBREAST, MIAS, and DDSM datasets. The images' overall size was adjusted to 227*227 pixels [18].

This work utilized a dataset [18] in which each image was annotated with the matching breast density. The DICOM files containing the original pictures from the mammography database were transformed into PNG files. The initial 106 photos have undergone the application of contrast-limited adaptive histogram equalization (CLAHE), a method used for image preparation. Furthermore, with CLAHE, the data is enhanced by rotating it at various angles (The values of θ are 30, 60, 90, 120, 150, 180, 210, 240,
270, 300, and 330 degrees). Subsequently, the original and 11 rotated images are flipped horizontally and vertically. This approach has also been discovered to mitigate the problem of overfitting.

A breast cancer tumour can be categorized as either benign, indicating that it poses no threat to one's health, or malignant, indicating that it has the potential to be destructive and deadly. Benign tumours are non-malignant as their cells closely resemble normal cells; they grow gradually and do not infiltrate nearby tissues or metastasize to other body areas. Malignant tumours are characterized by their malignant nature. If left untreated, malignant cells have the potential to metastasize and spread to other regions of the body, extending beyond the boundaries of the initial tumour. Malignant tumours are lethal due to their dramatically faster growth rate compared to benign tumours.

### Augmented MIAS Dataset

This study utilized the augmented MIAS dataset, which included PNG images. The dataset included 2456 samples labelled as Benign and 1440 as Malignant.

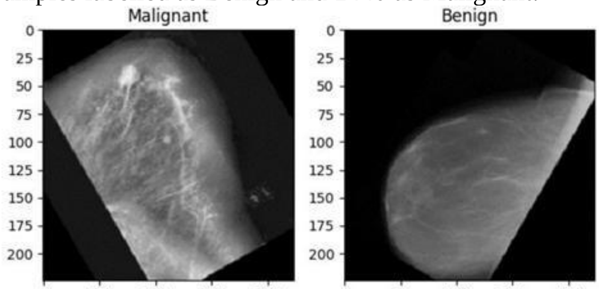


Figure 2. Some images from the Augmented MIAS dataset and their classes

Figure 2 shows two classes of mammogram images: benign and malignant. Benign is the type of tumour with good behaviour, while malignant is the tumour type with bad behaviour. Benigns often have a modest growth rate, a small geographic range, and a capsule surrounding them to prevent direct interaction with nearby tissues. Benign tumours do not exhibit metastases. Malignant is a type of tumour that can be considered cancer. A primary tumour is a malignant growth originating in a specific body location. Conversely, the tumours that develop in other parts of the body due to this tumour are called metastases.

### Augmented INBREAST Dataset

For this investigation, we utilized the augmented INBREAST dataset, comprising 2520 images classified as Benign and 5112 images classified as Malignant.

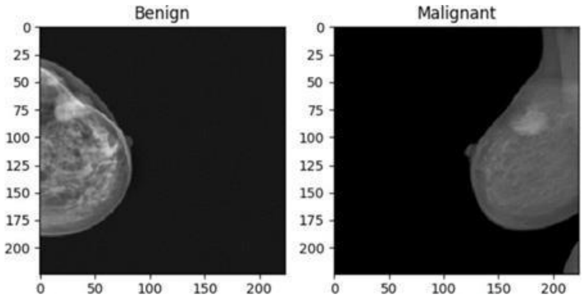


Figure 3. Some images extracted from the augmented INBREAST dataset and their corresponding class labels are displayed.

Figure 3 shows an image for each class of the augmented INBREAST dataset.

### Deep Learning Models

Artificial Intelligence technologies are improving day by day. Deep learning models are beneficial when there is a high level of computational complexity and a need to classify massive datasets. Over the past ten years, deep learning in histopathology has gained interest due to its state-of-the-art performance in tasks including classification and localization. Convolutional neural networks are deep learning frameworks that produce impressive results in tissue image processing. Deep learning makes it possible to learn directly from data. Deep learning is a method of image classification that uses many data to achieve highly successful results using a complex model. In this study, we are therefore using fourteen deep learning models, described in more detail below, to classify images of breast cancer. Apart from the existing models, we developed a new Hybrid Attention VGG model, to increase success in this classification problem.

For this purpose, we used Google Colab and L4 cloud GPU, which has 24 GB VRAM memory for calculation. We set the hyperparameter values as epoch=25 and optimizer="Adam". We used models trained on the ImageNet [19] dataset through fine-tuning and Transfer Learning.

Transfer Learning[20] is a machine learning technique that leverages acquired information from one activity to enhance performance on a related task. This approach has been extensively implemented in diverse domains, including computer vision, natural language processing, and speech recognition.

This study tested different hyperparameter values for both datasets and deep learning models as hyperparameter optimization. We tried a learning rate of 0.001 and 0.0001. We conducted our analysis using batch size values of 16, 32, and 64. To get better results, during the training phase of the model, we assign the same values to the hyperparameters for all models. Within the data splitting stage, a specific portion of the data set, amounting to 15%, is allocated as the test set. The validation set comprises 20% of the data. During the training phase, the models processed the data in PNG image format containing only two class labels: Benign and Malignant. In the study, we classified the characteristics of the tumour images using only images with tumours. We utilized accuracy, precision, recall, f1-measure, Cohen Kappa score, and Roc auc score as performance metrics to compare and evaluate the classification models. In addition, we utilized a confusion matrix to assess the outcomes.

**Vgg-16 :** VGG16 comprises three completely connected layers and thirteen interconnected convolutional layers. The initial two convolutional layers consist of 64 filters, each with a dimension of 3 x 3. Subsequently, there are two further layers, each composed of 128 filters, two layers with 256 filters, two layers with 512 filters, and ultimately, the last layer with 512 filters. Beyond the layers of convolution, there are three fully linked layers, each composed of 4096 neurons. Subsequently, there is a classification layer with 1000 neurons. [21].

**Vgg-19 :** VGG19 comprises 19 layers: 16 convolutional and three fully connected. The first two convolutional layers comprise 64 filters, measuring 3 × 3 pixels. These are then succeeded by two more layers, each containing 128 filters, followed by four layers with 256 filters, four layers with 512 filters, and finally, one layer with 512 filters. After the convolutional layers, there are three fully connected layers, each consisting of 4096 neurons, and a classification output layer with 1000 neurons. [22], [23].

**DenseNet-121 :** DenseNet contains two additional essential blocks in addition to the typical convolutional and pooling layers. The designs of the convolution block, pooling layer, transition layer, and classification layer are all shared by DenseNet's different versions. All DenseNet versions, however, have a distinct set of four DenseBlocks with various repeat times [24]. The first convolutional block consists of 64 filters with dimensions of 7 x 7 and a stride of two. Afterwards, a Max Pooling layer can be constructed with a stride of two and a 3 × 3 max pooling arrangement. A convolutional block is made up of the input layer, that is subsequently followed by the Batch Normalization, ReLu activation, and Conv2D layers. Every dense block has two convolutions with 1 × 1 and 3 × 3 kernel sizes. DenseNet-121 [24] is a specific type of neural network architecture. The blocks denoted as "dense Block1," "dense Block2," "dense Block3," and "dense Block4" are iterated 6, 12, 24, and 16 times, correspondingly.

**DenseNet-169 :** DenseNet-169 [24] is a convolutional neural network architecture. The blocks denoted as "dense Block1," "dense Block2," "dense Block3," and "dense Block4" are replicated 6, 12, 32, and 32 times, respectively.

**ResNet-50 :** ResNet-50 [25] is a deep CNN design that circumvents the vanishing gradient problem by learning from deep networks through residual connections. Its fifty levels include convolutional layers, batch normalization layers, ReLU activation functions, and fully connected layers. Additionally, ResNet50 uses a skip connection to bypass a few network layers and efficiently learns high-level and low-level features.

**ResNet-101 :** The ResNet-101 architecture is composed of 101 layers. Based on the Residual neural network learning approach, this architecture is considered one of the most advanced architectures for ImageNet [19] due to its depth. Compared to other architectures, the primary distinguishing characteristic of Resnet-101 is its optimization of the discrepancies between the input and required convolution qualities. Obtaining desired characteristics is more effortless and effective than obtaining alternative designs. Therefore, residual optimization can be performed to decrease the number of parameters in a more complex network. To attain a more optimal result, it is possible to reduce the number of layers by minimizing the number of parameters [26].

The ResNet architecture incorporates a ResBlock layer to transfer information from the previous layer to the new layer, enabling the learning of information not captured in the prior layer. The ResBlock layer in the Resnet design transmits residual values to the subsequent layer. This skip connection, placed between the weight layers and the Relu activation code at every two-layer activation, modifies the system's output [27].

**MobileNet :** A deep learning architecture called MobileNet [28] is appropriate for quickly and precisely analyzing medical images, particularly regarding BC diagnosis. MobileNet's focus on computational efficiency makes it possible to extract information from mammography images efficiently, facilitating the identification of minute patterns or anomalies linked to breast cancer. MobileNet is perfect for contexts with limited resources since it optimizes memory consumption and computational effort through depthwise separable convolutions. Integrating the ReLU6 activation mechanism further improves efficiency and compatibility with medical imaging devices. MobileNet presents a valuable option for BC analysis, yielding precise outcomes with minimal computing overhead.

**MobileNet-V2 :** MobileNet-V2 [29] builds upon the Depthwise Separable technique used in MobileNetV1 and incorporates the residual structure. It has been discovered that the Rectified Linear Unit (ReLU) leads to significant information loss in feature maps with just a few channels. As a result, linear bottlenecks and inverted residuals were developed as solutions. MobileNet-V2 maintains the structural simplicity of MobileNet-V1, enabling the same level of precision without requiring additional specialised procedures. MobileNet-V2 is specifically designed to investigate the capabilities of neural networks and create a network architecture that is both simple and efficient. The research primarily focuses on two areas: the utilisation of optimization approaches, such as evolutionary algorithms and reinforcement learning, for conducting framework searches [30] and the management of the "BottleNeck" Structure [31]. MobileNet-V2 incorporates two crucial innovations: line bottlenecks and inverted residuals and implementing the 3×3 depth-separable convolution. The linear bottleneck arises due to the linear transformation of the "manifold of interest" region, which may have a non- zero value following the ReLU process. Furthermore, following the ReLU activation function, a portion of the channel information will be discarded. The rationale behind incorporating inverted residuals is that the bottleneck already encompasses the essential information. Hence, the shortcut immediately connects the two bottlenecks. Furthermore, initially increasing the dimensionality, followed by feature extraction and subsequent dimensionality reduction, is employed because of the higher significance of information in the low-dimensional space.

**Inception-V3 :** Inception-V3 model's input layer supports shape images (299, 299, 3). The input image performs spatial dimension reduction and fundamental attribute extraction by utilizing two convolutional layers, max pooling and batch normalization. Convolutional filters and pooling techniques are present in the Inception modules. The outputs of the concatenated modules pass on to the subsequent module. Modules 5 and 11 of the Inception introduce two more classifiers. Each consists of two fully connected layers with ReLU activation, a global average pooling layer, and a classification softmax layer. The ultimate layer includes a global average pooling layer, a classification softmax output layer, and a fully connected layer [32].

**InceptionResNet-V2 :** InceptionResNet-V2 is the combination of the Inception and ResNet networks. In its 164 layers, skip connections improve gradient propagation during training. The stem module, classification layer, and several Inception-ResNet-A, B, and C modules utilize convolutional, pooling, and activation layers to analyse images. The Inception-ResNet-A, B, and C modules gather features at various scales using max pooling and 1 × 1 convolutions. Ultimately, the classification layer generates class predictions using a fully linked layer and a global average pooling layer [33].

**Xception:** Xception is a complex neural network structure that utilizes Depthwise Separable Convolutions. Google employee researchers developed this technology. Google introduced the concept of Inception modules in convolutional neural networks as a transitional stage between ordinary convolution and the depthwise separable convolution operation, which involves doing a depthwise convolution followed by a pointwise convolution [34].

**NasNetMobile :** Nasnet is a convolutional neural network (CNN) architecture designed to be scalable. It is developed using a process called neural architecture search. The architecture comprises fundamental building blocks called cells optimised via reinforcement learning [35]. A cell consists of limited processes, including separable convolutions and pooling, and is iterated numerous times based on the desired network capacity. The mobile version, known as Nasnet-Mobile, comprises 12 cells and has 5.3 million parameters and 564 million multiply-accumulates (MACs).

**EfficientNetV2B0 :** EfficientNetV2B0 [41] is the smallest and most fundamental variant of the EfficientNet [41] series, engineered as a convolutional neural network to optimize efficiency in deep learning applications. The model seeks high accuracy with a reduced number of parameters by the "compound scaling" method, which optimizes the equilibrium among width, depth, and resolution. Moreover, it is organized with Mobile Inverted Bottleneck Conv (MBConv) blocks, which provide rapid and efficient computing.

