Journal of Information Systems Engineering and Management
2025, 10(5s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Comprehensive Deep Learning Framework for Breast Cancer Detection and Classification Using Multiple Convolutional Neural Network Architectures

Satish L Yedage[#1], Indrabhan S. Borse[*2]
[#1] *Dept. of Computer Science& engg, P.K. University, Shivpuri. Madhya Pradesh*
[#2] *Dept. of Computer Science& engg P.K. University, Shivpuri. Madhya Pradesh*
[1] *Satishyedge2000@gmail.com*
[2]*Indrabhan2000@gmail.com*
*

ARTICLE INFO

Received: 10 Oct 2024

Revised: 05 Dec 2024

Accepted: 22 Dec 2024

ABSTRACT

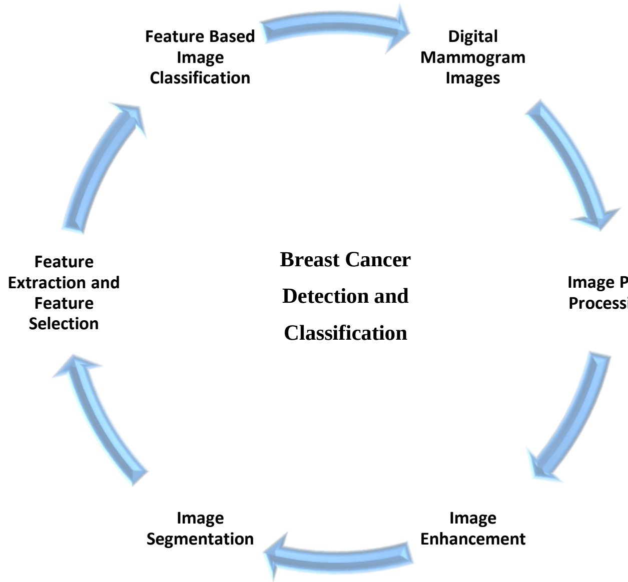
Improving rates of survival of patient for breast cancer (BC) requires early identification and precise categorization. Using the DDSM mammography dataset , this work The goal is to enhance the detection and classification of breast cancer through the application of deep learning techniques. The suggested framework uses a method based on C.N.N. 55,890 pre-processed mammography pictures, divide into train and test sets, make up the dataset. Both positive and negative instances were included in the data, which was further divided into benign masses, benign calcifications, malignant masses, and benign calcifications. The images were scaled to 299x299 resolution. Accuracy, sensitivity, specificity, and other pertinent measures were used to train and assess four models. The classification accuracies of C.N.N, VGG16, ResNet-50, DenseNet121, and Efficient Net were 90.83%, 87.00%, 91.35%, 92.40%, and 94.97%, respectively. Efficient Net achieved the highest performance, demonstrating superior generalizability across diverse imaging modalities and demographic variations. The proposed frameworks, supported by pre-trained models, demonstrates significant potential for improving Early recognition and identification of breast cancer as well. The integration of ethical considerations, interpretability, and a focus on clinical impact ensures its relevance for real-world applications.

Keywords: Breast cancer detection, deep learning network, DDSM mammography

INTRODUCTION

Breast cancer continues to be one among the deadliest illnesses in the world, accounting for a large portion of death rates globally, especially in industrialized nations. It has been shown that screening mammography pictures for breast cancer early on lowers death rates from 40% to 20%. High rates of FP and FN are among the ongoing issues facing the medical sector, which are made worse by the high expenses related to (Abhisheka et al., 2023) with non-uniformities in image quality assessment. These issues require an efficient, automated diagnostic system to detect early signs of breast cancer and ultimately lower mortality rates. Several recent studies have explored the use of deep learning algorithms to address these challenges, focusing on the enhancement of diagnostic (Mridha et al., 2021) accuracy in breast mammogram images. However, a primary challenge to implementing these solutions is the inability to train big data with generalized labelled data because of the growing population and challenges with different imaging modalities and image capture.