The model exhibits remarkable efficacy, especially in image classification, object detection, and image segmentation. Employing data augmentation techniques during training enables favourable outcomes even with limited data. EfficientNetV2b0, providing a robust alternative for research and practical applications, has a significant position in deep learning.

**EfficientNetV2L :** EfficientNetV2L [41], a larger and more potent iteration of the EfficientNet series, delivers enhanced performance on intricate jobs owing to its increased number of layers and parameters. EfficientNetV2L employs a "compound scaling" methodology to improve the equilibrium among breadth, depth, and resolution, whereas EfficientNetV2B0 is a more compact model intended for operation with reduced resources. Furthermore, EfficientNetV2L leverages enhanced training methodologies and optimised MBConv blocks, enabling it to attain superior accuracy with extensive datasets. Thus, EfficientNetV2L is superior for large-scale applications, while EfficientNetV2B0 delivers rapid and efficient outcomes with reduced resource demands.

**Hybrid Attention VGG (Proposed Method):**

Initially, in creating this proposed model, VGG-16, VGG-19, ResNet50, and ResNet101 were experimented with as basis models. After comparing the performances of the models produced from these basic models on two datasets, it was decided to choose VGG16 as the base model. When we take the VGG16 model as the base, it has been observed that the Hybrid Attention Network is more successful in most hyperparameter values and both datasets.

**Model Architecture:**

The HybridAttentionVGG model is constructed by extending the pre-trained VGG16 model with the following architecture:

The VGG16 model, in its base form, is utilised without including the completely connected layers at the top. It acts as the feature extractor, using its deep and well-established convolutional layers to capture intricate visual features.

The Hybrid Attention Block is a newly introduced one that follows the VGG16 base architecture. This block applies Global Average Pooling to capture the overall context of the input.

It employs dense layers to acquire channel-wise attention weights highlighting the most significant characteristics.

The original and attention-refined feature maps are combined using a skip link, guaranteeing raw and refined information preservation.

Flatten Layer: Transforms the two-dimensional output from the attention block into a one-dimensional feature vector, which can be used as input for dense layers.

A fully connected layer is included, incorporating Batch Normalization and Dropout techniques to enhance feature learning and regularization.

The output layer comprises a dense layer employing a softmax activation function for classification.

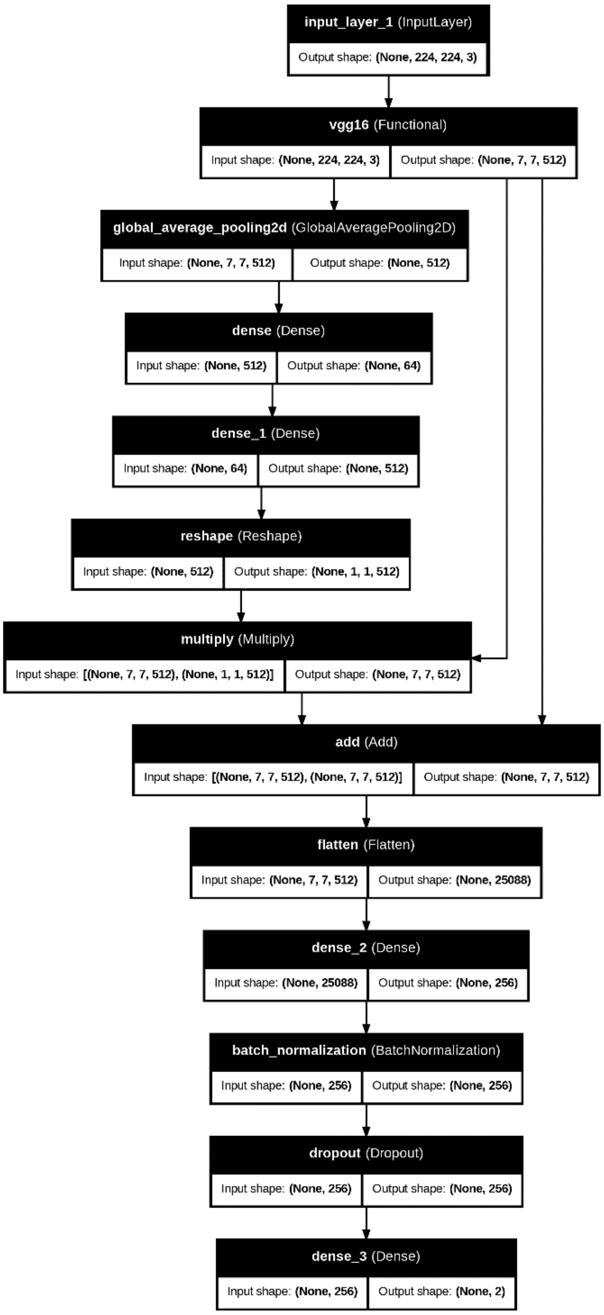

input_layer_1 (InputLayer)
Output shape: (None, 224, 224, 3)
vgg16 (Functional)
Input shape: (None, 224, 224, 3)
Output shape: (None, 7, 7, 512)
global_average_pooling2d (GlobalAveragePooling2D)
Input shape: (None, 7, 7, 512)
Output shape: (None, 512)
dense (Dense)
Input shape: (None, 512)
Output shape: (None, 64)
dense_1 (Dense)
Input shape: (None, 64)
Output shape: (None, 512)
reshape (Reshape)
Input shape: (None, 512)
Output shape: (None, 1, 1, 512)
multiply (Multiply)
Input shape: [(None, 7, 7, 512), (None, 1, 1, 512)]
Output shape: (None, 7, 7, 512)
add (Add)
Input shape: [(None, 7, 7, 512), (None, 7, 7, 512)]
Output shape: (None, 7, 7, 512)
flatten (Flatten)
Input shape: (None, 7, 7, 512)
Output shape: (None, 25088)
dense_2 (Dense)
Input shape: (None, 25088)
Output shape: (None, 256)
batch_normalization (BatchNormalization)
Input shape: (None, 256)
Output shape: (None, 256)
dropout (Dropout)
Input shape: (None, 256)
Output shape: (None, 256)
dense_3 (Dense)
Input shape: (None, 256)
Output shape: (None, 2)

Figure 4. The layers of the proposed model and the connections between the architecture of the proposed model

Figure 4 depicts the structure of this model. The HybridAttentionVGG model improves upon the pre-trained VGG16 architecture by incorporating a novel Hybrid Attention Block to enhance learning features. This block utilizes global average pooling to capture the overall context and employs dense layers to calculate attention weights for each channel, highlighting significant aspects. The model uses a skip connection to combine the initial VGG16 output with the attention-refined features, preserving unprocessed and enhanced information to enhance learning. After that, the model employs flattened and fully linked layers using Batch Normalization and

Dropout to analyse the features further before performing classification. The new integration of attention techniques and classic CNN layers achieves a favourable equilibrium between simplicity and improved performance, rendering it highly suitable for applications requiring concentrated feature extraction.

**Essential Architectural Elements:**

The VGG16 base model is a powerful feature extractor pre-trained on the ImageNet[19] dataset. It is well-regarded for its straightforward design and effectiveness in extracting hierarchical features.

The Hybrid Attention Block is this concept's core breakthrough. The system dynamically adjusts the feature maps, taking into account their significance, and then merges them with the original features using a skip connection.

A dense layer with batch normalization and dropout enhances the recovered features from the attention block and mitigates the risk of overfitting.

The Softmax Output Layer generates probability distributions for classification problems.

**Significance of this Model:**

Blends straightforwardness and ingenuity: This model successfully integrates VGG16's straightforwardness with a lightweight yet potent attention mechanism, effectively closing the divide between user-friendliness and improved performance.

Enhanced Feature Learning: By incorporating an attention mechanism, the model acquires the ability to concentrate on the most significant characteristics. This has the potential to enhance performance on tasks that require highlighting critical aspects, such as object detection and medical imaging.

The model is both flexible and lightweight, as it does not substantially increase computational complexity compared to the base VGG16. This makes it well-suited for applications with limited resources.

**Benefits of HybridAttentionVGG:**

The attention block improves feature representation by selectively emphasising the most pertinent regions of the image. This results in improved generalisation, particularly on intricate datasets where not all features have the same significance level.

The benefits of skip connections lie in the ability to merge raw and refined features, allowing the model to preserve a broader range of information. This reduces the likelihood of losing potentially valuable information that the attention mechanism could overlook.

Regularisation techniques such as Batch Normalization and Dropout are employed in the dense layers to enhance training stability and mitigate overfitting, which is especially crucial in deep learning models.

The use of a pre-trained VGG16 model in transfer learning becomes advantageous, particularly in scenarios with a scarcity of data. The model can rapidly adjust to new tasks by making small adjustments to a smaller number of layers.

Effortlessness and straightforwardness of implementation: The architecture is uncomplicated to execute, alter, and comprehend. This makes it an excellent option for practitioners exploring attention mechanisms without delving into excessively intricate models such as transformers.

**Drawbacks of Hybrid Attention:**

Possible Exaggeration of Specific Features: The attention mechanism can excessively concentrate on some features, potentially disregarding other features that may be less apparent but nonetheless significant for certain activities.

Limitations of Dependency on VGG16: While serving as a robust benchmark, VGG16 is rather outdated compared to more recent architectures like ResNet or EfficientNet. The model may not utilise certain advanced strategies included in such models, such as residual connections or advanced activation functions.

**Utilisation Scenarios and Prospective Implementations:**

Medical imaging: The attention mechanism can be utilised in tasks such as tumour identification to prioritise the most significant regions over others, hence enhancing the focus on the most relevant parts.

Object Detection and Localisation: This approach applies when certain regions of a picture hold more importance than others, such as identifying particular objects in crowded scenes.
Satellite Image Analysis: Attention processes can be used in remote sensing to improve predictions by enhancing the ability to differentiate between small characteristics in huge images.

**Performance Metrics**

Fourteen deep-learning models and the proposed model were analyzed using augmented mammogram images from MIAS and INBREAST datasets. Comparisons with other deep learning models have been made to demonstrate the algorithm's superiority in breast cancer diagnosis. This classification's most commonly used comparison criteria are accuracy, precision, recall, F1-score, Cohen Kappa score, Roc Auc score and confusion matrices.

**Accuracy :** Accuracy is the ratio of correctly identified samples in the evaluation dataset to the total number of samples. This metric is frequently used in machine learning applications in the medical field, but it is also notorious for its potential to mislead when dealing with imbalanced class distributions. This is because reaching high accuracy can be easily accomplished by assigning all samples to the dominant class. The accuracy is limited to a range of 0 to 1. A value of 1 indicates that all positive and negative samples are correctly predicted, while 0 indicates that none of the positive or negative samples are predicted correctly.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

**Precision :** Precision is the measure of the ratio of correctly identified samples to all samples assigned to a particular class, indicating the proportion of relevant retrieved samples. It is a quantitative measure that fluctuates between 0 and 1. A precision score of 1 signifies that all samples in the class were accurately predicted, while a score of 0 indicates that no valid predictions were produced.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

**Recall :** The recall, often referred to as the sensitivity or True Positive Rate (TPR), represents the proportion of positive samples that are accurately classified. It is computed by dividing the number of correctly classified positive samples by the total number allocated to the positive class. The recall metric is defined within the range of [0, 1], with 1 indicating a precise prediction of the positive class and 0 indicating an inaccurate prediction of all positive class samples. This statistic is considered one of the most crucial in medical studies because it aims to minimize missed positive cases, resulting in a high recall rate.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

**F1 Score :** The F1 score is computed by calculating the mean harmonic of precision and recall, resulting in a measure penalizing excessive values of either metric. This metric exhibits asymmetry between the classes, meaning that its value is contingent upon the designation of one class as positive and the other as negative. For instance, if there is a significant positive class and a classifier inclined towards this majority, the F1 score, which is directly related to the true positive (TP) rate, would be high. Modifying the class labels to make the negative class the dominant one and introducing a bias towards the negative class in the classifier will decrease the F1 score, even though there have been no changes in the data or the distribution of classes. The F1-score is constrained within the range of [0, 1], with 1 indicating the highest precision and recall and 0 indicating no precision or recall.

$$F1\,Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

**Cohen Kappa Score :** The reliability of raters for inter-rater and intra-rater agreement in categorising data can be assessed using Cohen's Kappa value (K). Due to its consideration of the possibility of coincidental agreement, most individuals perceive it as a more precise method to measure agreement than a straightforward percentage agreement. While it can be adapted for situations involving more than two raters, it is commonly used in contexts where there are just two raters. In binary classification models, one rater acts as the classification model, while the second is an observer who knows the true classifications for each record or dataset. Cohen's Kappa considers the level of agreement amongst raters in terms of both true positives and negatives, as well as false positives and negatives. Cohen and Kappa can assess overall agreement and agreement by considering random factors. The Cohen's Kappa score (κ) is a metric used to evaluate the performance of classification models by measuring the level of agreement between two raters: a real-world observer and the classification model. It considers both the perfect agreement and the agreement that could occur by chance. Po is the measured level of agreement between the raters. Pe represents the probability of obtaining an agreement by chance.

$$\kappa = \frac{Po - Pe}{1 - Pe} \tag{5}$$

**Roc Auc Score :** The Receiver Operating Characteristics (ROC) is a statistical measure that evaluates the performance of a binary classification model. The ROC curve is a visual depiction of the performance of a binary classification model. ROC is an acronym for receiver operating characteristics. The function visually illustrates the correlation between the true positive rate (TPR) and the false positive rate (FPR) at different categorisation thresholds.

The Area Under Curve (AUC) is a quantitative measure representing the extent of the region bounded by the ROC curve. The metric evaluates the holistic effectiveness of the binary classification model. Since both the true positive rate (TPR) and the false positive rate (FPR) have values between 0 and 1, the area under the curve (AUC) will also always fall within this range. A higher value of AUC indicates superior model performance. The primary objective is to optimize the area to achieve the maximum true positive rate (TPR) and the lowest false positive rate (FPR) at the specified threshold. The AUC quantifies the likelihood that the model would assign a higher predicted probability to a randomly selected positive case than a randomly selected negative instance.

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

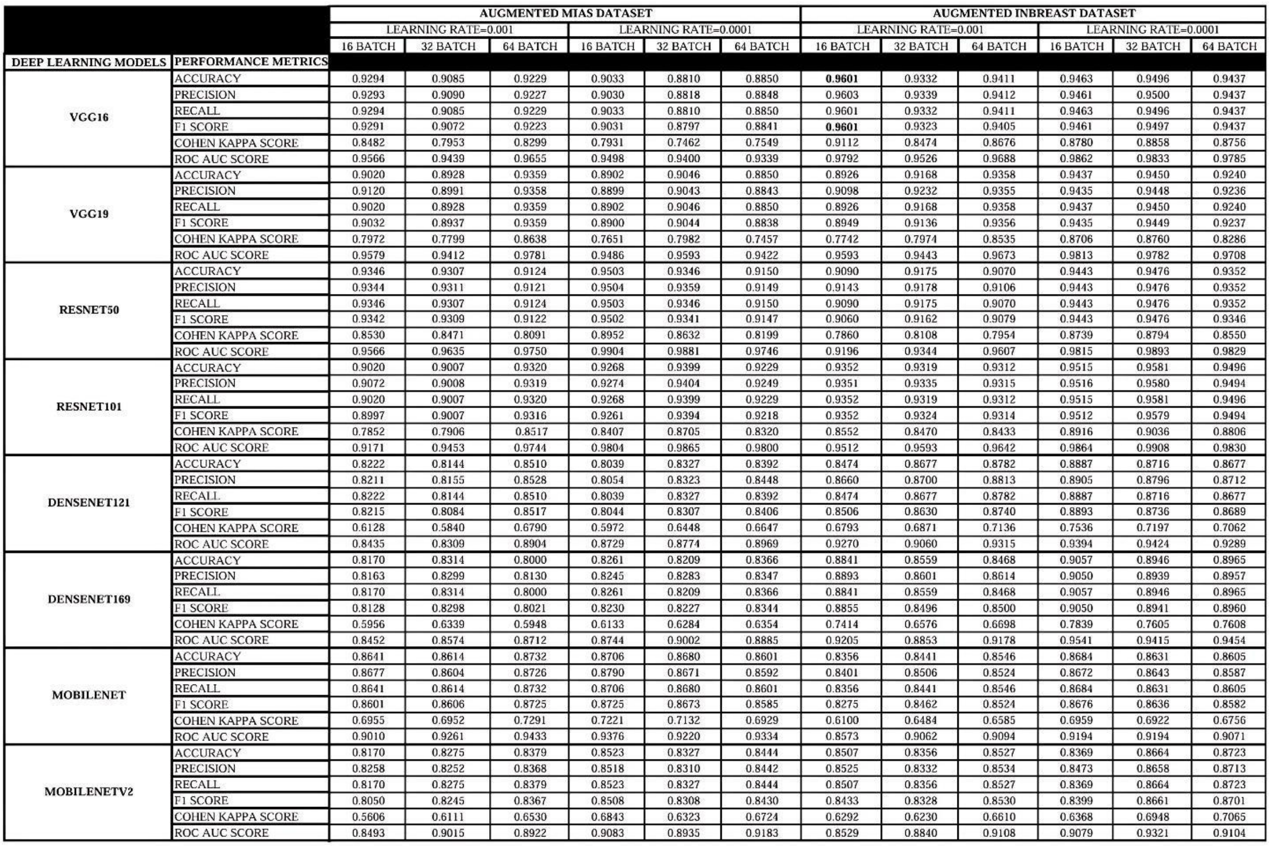
$$FPR = \frac{FP}{FP + \mathrm{T}N} \tag{7}$$

**Confusion Matrix :** Confusion matrices are utilised to assess the efficacy of machine learning algorithms by contrasting their predictions with the data's actual labels. The predictions are organised in a grid structure, with rows representing the predicted classes and columns representing the actual classes. This configuration facilitates the comprehension of the model's performance in classified various classes. The matrix has measures such as True Positive (accurately predicted positive samples), True Negative (accurately predicted negative samples), False Positive (incorrectly classified negative samples), and False

Negative (incorrectly classified positive samples). The selection of the null hypothesis determines how these metrics are interpreted within the matrix.

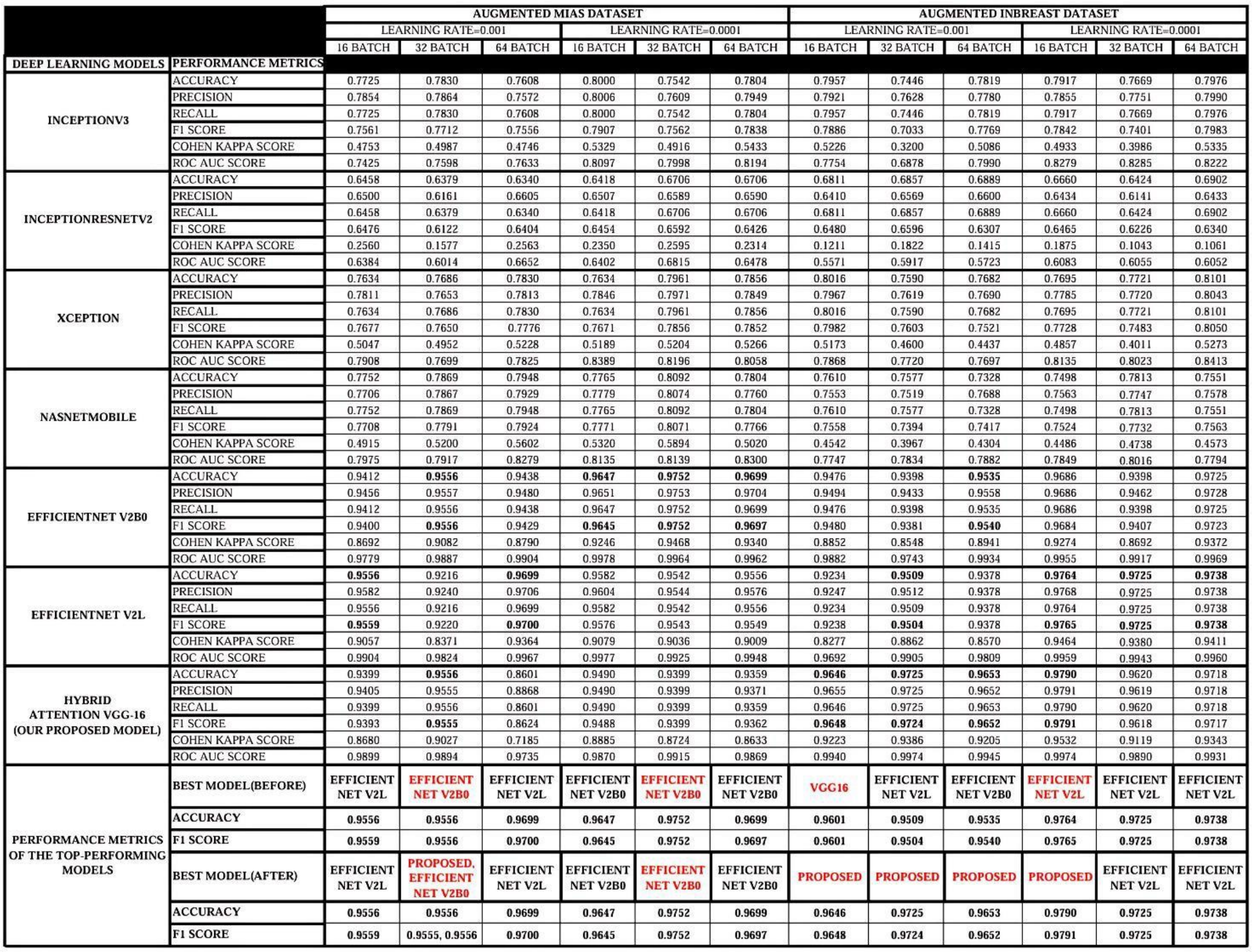
## Results

In our study, the models were trained on two datasets: the Augmented MIAS and the Augmented INBREAST. In addition to the existing fourteen deep learning models, our proposed HybridAttentionVGG model has utilized different hyperparameter values with learning rates of 0.001 and 0.0001, and batch sizes of 16, 32, and 64. The model was trained using the pre-trained ImageNet[19] model by transfer learning, utilizing the Adam optimizer for 25 epochs. The accuracy, precision, recall, F1-score, Cohen's kappa score, and ROC AUC score values of these models are shown in Figure 5 and Figure 6.