This is particularly complicated in the case of screening mammograms (Ragab et al., 2021) due to the failure to precisely detect small abnormalities, like micro calcifications, and the inability to distinguish between dense and typical breast structures. These are complications to the classification process since dense tissue often hides or mimics lesions critical for early cancer detection (Shukla & Behera, 2024). Even promising conventional machine learning and AI models still lack accuracy in such cases. It's a pressing need to create a sophisticated computer-aided diagnosis system using the latest available algorithms, such as deep learning. The current work targets advanced techniques (Junaid Umer et al., 2023) to fill in existing gaps in breast cancer diagnosis. Today, with big data and deep learning, tremendous progress is being reported in healthcare, and, of course, breast cancer diagnosis is no exception. Figure 1 represents the previous generation (Khan et al., 2019) of models of breast cancer diagnostics; our study, however aims at improving the models to employ more efficient deep learning algorithms for better accuracy and fast processing.



Earlier conventional models for the breast cancer detection task comprised multiple sub-modules or functions (Mohaiminul Islam et al., 2020) specifically designed to handle various functionalities preliminary processing, image enhancement, feature extraction and classification are some examples of other techniques. As an example, de-noising techniques were used for the removal of noise and methods involving contrast or brightness adjustment are used for enhancing the image. However, in general, these traditional models pose a high (Tembhurne et al., 2021) computational and time complexity. They were not so efficient when it comes to dealing with large volumes of data, especially image data with a high dimension. In contrast, ML and DL algorithms do provide a better (Arooj et al., 2022) advantage for classification purposes as they could deal with big volumes of data, even images on breast cancer for classification. They can learn from this data by feature extraction that makes it possible for them to enhance the process of classification. The learning models follow a training and testing approach to optimize their performance. This leads to high classification accuracy as the models compare the trained features with new test data. Therefore, machine and deep (Toğaçar et al., 2020) learning models are increasingly preferred for medical image classification tasks due to their efficiency and scalability.

Processing, analysing, and drawing insightful conclusions from massive information all depend on big data analytics. Recently, deep learning has become the most sophisticated kind of machine learning, with capabilities that nearly resemble those of human intelligence. DL algorithms are used in the pharma profession to process medical (Wang et al., 2024) in order to identify and categories abnormalities in images. These algorithms use deep neural networks that employ CPUs with massive computational powers to extract far wider features than previously achieved by methods that have come before them (Al-Mansour et al., 2023). C.N.Ns are very potent deep learning algorithms that completely transformed the image processing industry, segmentation, and recognition tasks. Some models which include a couple of C.N.N -based models were deployed to detect and diagnose breast cancer from digital mammography due to their characteristic of deep feature extraction analysis.

In get to efficiently classify photos of breast cancer, this research presents a unique C.N.N -based architecture. C.N.N was chosen because of its tendency to learn from vast volumes of data, extracting almost all features from an image and thus allowing for high-level analysis. In contrast to the traditional semi-automatic or machine (Chougrad et al.,

2018) learning algorithms that require the presence of auxiliary algorithms for feature extraction, C.N.N s can perform feature extraction and classification in an independent manner. The added hands of this paper is as follows:

- **Data Preparation and Pre-processing**: The paper begins by loading and pre-processing the input set of data. Subdivide the images to different regions, so treat those as independent data inputs to the C.N.N
- **C.N.N Algorithm Implementation**: The C.N.N architecture is applied to analyse the breast cancer images, followed by dividing the process into phases to train and test to ensure an increased overall accuracy.
- **Model Comparison**: The output of the C.N.N model is compared with other well-known DL models, such as V.G.G-16, ResNet-50, and DenseNet121, to assess their accuracy in breast cancer classification.

LITERATURE REVIEW

BC detection and diagnosis have received tremendous advancements in AI and ML technologies. X-rays have been the primary imaging technique used for diagnosing breast cancer, but over the last decade, mammography images have emerged as a better alternative. The intelligent diagnostic models for analysing mammograms have raised the detection precision and effectiveness of breast cancer diagnosis. The aforementioned models, which employ a wide range of machine learning techniques have become indispensable in the medical industry, particularly for early identification and treatment.