| | | AUGMENTED MIAS DATASET | | | | | | AUGMENTED INBREAST DATASET | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LEARNING RATE=0.001 | | | LEARNING RATE=0.0001 | | | LEARNING RATE=0.001 | | | LEARNING RATE=0.0001 | | |
| | | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH |
| DEEP LEARNING MODELS | PERFORMANCE METRICS | | | | | | | | | | | | |
| VGG16 | ACCURACY | 0.9294 | 0.9085 | 0.9229 | 0.9033 | 0.8810 | 0.8850 | **0.9601** | 0.9332 | 0.9411 | 0.9463 | 0.9496 | 0.9437 |
| | PRECISION | 0.9293 | 0.9090 | 0.9227 | 0.9030 | 0.8818 | 0.8848 | 0.9603 | 0.9339 | 0.9412 | 0.9461 | 0.9500 | 0.9437 |
| | RECALL | 0.9294 | 0.9085 | 0.9229 | 0.9033 | 0.8810 | 0.8850 | 0.9601 | 0.9332 | 0.9411 | 0.9463 | 0.9496 | 0.9437 |
| | F1 SCORE | 0.9291 | 0.9072 | 0.9223 | 0.9031 | 0.8797 | 0.8841 | **0.9601** | 0.9323 | 0.9405 | 0.9461 | 0.9497 | 0.9437 |
| | COHEN KAPPA SCORE | 0.8482 | 0.7953 | 0.8299 | 0.7931 | 0.7462 | 0.7549 | 0.9112 | 0.8474 | 0.8676 | 0.8780 | 0.8858 | 0.8756 |
| | ROC AUC SCORE | 0.9566 | 0.9439 | 0.9655 | 0.9498 | 0.9400 | 0.9339 | 0.9792 | 0.9526 | 0.9688 | 0.9862 | 0.9833 | 0.9785 |
| VGG19 | ACCURACY | 0.9020 | 0.8928 | 0.9359 | 0.8902 | 0.9046 | 0.8850 | 0.8926 | 0.9168 | 0.9358 | 0.9437 | 0.9450 | 0.9240 |
| | PRECISION | 0.9120 | 0.8991 | 0.9358 | 0.8899 | 0.9043 | 0.8843 | 0.9098 | 0.9232 | 0.9355 | 0.9435 | 0.9448 | 0.9236 |
| | RECALL | 0.9020 | 0.8928 | 0.9359 | 0.8902 | 0.9046 | 0.8850 | 0.8926 | 0.9168 | 0.9358 | 0.9437 | 0.9450 | 0.9240 |
| | F1 SCORE | 0.9032 | 0.8937 | 0.9359 | 0.8900 | 0.9044 | 0.8838 | 0.8949 | 0.9136 | 0.9356 | 0.9435 | 0.9449 | 0.9237 |
| | COHEN KAPPA SCORE | 0.7972 | 0.7799 | 0.8638 | 0.7651 | 0.7982 | 0.7457 | 0.7742 | 0.7974 | 0.8535 | 0.8706 | 0.8760 | 0.8286 |
| | ROC AUC SCORE | 0.9579 | 0.9412 | 0.9781 | 0.9486 | 0.9593 | 0.9422 | 0.9593 | 0.9443 | 0.9673 | 0.9813 | 0.9782 | 0.9708 |
| RESNET50 | ACCURACY | 0.9346 | 0.9307 | 0.9124 | 0.9503 | 0.9346 | 0.9150 | 0.9090 | 0.9175 | 0.9070 | 0.9443 | 0.9476 | 0.9352 |
| | PRECISION | 0.9344 | 0.9311 | 0.9121 | 0.9504 | 0.9359 | 0.9149 | 0.9143 | 0.9178 | 0.9106 | 0.9443 | 0.9476 | 0.9352 |
| | RECALL | 0.9346 | 0.9307 | 0.9124 | 0.9503 | 0.9346 | 0.9150 | 0.9090 | 0.9175 | 0.9070 | 0.9443 | 0.9476 | 0.9352 |
| | F1 SCORE | 0.9342 | 0.9309 | 0.9122 | 0.9502 | 0.9341 | 0.9147 | 0.9060 | 0.9162 | 0.9079 | 0.9443 | 0.9476 | 0.9346 |
| | COHEN KAPPA SCORE | 0.8530 | 0.8471 | 0.8091 | 0.8952 | 0.8632 | 0.8199 | 0.7860 | 0.8108 | 0.7954 | 0.8739 | 0.8794 | 0.8550 |
| | ROC AUC SCORE | 0.9566 | 0.9635 | 0.9750 | 0.9904 | 0.9881 | 0.9746 | 0.9196 | 0.9344 | 0.9607 | 0.9815 | 0.9893 | 0.9829 |
| RESNET101 | ACCURACY | 0.9020 | 0.9007 | 0.9320 | 0.9268 | 0.9399 | 0.9229 | 0.9352 | 0.9319 | 0.9312 | 0.9515 | 0.9581 | 0.9496 |
| | PRECISION | 0.9072 | 0.9008 | 0.9319 | 0.9274 | 0.9404 | 0.9249 | 0.9351 | 0.9335 | 0.9315 | 0.9516 | 0.9580 | 0.9494 |
| | RECALL | 0.9020 | 0.9007 | 0.9320 | 0.9268 | 0.9399 | 0.9229 | 0.9352 | 0.9319 | 0.9312 | 0.9515 | 0.9581 | 0.9496 |
| | F1 SCORE | 0.8997 | 0.9007 | 0.9316 | 0.9261 | 0.9394 | 0.9218 | 0.9352 | 0.9324 | 0.9314 | 0.9512 | 0.9579 | 0.9494 |
| | COHEN KAPPA SCORE | 0.7852 | 0.7906 | 0.8517 | 0.8407 | 0.8705 | 0.8320 | 0.8552 | 0.8470 | 0.8433 | 0.8916 | 0.9036 | 0.8806 |
| | ROC AUC SCORE | 0.9171 | 0.9453 | 0.9744 | 0.9804 | 0.9865 | 0.9800 | 0.9512 | 0.9593 | 0.9642 | 0.9864 | 0.9908 | 0.9830 |
| DENSENET121 | ACCURACY | 0.8222 | 0.8144 | 0.8510 | 0.8039 | 0.8327 | 0.8392 | 0.8474 | 0.8677 | 0.8782 | 0.8887 | 0.8716 | 0.8677 |
| | PRECISION | 0.8211 | 0.8155 | 0.8528 | 0.8054 | 0.8323 | 0.8448 | 0.8660 | 0.8700 | 0.8813 | 0.8905 | 0.8796 | 0.8712 |
| | RECALL | 0.8222 | 0.8144 | 0.8510 | 0.8039 | 0.8327 | 0.8392 | 0.8474 | 0.8677 | 0.8782 | 0.8887 | 0.8716 | 0.8677 |
| | F1 SCORE | 0.8215 | 0.8084 | 0.8517 | 0.8044 | 0.8307 | 0.8406 | 0.8506 | 0.8630 | 0.8740 | 0.8893 | 0.8736 | 0.8689 |
| | COHEN KAPPA SCORE | 0.6128 | 0.5840 | 0.6790 | 0.5972 | 0.6448 | 0.6647 | 0.6793 | 0.6871 | 0.7136 | 0.7536 | 0.7197 | 0.7062 |
| | ROC AUC SCORE | 0.8435 | 0.8309 | 0.8904 | 0.8729 | 0.8774 | 0.8969 | 0.9270 | 0.9060 | 0.9315 | 0.9394 | 0.9424 | 0.9289 |
| DENSENET169 | ACCURACY | 0.8170 | 0.8314 | 0.8000 | 0.8261 | 0.8209 | 0.8366 | 0.8841 | 0.8559 | 0.8468 | 0.9057 | 0.8946 | 0.8965 |
| | PRECISION | 0.8163 | 0.8299 | 0.8130 | 0.8245 | 0.8283 | 0.8347 | 0.8893 | 0.8601 | 0.8614 | 0.9050 | 0.8939 | 0.8957 |
| | RECALL | 0.8170 | 0.8314 | 0.8000 | 0.8261 | 0.8209 | 0.8366 | 0.8841 | 0.8559 | 0.8468 | 0.9057 | 0.8946 | 0.8965 |
| | F1 SCORE | 0.8128 | 0.8298 | 0.8021 | 0.8230 | 0.8227 | 0.8344 | 0.8855 | 0.8496 | 0.8500 | 0.9050 | 0.8941 | 0.8960 |
| | COHEN KAPPA SCORE | 0.5956 | 0.6339 | 0.5948 | 0.6133 | 0.6284 | 0.6354 | 0.7414 | 0.6576 | 0.6698 | 0.7839 | 0.7605 | 0.7608 |
| | ROC AUC SCORE | 0.8452 | 0.8574 | 0.8712 | 0.8744 | 0.9002 | 0.8885 | 0.9205 | 0.8853 | 0.9178 | 0.9541 | 0.9415 | 0.9454 |
| MOBILENET | ACCURACY | 0.8641 | 0.8614 | 0.8732 | 0.8706 | 0.8680 | 0.8601 | 0.8356 | 0.8441 | 0.8546 | 0.8684 | 0.8631 | 0.8605 |
| | PRECISION | 0.8677 | 0.8604 | 0.8726 | 0.8790 | 0.8671 | 0.8592 | 0.8401 | 0.8506 | 0.8524 | 0.8672 | 0.8643 | 0.8587 |
| | RECALL | 0.8641 | 0.8614 | 0.8732 | 0.8706 | 0.8680 | 0.8601 | 0.8356 | 0.8441 | 0.8546 | 0.8684 | 0.8631 | 0.8605 |
| | F1 SCORE | 0.8601 | 0.8606 | 0.8725 | 0.8725 | 0.8673 | 0.8585 | 0.8275 | 0.8462 | 0.8524 | 0.8676 | 0.8636 | 0.8582 |
| | COHEN KAPPA SCORE | 0.6955 | 0.6952 | 0.7291 | 0.7221 | 0.7132 | 0.6929 | 0.6100 | 0.6484 | 0.6585 | 0.6959 | 0.6922 | 0.6756 |
| | ROC AUC SCORE | 0.9010 | 0.9261 | 0.9433 | 0.9376 | 0.9220 | 0.9334 | 0.8573 | 0.9062 | 0.9094 | 0.9194 | 0.9194 | 0.9071 |
| MOBILENETV2 | ACCURACY | 0.8170 | 0.8275 | 0.8379 | 0.8523 | 0.8327 | 0.8444 | 0.8507 | 0.8356 | 0.8527 | 0.8369 | 0.8664 | 0.8723 |
| | PRECISION | 0.8258 | 0.8252 | 0.8368 | 0.8518 | 0.8310 | 0.8442 | 0.8525 | 0.8332 | 0.8534 | 0.8473 | 0.8658 | 0.8713 |
| | RECALL | 0.8170 | 0.8275 | 0.8379 | 0.8523 | 0.8327 | 0.8444 | 0.8507 | 0.8356 | 0.8527 | 0.8369 | 0.8664 | 0.8723 |
| | F1 SCORE | 0.8050 | 0.8245 | 0.8367 | 0.8508 | 0.8308 | 0.8430 | 0.8433 | 0.8328 | 0.8530 | 0.8399 | 0.8661 | 0.8701 |
| | COHEN KAPPA SCORE | 0.5606 | 0.6111 | 0.6530 | 0.6843 | 0.6323 | 0.6724 | 0.6292 | 0.6230 | 0.6610 | 0.6368 | 0.6948 | 0.7065 |
| | ROC AUC SCORE | 0.8493 | 0.9015 | 0.8922 | 0.9083 | 0.8935 | 0.9183 | 0.8529 | 0.8840 | 0.9108 | 0.9079 | 0.9321 | 0.9104 |

Figure 5. The performance metrics of the first eight deep learning models trained on two different datasets and the different hyperparameter values of these models.



| | | AUGMENTED MIAS DATASET | | | | | | AUGMENTED INBREAST DATASET | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LEARNING RATE=0.001 | | | LEARNING RATE=0.0001 | | | LEARNING RATE=0.001 | | | LEARNING RATE=0.0001 | | |
| | | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH |
| DEEP LEARNING MODELS | PERFORMANCE METRICS | | | | | | | | | | | | |
| INCEPTIONV3 | ACCURACY | 0.7725 | 0.7830 | 0.7608 | 0.8000 | 0.7542 | 0.7804 | 0.7957 | 0.7446 | 0.7819 | 0.7917 | 0.7669 | 0.7976 |
| | PRECISION | 0.7854 | 0.7864 | 0.7572 | 0.8006 | 0.7609 | 0.7949 | 0.7921 | 0.7628 | 0.7780 | 0.7855 | 0.7751 | 0.7990 |
| | RECALL | 0.7725 | 0.7830 | 0.7608 | 0.8000 | 0.7542 | 0.7804 | 0.7957 | 0.7446 | 0.7819 | 0.7917 | 0.7669 | 0.7976 |
| | F1 SCORE | 0.7561 | 0.7712 | 0.7556 | 0.7907 | 0.7562 | 0.7838 | 0.7886 | 0.7033 | 0.7769 | 0.7842 | 0.7401 | 0.7983 |
| | COHEN KAPPA SCORE | 0.4753 | 0.4987 | 0.4746 | 0.5329 | 0.4916 | 0.5433 | 0.5226 | 0.3200 | 0.5086 | 0.4933 | 0.3986 | 0.5335 |
| | ROC AUC SCORE | 0.7425 | 0.7598 | 0.7633 | 0.8097 | 0.7998 | 0.8194 | 0.7754 | 0.6878 | 0.7990 | 0.8279 | 0.8285 | 0.8222 |
| INCEPTIONRESNETV2 | ACCURACY | 0.6458 | 0.6379 | 0.6340 | 0.6418 | 0.6706 | 0.6706 | 0.6811 | 0.6857 | 0.6889 | 0.6660 | 0.6424 | 0.6902 |
| | PRECISION | 0.6500 | 0.6161 | 0.6605 | 0.6507 | 0.6589 | 0.6590 | 0.6410 | 0.6569 | 0.6600 | 0.6434 | 0.6141 | 0.6433 |
| | RECALL | 0.6458 | 0.6379 | 0.6340 | 0.6418 | 0.6706 | 0.6706 | 0.6811 | 0.6857 | 0.6889 | 0.6660 | 0.6424 | 0.6902 |
| | F1 SCORE | 0.6476 | 0.6122 | 0.6404 | 0.6454 | 0.6592 | 0.6426 | 0.6480 | 0.6596 | 0.6307 | 0.6465 | 0.6226 | 0.6340 |
| | COHEN KAPPA SCORE | 0.2560 | 0.1577 | 0.2563 | 0.2350 | 0.2595 | 0.2314 | 0.1211 | 0.1822 | 0.1415 | 0.1875 | 0.1043 | 0.1061 |
| | ROC AUC SCORE | 0.6384 | 0.6014 | 0.6652 | 0.6402 | 0.6815 | 0.6478 | 0.5571 | 0.5917 | 0.5723 | 0.6083 | 0.6055 | 0.6052 |
| XCEPTION | ACCURACY | 0.7634 | 0.7686 | 0.7830 | 0.7634 | 0.7961 | 0.7856 | 0.8016 | 0.7590 | 0.7682 | 0.7695 | 0.7721 | 0.8101 |
| | PRECISION | 0.7811 | 0.7653 | 0.7813 | 0.7846 | 0.7971 | 0.7849 | 0.7967 | 0.7619 | 0.7690 | 0.7785 | 0.7720 | 0.8043 |
| | RECALL | 0.7634 | 0.7686 | 0.7830 | 0.7634 | 0.7961 | 0.7856 | 0.8016 | 0.7590 | 0.7682 | 0.7695 | 0.7721 | 0.8101 |
| | F1 SCORE | 0.7677 | 0.7650 | 0.7776 | 0.7671 | 0.7856 | 0.7852 | 0.7982 | 0.7603 | 0.7521 | 0.7728 | 0.7483 | 0.8050 |
| | COHEN KAPPA SCORE | 0.5047 | 0.4952 | 0.5228 | 0.5189 | 0.5204 | 0.5266 | 0.5173 | 0.4600 | 0.4437 | 0.4857 | 0.4011 | 0.5273 |
| | ROC AUC SCORE | 0.7908 | 0.7699 | 0.7825 | 0.8389 | 0.8196 | 0.8058 | 0.7868 | 0.7720 | 0.7697 | 0.8135 | 0.8023 | 0.8413 |
| NASNETMOBILE | ACCURACY | 0.7752 | 0.7869 | 0.7948 | 0.7765 | 0.8092 | 0.7804 | 0.7610 | 0.7577 | 0.7328 | 0.7498 | 0.7813 | 0.7551 |
| | PRECISION | 0.7706 | 0.7867 | 0.7929 | 0.7779 | 0.8074 | 0.7760 | 0.7553 | 0.7519 | 0.7688 | 0.7563 | 0.7747 | 0.7578 |
| | RECALL | 0.7752 | 0.7869 | 0.7948 | 0.7765 | 0.8092 | 0.7804 | 0.7610 | 0.7577 | 0.7328 | 0.7498 | 0.7813 | 0.7551 |
| | F1 SCORE | 0.7708 | 0.7791 | 0.7924 | 0.7771 | 0.8071 | 0.7766 | 0.7558 | 0.7394 | 0.7417 | 0.7524 | 0.7732 | 0.7563 |
| | COHEN KAPPA SCORE | 0.4915 | 0.5200 | 0.5602 | 0.5320 | 0.5894 | 0.5020 | 0.4542 | 0.3967 | 0.4304 | 0.4486 | 0.4738 | 0.4573 |
| | ROC AUC SCORE | 0.7975 | 0.7917 | 0.8279 | 0.8135 | 0.8139 | 0.8300 | 0.7747 | 0.7834 | 0.7882 | 0.7849 | 0.8016 | 0.7794 |
| EFFICIENTNET V2B0 | ACCURACY | 0.9412 | **0.9556** | 0.9438 | **0.9647** | **0.9752** | **0.9699** | 0.9476 | 0.9398 | **0.9535** | 0.9686 | 0.9398 | 0.9725 |
| | PRECISION | 0.9456 | 0.9557 | 0.9480 | 0.9651 | 0.9753 | 0.9704 | 0.9494 | 0.9433 | 0.9558 | 0.9686 | 0.9462 | 0.9728 |
| | RECALL | 0.9412 | 0.9556 | 0.9438 | 0.9647 | 0.9752 | 0.9699 | 0.9476 | 0.9398 | 0.9535 | 0.9686 | 0.9398 | 0.9725 |
| | F1 SCORE | 0.9400 | **0.9556** | 0.9429 | **0.9645** | **0.9752** | **0.9697** | 0.9480 | 0.9381 | **0.9540** | 0.9684 | 0.9407 | 0.9723 |
| | COHEN KAPPA SCORE | 0.8692 | 0.9082 | 0.8790 | 0.9246 | 0.9468 | 0.9340 | 0.8852 | 0.8548 | 0.8941 | 0.9274 | 0.8692 | 0.9372 |
| | ROC AUC SCORE | 0.9779 | 0.9887 | 0.9904 | 0.9978 | 0.9964 | 0.9962 | 0.9882 | 0.9743 | 0.9934 | 0.9955 | 0.9917 | 0.9969 |
| EFFICIENTNET V2L | ACCURACY | **0.9556** | 0.9216 | **0.9699** | 0.9582 | 0.9542 | 0.9556 | 0.9234 | **0.9509** | 0.9378 | **0.9764** | **0.9725** | **0.9738** |
| | PRECISION | 0.9582 | 0.9240 | 0.9706 | 0.9604 | 0.9544 | 0.9576 | 0.9247 | 0.9512 | 0.9378 | 0.9768 | 0.9725 | 0.9738 |
| | RECALL | 0.9556 | 0.9216 | 0.9699 | 0.9582 | 0.9542 | 0.9556 | 0.9234 | 0.9509 | 0.9378 | 0.9764 | 0.9725 | 0.9738 |
| | F1 SCORE | **0.9559** | 0.9220 | **0.9700** | 0.9576 | 0.9543 | 0.9549 | 0.9238 | **0.9504** | 0.9378 | **0.9765** | **0.9725** | **0.9738** |
| | COHEN KAPPA SCORE | 0.9057 | 0.8371 | 0.9364 | 0.9079 | 0.9036 | 0.9009 | 0.8277 | 0.8862 | 0.8570 | 0.9464 | 0.9380 | 0.9411 |
| | ROC AUC SCORE | 0.9904 | 0.9824 | 0.9967 | 0.9977 | 0.9925 | 0.9948 | 0.9692 | 0.9905 | 0.9809 | 0.9959 | 0.9943 | 0.9960 |
| HYBRID ATTENTION VGG-16 (OUR PROPOSED MODEL) | ACCURACY | 0.9399 | **0.9556** | 0.8601 | 0.9490 | 0.9399 | 0.9359 | **0.9646** | **0.9725** | **0.9653** | **0.9790** | 0.9620 | 0.9718 |
| | PRECISION | 0.9405 | 0.9555 | 0.8868 | 0.9490 | 0.9399 | 0.9371 | 0.9655 | 0.9725 | 0.9652 | 0.9791 | 0.9619 | 0.9718 |
| | RECALL | 0.9399 | 0.9556 | 0.8601 | 0.9490 | 0.9399 | 0.9359 | 0.9646 | 0.9725 | 0.9653 | 0.9790 | 0.9620 | 0.9718 |
| | F1 SCORE | 0.9393 | **0.9555** | 0.8624 | 0.9488 | 0.9399 | 0.9362 | **0.9648** | **0.9724** | **0.9652** | **0.9791** | 0.9618 | 0.9717 |
| | COHEN KAPPA SCORE | 0.8680 | 0.9027 | 0.7185 | 0.8885 | 0.8724 | 0.8633 | 0.9223 | 0.9386 | 0.9205 | 0.9532 | 0.9119 | 0.9343 |
| | ROC AUC SCORE | 0.9899 | 0.9894 | 0.9735 | 0.9870 | 0.9915 | 0.9869 | 0.9940 | 0.9974 | 0.9945 | 0.9974 | 0.9890 | 0.9931 |
| PERFORMANCE METRICS OF THE TOP-PERFORMING MODELS | **BEST MODEL(BEFORE)** | **EFFICIENT NET V2L** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2L** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2B0** | **VGG16** | **EFFICIENT NET V2L** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2L** | **EFFICIENT NET V2L** | **EFFICIENT NET V2L** |
| | **ACCURACY** | **0.9556** | **0.9556** | **0.9699** | **0.9647** | **0.9752** | **0.9699** | **0.9601** | **0.9509** | **0.9535** | **0.9764** | **0.9725** | **0.9738** |
| | **F1 SCORE** | **0.9559** | **0.9556** | **0.9700** | **0.9645** | **0.9752** | **0.9697** | **0.9601** | **0.9504** | **0.9540** | **0.9765** | **0.9725** | **0.9738** |
| | **BEST MODEL(AFTER)** | **EFFICIENT NET V2L** | **PROPOSED, EFFICIENT NET V2B0** | **EFFICIENT NET V2L** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2B0** | **EFFICIENT NET V2B0** | **PROPOSED** | **PROPOSED** | **PROPOSED** | **PROPOSED** | **EFFICIENT NET V2L** | **EFFICIENT NET V2L** |
| | **ACCURACY** | **0.9556** | **0.9556** | **0.9699** | **0.9647** | **0.9752** | **0.9699** | **0.9646** | **0.9725** | **0.9653** | **0.9790** | **0.9725** | **0.9738** |
| | **F1 SCORE** | **0.9559** | **0.9555, 0.9556** | **0.9700** | **0.9645** | **0.9752** | **0.9697** | **0.9648** | **0.9724** | **0.9652** | **0.9791** | **0.9725** | **0.9738** |

Figure 6. The performance metrics of six deep learning models trained later and the performance metrics of our proposed Hybrid Attention VGG model, comparing all models to determine some of the best models before and after our proposed model, as well as the Accuracy and F1-score values of these models.

Figure 6 displays the preceding and subsequent models that achieved the some of the highest level of success, alongside the Accuracy and F1-Score values employed for their evaluation and comparison. The models highlighted in red represent our models and some of the most influential models for each dataset.

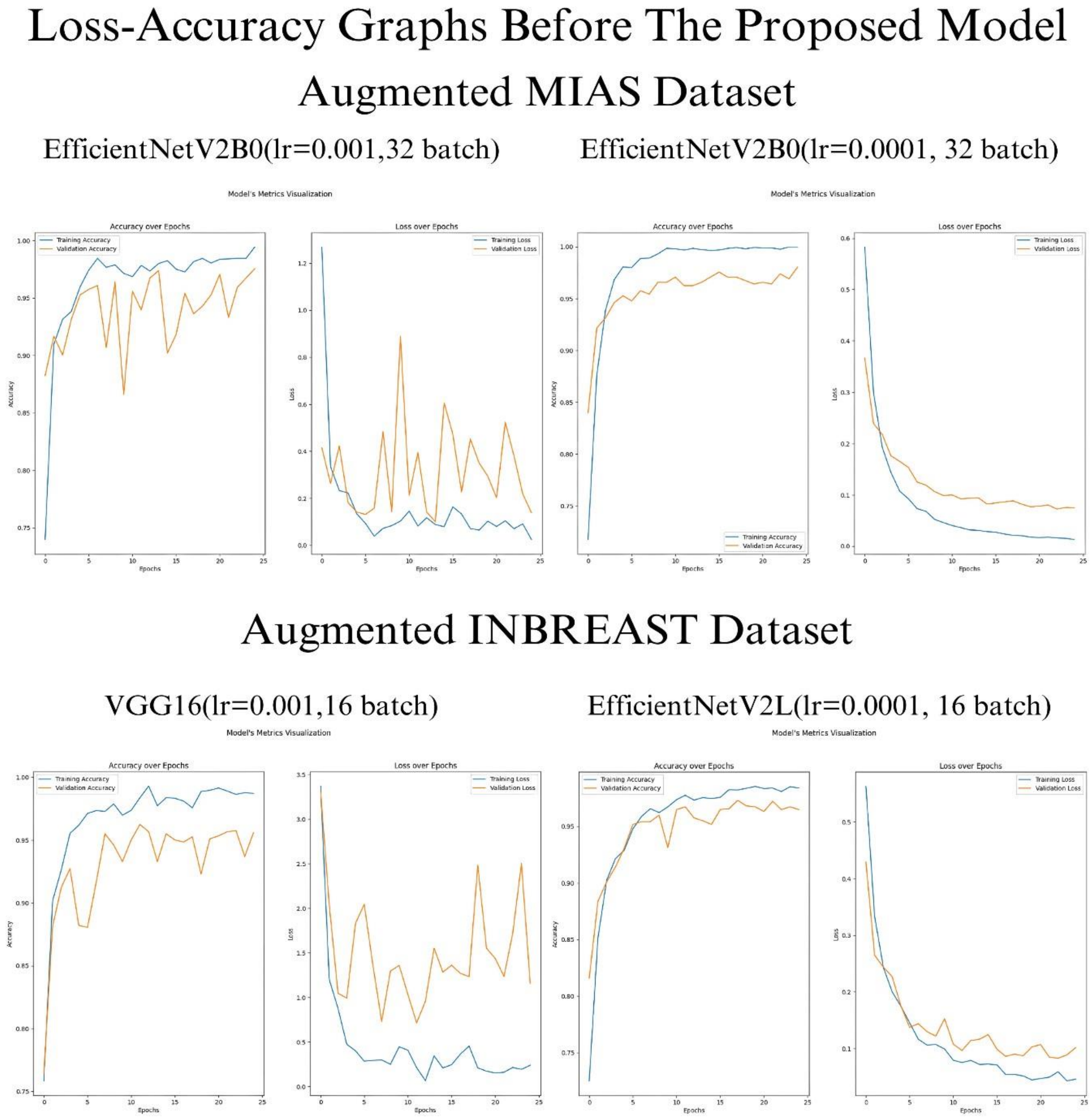


Figure 7. Before the proposed model, the loss and accuracy graphs of some of the most successful models based on the learning rate as a hyperparameter for two different datasets.

In Figure 7, the loss and accuracy values for training and validation are presented for some of the most successful models among the fourteen deep learning models before the proposed model, in both datasets and learning rates.

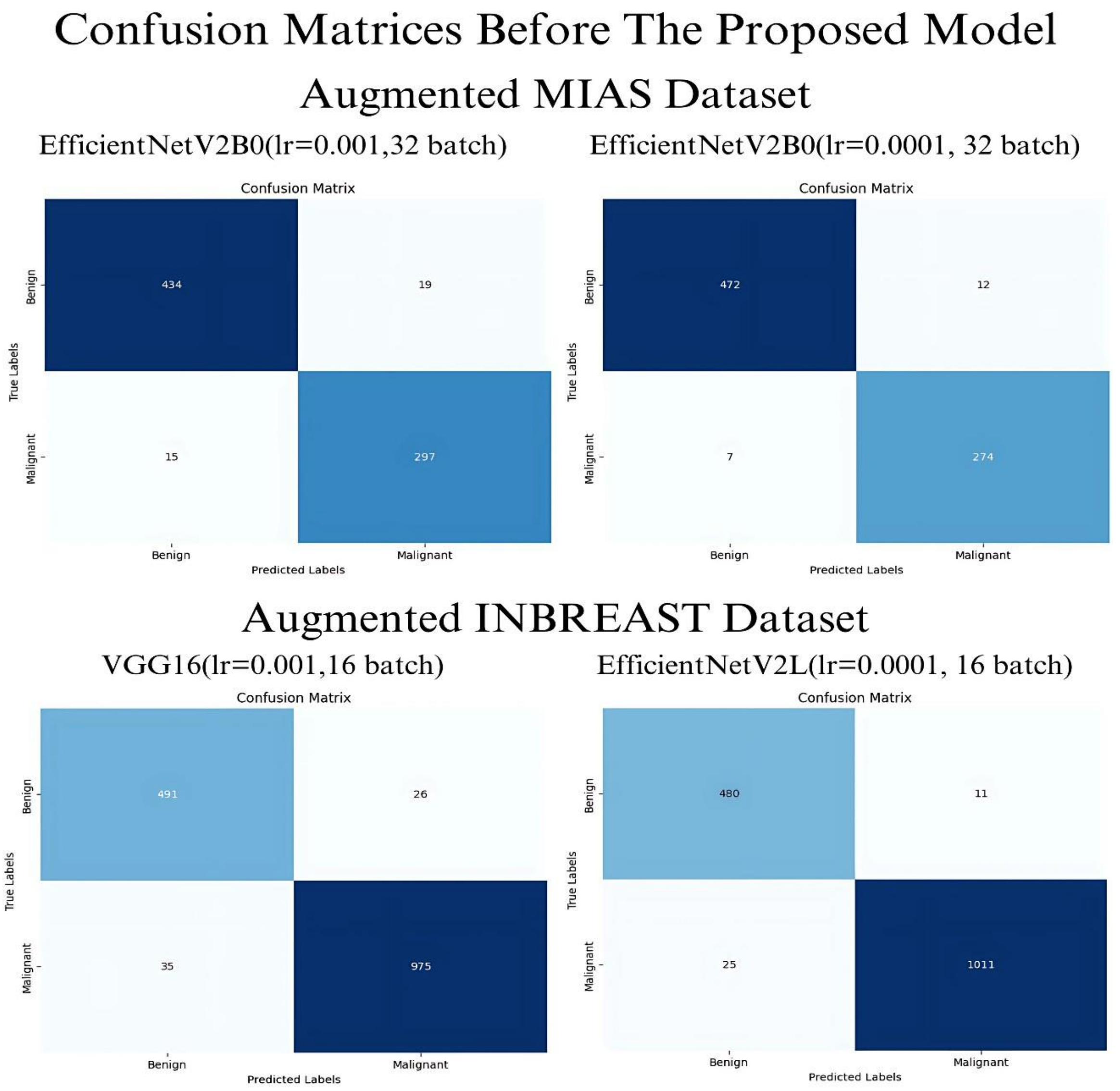


Figure 8. Before the suggested method, the confusion matrices of some of the most influential models were analyzed using the learning rate as a hyperparameter for two distinct datasets.

Figure 8 shows the confusion matrix values for some of the most successful models among the fourteen deep learning models for both datasets and the learning rate before the proposed model.

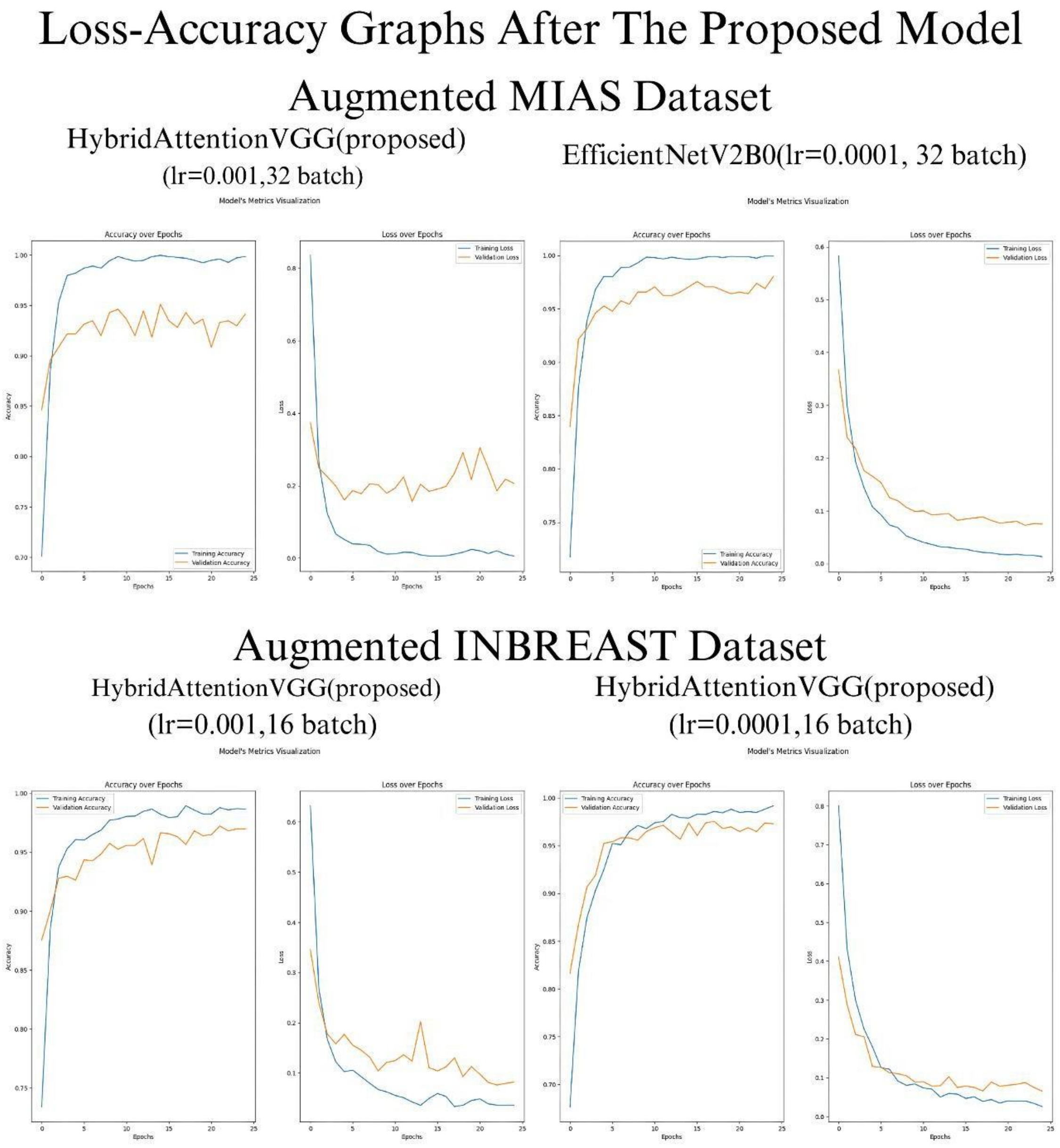


Figure 9. After the proposed model, the loss and accuracy graphs of some of the most successful models based on the learning rate as a hyperparameter for two different datasets.

Figure 9 displays the loss and accuracy values for training and validation of some of the most successful models out of fifteen deep learning models. The proposed model was incorporated following training on both datasets with diverse hyperparameter values, including the learning rate. The graph presents these values based on the number of epochs.

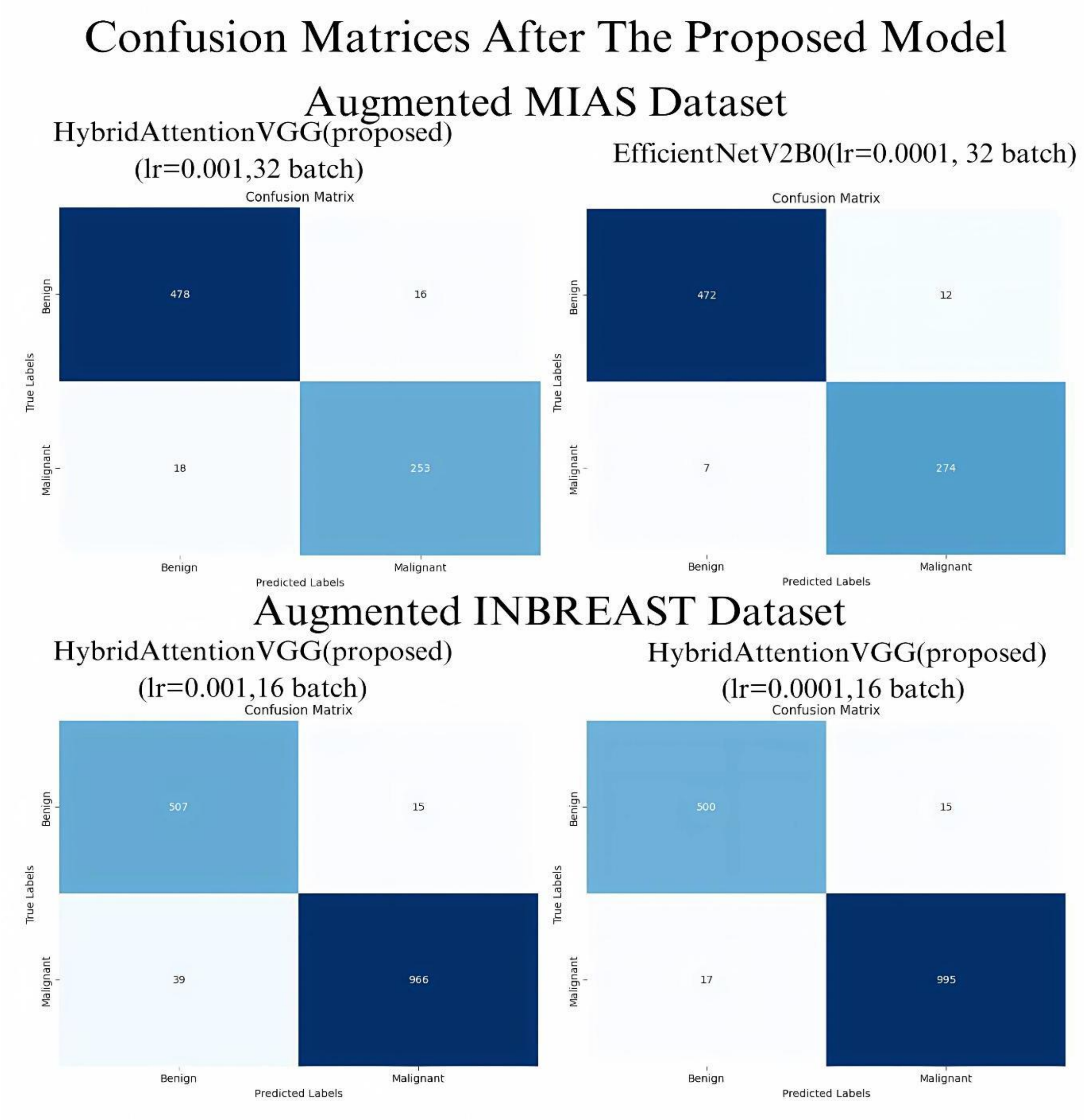


Figure 10. Following the suggested methodology, the confusion matrices of some of the best effective models are presented for two distinct datasets, considering the learning rate as a hyperparameter.

The confusion matrix values for some of the best-performing models out of fifteen deep-learning models are shown in Figure 10. These values were obtained by including the suggested model in both datasets and varying the learning rate.

**GRAD-CAM:** GRAD-CAM(Gradient-weighted Class Activation Mapping) is a method employed to visualize the regions of interest that a deep learning model prioritizes during decision-making, particularly in image classification tasks. It enhances the transparency of the model's decision-making process by indicating which features influenced the classification of a specific class. GRAD-CAM enhances model reliability by examining accurate and inaccurate classifications, pinpointing areas of focus, and resolving potential problems. In medical image analysis, it can improve clinical applications by confirming whether the model is concentrating on pertinent areas.

GRAD-CAM utilizes gradients from the last convolutional layer to produce a heatmap that emphasises the pixels deemed significant by the model. The places that significantly influence the model's choice are represented in warm hues (red, orange), whilst regions of lesser importance are depicted in cooler tones (blue). The results assess the model's accuracy in classifications and its attention to significant regions within the image.

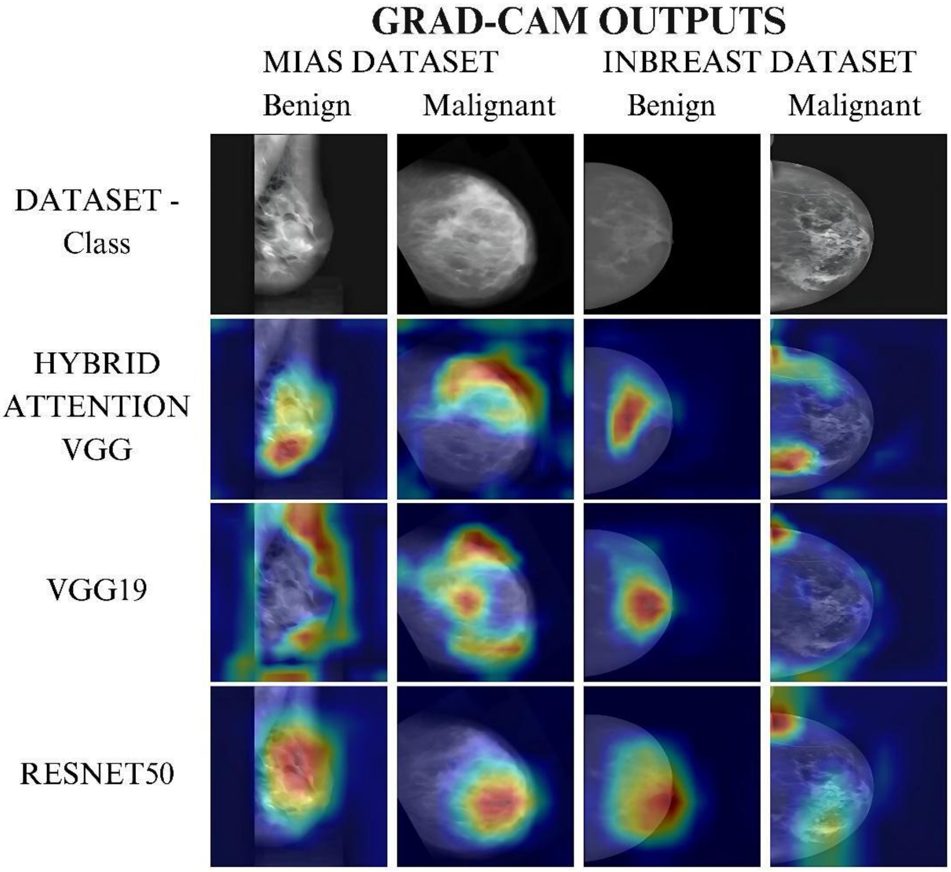


Figure 11. GRAD-CAM outputs of mammography images from both categories in two datasets for the VGG19, RESNET50 models, and our suggested model.

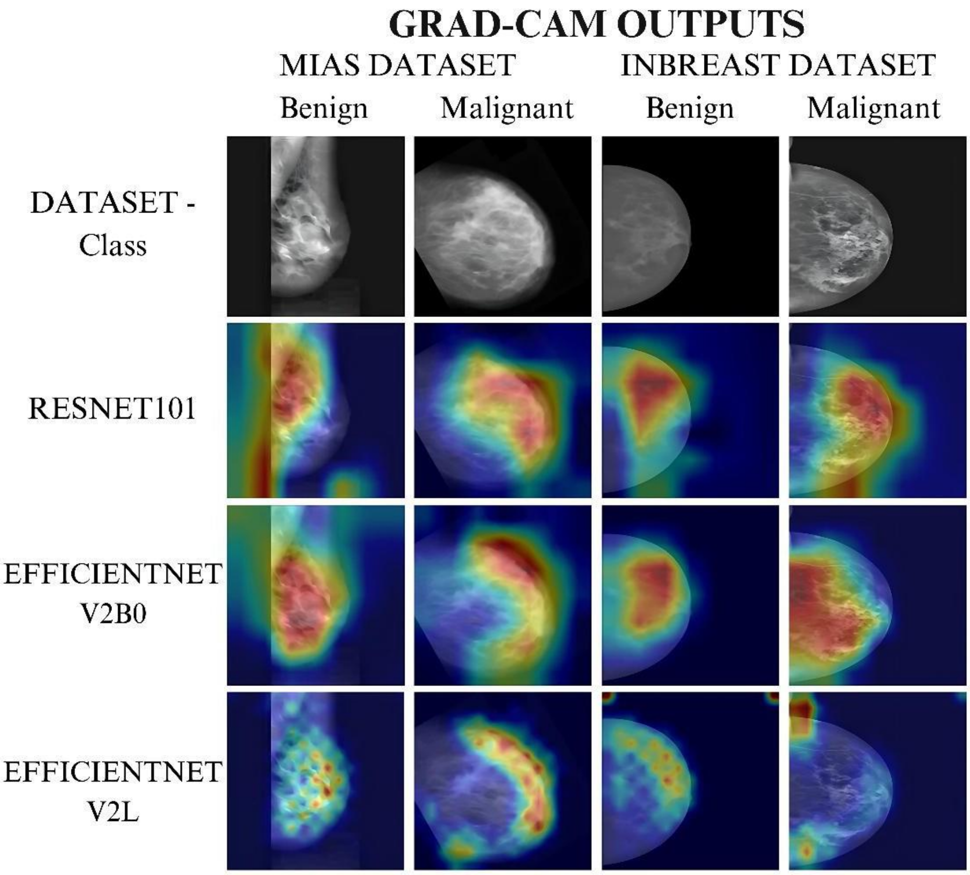
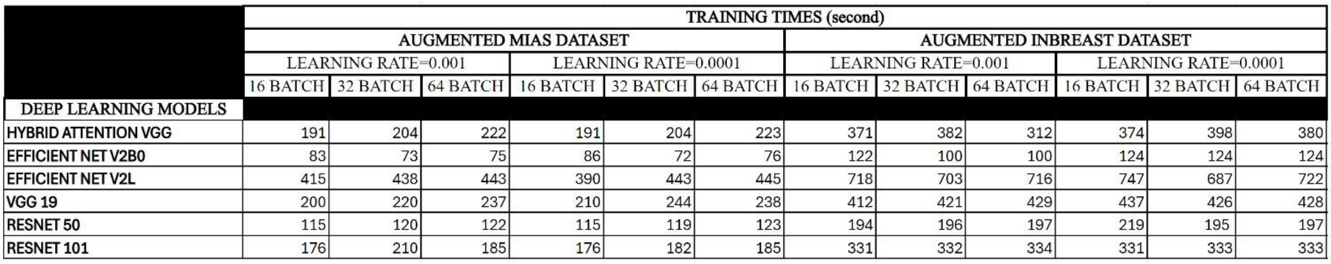


Figure 12. Mammogram images from two distinct classes sourced from both datasets, along with the GRAD-CAM outputs of the RESNET101, EFFICIENTNETV2B0, and EFFICIENTNETV2L models.

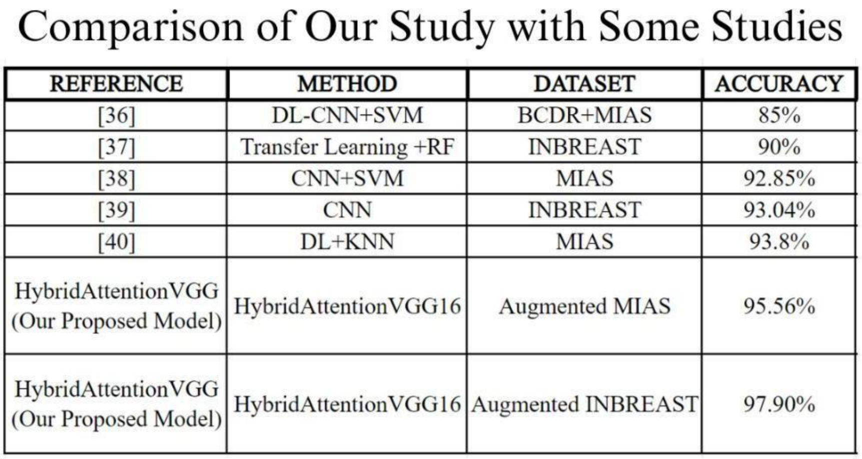
Figures 11 and 12 illustrate the GRAD-CAM outputs of the decision structures of the highest-performing models on mammograms from each dataset and class. In the GRAD-CAM output, the areas where the model focuses are depicted with varying color tones. Warm colors (red, orange, yellow) highlight the regions that contribute the most to the model's decision, reflecting the key features the model considers important. On the other hand, cool colors (blue, purple) show the areas that the model pays less attention to or disregards. This visualization clearly illustrates which parts of the image the model takes into account during the classification process.



|  | TRAINING TIMES (second) |  |  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | AUGMENTED MIAS DATASET |  |  |  |  |  | AUGMENTED INBREAST DATASET |  |  |  |  |  |
|  | LEARNING RATE=0.001 |  |  | LEARNING RATE=0.0001 |  |  | LEARNING RATE=0.001 |  |  | LEARNING RATE=0.0001 |  |  |
|  | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH | 16 BATCH | 32 BATCH | 64 BATCH |
| DEEP LEARNING MODELS |  |  |  |  |  |  |  |  |  |  |  |  |
| HYBRID ATTENTION VGG | 191 | 204 | 222 | 191 | 204 | 223 | 371 | 382 | 312 | 374 | 398 | 380 |
| EFFICIENT NET V2B0 | 83 | 73 | 75 | 86 | 72 | 76 | 122 | 100 | 100 | 124 | 124 | 124 |
| EFFICIENT NET V2L | 415 | 438 | 443 | 390 | 443 | 445 | 718 | 703 | 716 | 747 | 687 | 722 |
| VGG 19 | 200 | 220 | 237 | 210 | 244 | 238 | 412 | 421 | 429 | 437 | 426 | 428 |
| RESNET 50 | 115 | 120 | 122 | 115 | 119 | 123 | 194 | 196 | 197 | 219 | 195 | 197 |
| RESNET 101 | 176 | 210 | 185 | 176 | 182 | 185 | 331 | 332 | 334 | 331 | 333 | 333 |

Figure 13. The training durations of high-performing models

Although the attention mechanism usually increases computational cost, the training times for Hybrid Attention VGG were observed to be average compared to other high-performing models. It is shown in Figure 13. When examining other computational costs by looking at a few hyperparameter values, it has been observed that the proposed model's resource usage is also reasonable.



Comparison of Our Study with Some Studies

| REFERENCE | METHOD | DATASET | ACCURACY |
|---|---|---|---|
| [36] | DL-CNN+SVM | BCDR+MIAS | 85% |
| [37] | Transfer Learning +RF | INBREAST | 90% |
| [38] | CNN+SVM | MIAS | 92.85% |
| [39] | CNN | INBREAST | 93.04% |
| [40] | DL+KNN | MIAS | 93.8% |
| HybridAttentionVGG (Our Proposed Model) | HybridAttentionVGG16 | Augmented MIAS | 95.56% |
| HybridAttentionVGG (Our Proposed Model) | HybridAttentionVGG16 | Augmented INBREAST | 97.90% |

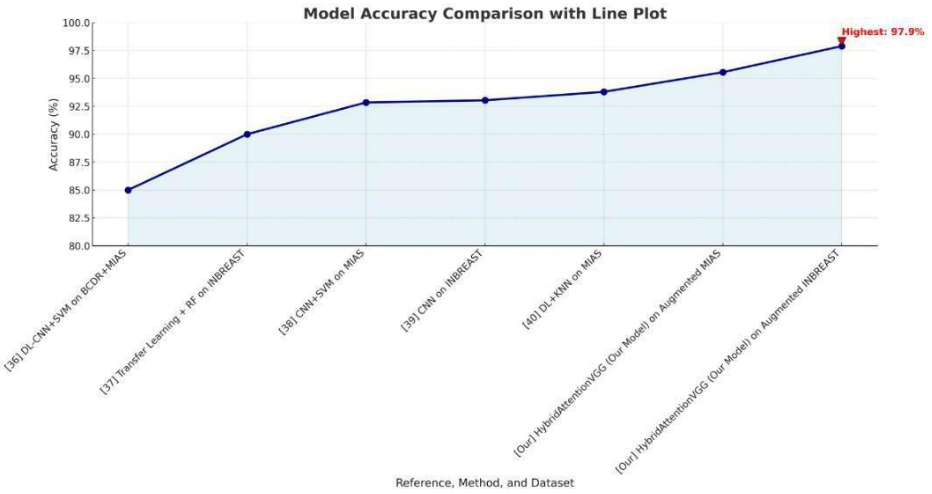


Figure 14. Comparing our findings with other studies in the literature

Figure 14 compares the Accuracy values of the most successful results obtained from hyperparameter optimization of our studies with some studies conducted in this literature. The line plot below the image emphasizes this comparison even more.

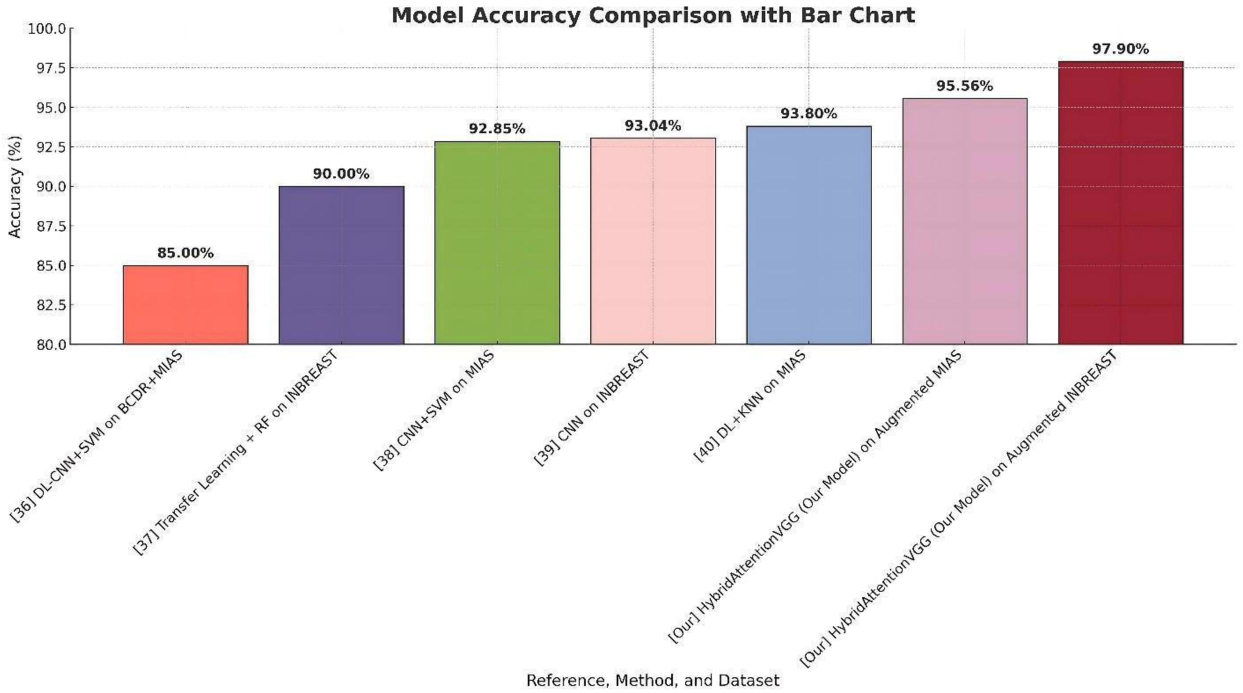


Figure 15. Bar plot graph of the studies in the literature and our work.

Figure 15 shows the bar plot graph of studies conducted in the literature on breast cancer classification and the studies we have carried out.

## Discussions

During the preliminary investigation conducted before proposing the model, the performances of various models, including VGG16, VGG19, RESNET50, RESNET101, EfficientNetV2B0 and EfficientNetV2L were compared based on Accuracy and F1 scores. These models demonstrated exceptional performance concerning the dataset and hyperparameter values. Even so, the proposed model demonstrated outstanding performance across a wide range of hyperparameter values and surpassed the performance of the other models in some hyperparameter values, especially on the Augmented INBREAST dataset. EfficientNetV2B0 and EfficientNetV2L models have demonstrated superior performance, attaining above 0.95 accuracy using diverse hyperparameter configurations prior to the evaluation of the proposed model. The models exhibiting the highest accuracy values are these models.

Before suggesting a model in the augmented MIAS dataset, it was seen that RESNET50 obtained one of the superior performance with a learning rate of 0.001 and a small batch size, whereas VGG19 performed well with a batch size of 64. RESNET50 achieved one of the superior performance using a learning rate of 0.0001 and a batch size 16. However, RESNET101 outperformed it by achieving even better results with bigger batch sizes. Within the Augmented INBREAST dataset, the VGG16 model exhibited notable performance with a learning rate of 0.001. However RESNET101 model displayed one of the superior performance with a learning rate of 0.0001.

Upon assessing the performance of deep learning models using the proposed HybridAttentionVGG model, it has been observed that this newly recommended model outperforms others in various hyperparameter configurations. Our suggested model exhibits equivalent accuracy to the EfficientNetV2B0 model, utilising a learning rate of 0.001 and a batch size of 32 on the Augmented MIAS dataset. The model enhanced performance across all batch size parameters on the Augmented INBREAST dataset with a learning rate of 0.001. With a learning rate of 0.0001, the accuracy has improved with a batch size of 16. The augmented INBREAST dataset enhanced performance in four of six distinct hyperparameter configurations. The usefulness of our newly proposed model has been shown by achieving superior performance in five out of twelve scenarios, including two datasets and six distinct hyperparameter values.

Before the model was proposed, the model with the highest accuracy value was EfficientNetV2L with an accuracy value of 0.9764, whereas after including our developed model, the model with the highest performance is the HybridAttentionVGG model with an accuracy value of 0.9790.

As shown in Figures 14 and 15, when compared to other breast cancer classification studies in the literature, our proposed new model, the HybridAttentionVGG model, has been demonstrated to be a successful model with accuracy values of 0.9556 and 0.9790 on the augmented MIAS and the augmented INBREAST datasets, respectively.

The HybridAttentionVGG model surpasses conventional models by integrating the robust feature extraction skills of VGG16 with a unique attention method that dynamically highlights significant features, guaranteeing more concentrated and resilient learning. The unique skip connections of the model preserve both the original and improved information, resulting in a well-balanced strategy that improves accuracy without substantially increasing computational complexity. Accordingly, the performance measure values of this model have surpassed those of other models.

## Conclusions

The HybridAttentionVGG model integrates the straightforwardness and resilience of the VGG16 architecture with an innovative attention mechanism that improves feature learning by prioritizing significant regions in the image. It offers a balanced solution between typical CNN models, which may miss important subtle features, and more computationally expensive designs like ResNet or Vision Transformers. The proposed model incorporates a streamlined attention mechanism and skip connections to preserve original and enhanced characteristics, providing a flexible and computationally practical option for various image classification tasks.

The proposed model can compete with those within the EfficientNet architecture. Nearly all hyperparameter values exhibit comparable or superior accuracy across the two datasets. Particularly on the augmented INBREAST dataset, it has exceeded the performance of EfficientNet models across numerous hyperparameter configurations. The model's training duration is intermediate between the training durations of the two EfficientNet models, considering the computational cost. The suggested model exhibits both low computational requirements and high accuracy.

The study has shown that this model is highly effective and helpful for classifying both benign and malignant cancers, consistently outperforming other models in most circumstances.

The forthcoming study will investigate the influence on classification accuracy by employing diverse data obtained by creating synthetic data from a derived dataset in deep learning models and evaluating the effectiveness of using synthetic data by comparing these datasets. Generating synthetic data will address the imbalanced data problem and find a solution.

## References

[1] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, ‘Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study’, *Journal of Imaging*, vol. 5, no. 3, 2019.

[2] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. Guevara Lopez, ‘Representation learning for mammography mass lesion classification with convolutional neural networks’, *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248–257, 2016.

[3] D. Moura *et al.*, ‘Benchmarking Datasets for Breast Cancer Computer-Aided Diagnosis (CADx)’, 11 2013, vol. 8258, pp. 326–333.

[4] B. Huynh, H. Li, and M. Giger, ‘Digital mammographic tumor classification using transfer learning from deep convolutional neural networks’, *Journal of Medical Imaging (Bellingham, Wash. ),* vol. 3, p. 034501, 07 2016.

[5] A. Elbagoury, ‘Breast Infrared Thermography Segmentation Based on Adaptive Tuning of a Fully Convolutional Network’, *Current Medical Imaging Reviews*, vol. 16, pp. 611–621, 05 2020.

[6] Z. Jiao, X. Gao, Y. Wang, and J. Li, ‘A Deep Feature Based Framework for Breast Masses Classification’, *Neurocomputing*, vol. 197, 03 2016.

[7] F. F. Ting, Y. J. Tan, and K. S. Sim, ‘Convolutional neural network improvement for breast cancer classification’, *Expert Systems with Applications*, vol. 120, pp. 103–115, 2019.

[8] A. Rampun, B. Scotney, P. Morrow, and H. Wang, ‘Breast Mass Classification in Mammograms using Ensemble Convolutional Neural Networks’, 09 2018,
pp. 1–6.

[9] V. R. and M. S, ‘DIscrete wavelet transform based principal component averaging fusion for medical images’, *AEU - International Journal of Electronics and Communications*, vol. 69, pp. 896–902, 04 2015.

[10] D. Ragab, M. Sharkas, S. Marshall, and J. Ren, ‘Breast cancer detection using deep convolutional neural networks and support vector machines’, *PeerJ*, vol. 7,
p. e6201, 01 2019.

[11] H.-C. Shin *et al.*, ‘Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning’, *IEEE Transactions on Medical Imaging*, vol. 35, 02 2016.

[12] P. Ballester and R. Araujo, “On the Performance of GoogLeNet and AlexNet Applied to Sketches”, *AAAI*, vol. 30, no. 1, Feb. 2016.

[13] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks.,” in *NIPS*, 2016, pp. 379–387.

[14] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge.,” *CoRR*, vol. abs/1409.0575, 2014.

[15] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, ‘A curated mammography data set for use in computer-aided detection and diagnosis research’, *Scientific data*, vol. 4, p. 170177, Dec. 2017.

[16] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, ‘Revisiting Unreasonable Effectiveness of Data in Deep Learning Era’, in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 843–852.

[17] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, ‘Understanding data augmentation for classification: when to warp?’, *CoRR*, vol. abs/1609.08764, 2016.

[18] M.-L. Huang and T.-Y. Lin, ‘Dataset of breast mammography images with masses’, Data in Brief, vol. 31, p. 105928, 2020.

[19] F. Zhuang et al., ‘A Comprehensive Survey on Transfer Learning’, arXiv [cs.LG]. 2020.

[20] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[21] S. Liu and W. Deng, ‘Very deep convolutional neural network based image classification using small training sample size’, *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 730–734, 2015.

[22] V. Sudha and D. Ganeshbabu, ‘A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning’, *Computers, Materials & Continua*, vol. 66, pp. 827–842, 01 2020.

[23] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ‘Densely Connected Convolutional Networks’, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[25] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[26] Y.-M. Chung, C.-S. Hu, A. Lawson, and C. D. Smyth, ‘TopoResNet: A hybrid deep learning architecture and its application to skin lesion classification’, *CoRR*, vol. abs/1905.08607, 2019.

[27] M. Koç and R. Özdemir, ‘Enhancing Facial Expression Recognition in the Wild with Deep Learning Methods Using a New Dataset: RidNet’, *RidNet. Bilecik Seyh Edebali University Journal of Science*, vol. 6, no. 2, pp. 384–396, 2019.

[28] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.,” *CoRR*, vol. abs/1704.04861, 2017.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ‘MobileNetV2: Inverted Residuals and Linear Bottlenecks’, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[30] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, ‘Learning Transferable Architectures for Scalable Image Recognition’, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8697–8710.

[31] M. Wang, B. Liu, and H. Foroosh, ‘Design of Efficient Convolutional Layers using Single Intra-channel Convolution, Topological Subdivisioning and Spatial “Bottleneck” Structure’, *arXiv: Computer Vision and Pattern Recognition*, 2016.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ‘Rethinking the Inception Architecture for Computer Vision’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, ‘Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning’, *AAAI Conference on Artificial Intelligence*, vol. 31, 02 2016.

[34] F. Chollet, ‘Xception: Deep Learning with Depthwise Separable Convolutions’, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.

[35] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, ‘Learning Transferable Architectures for Scalable Image Recognition’, arXiv [cs.CV]. 2018.

[36] Saraswathi Duraisamy and Srinivasan Emperumal, “Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier.” IET Computer Vision vol. 11, no. 8, PP. 656-662, July 2017.

[37] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for analysing masses in mammograms with minimal user intervention," *Med. Image Anal.*, vol. 37, pp. 114-128, 2017.

[38] Jaffar, M. A., "Deep learning based computer aided diagnosis system for breast mammograms," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 7, pp. 286-290, 2017.

[39] E. M. F. El Houby and N. I. R. Yassin, ‘Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks’, *Biomed. Signal Process. Control*, vol. 70, no. 102954, p. 102954, Sep. 2021.

[40] P. Kaur, G. Singh, and P. Kaur, ‘Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification’, Inform. Med. Unlocked, vol. 16, no. 100151, p. 100151, 2019.

[41] M. Tan and Q. V. Le, ‘EfficientNetV2: Smaller Models and Faster Training’, CoRR, vol. abs/2104.00298, 2021.