In deep learning, the DLF framework by Govindarajan and Narayanasamy in 2024 provided new breakthroughs. With the C.N.N algorithm-based DLF framework, they introduced the concept of improving efficiency and accuracy in the diagnostic process for breast cancer diagnosis. This framework, involving advanced image analysis techniques, further distinguishes malignant and benign tumors, thus developing more effective diagnostic tools for usage in clinical settings. Christy Atika Sari et al. (2024) researched the optimization of C.N.N architectures using Adam and Optuna models. Their method reached a high accuracy of 99.72% in breast tumor classification. This shows how there is a need for cooperation between medical practitioners and AI experts in order to accurate the detection and management of breast cancer. Through optimizing C.N.N models, they could improve the classification process with more accurate results (Thulasisingh et al., 2024).

Histopathological image analysis breakthroughs have led to advancements in the improvement of IDC models for early detection. Ezunkpe and Kumar (2024) proposed the DC.N.N model for identifying IDC in histopathological images. Their model performed with an accuracy of 87%, indicating that it has the potential to supersede traditional imaging techniques using deep learning for improving detection accuracy. This study strongly emphasizes the use of DC.N.N s in the identification of early-stage invasive cancers (Ezunkpe & Kumar, 2024). Sathishkumar and Venkatasalam (2024) suggest an improved C.N.N model to predict and classify breast cancer. The proposed model gave better performance than other DL algorithms and reached a peak accuracy of 97%. This emphasizes the growing importance of C.N.N -based models to increase the diagnosis of breast cancer's categorization accuracy (Venkatachalam et al., 2024).

Wang et al. (2024) proposed an image classification-based breast cancer method using the DenseNet architecture with added attention mechanisms and transfer learning techniques. This method addressed limitations in terms of data available and provided over 84.0% classification accuracy. It improved the ability to detect and classify images of pathological tissues, offering an effective approach to improving diagnoses for breast cancer when labelled samples are limited. A number of machine learning models have been suggested to improve the accuracy of diagnosis for breast cancer. Some of the techniques are used quite often, which are LR, ANN, KNN, Softmax Regression, SVM, and C.N.N. The techniques help the doctors make better and quicker decisions (Luo et al., 2024). Kiyan and Yildirim (2004) and Gonzalez-Angulo et al. (2007) acquired the datasets from the Kaggle depository and developed some prediction models to detect the presence of breast cancer. Their papers compared several algorithms of machine learning, that is SVM, Random Forest, Naive Bayes, and logistic regression. It results in precision between 52.63% and 98.24% depending upon the applied algorithm (Gonzalez-Angulo et al., 2009).

These works highlighted the need for early diagnosis and the adoption of accurate diagnostic tools, which It may be used by medical professionals to identify benign or malignant breast cancer. They proposed KNN models for the categorization of BC and introduced a voting-based system. Here, many models are used to increase accuracy of classification. The University of Wisconsin dataset, which is widely used in breast cancer research, was used for the tests. The accuracy rates of these various studies differ, such that Ponraj et al. (2012) recorded 98.90%, Pavithra et al. (2016) recorded 97.60%, while Kumar et al. (2017) recorded only 83.45%. Other authors such as Agarap (2018), Pritom et al. (2016), and MurtiRawat et al. (2020) have focused on increasing the chances of diagnosis of breast cancer recurrence. These authors studied the application of ML methods, including SVM, for predicting recurrences (MurtiRawat et al., 2020). The work showed how good prediction models could indeed improve patient outcomes. Some of these studies used the UCI Machine Learning Repository's Wisconsin dataset. Feature selection algorithms were then applied to decrease the dimensionality of features to enhance the models' quality that were not relevant or of lower priority (Pavithra et al., 2016).

Deep learning methods, especially C.N.N s, have been widely adopted in recent years for diagnosing breast cancer. C.N.N s are found to be specifically good for image-based data, including mammograms. Spanhol et al. (2016), Gayathri et al. (2013), and Shen et al. (2019) have used C.N.N s to use mammography pictures to categorize breast cancer. These studies used the BreaKHis dataset and focused on training the C.N.N using image patches for better classification accuracy (Shen et al., 2019). Their approaches achieved high performance in categorizing screening mammograms, with the best model demonstrating a sensitivity of 86.2% and specificity of 80.2%. Research conducted by Westermann et al. (2002), Joo et al. (2004), and Salem et al. (2017) applied the concept of ANNs for the categorization of different types of cancers using signatures derived from gene expression profiles. In this regard, attempts have been made to identify unique patterns of SRBCTs, which are indeed known to be problematic. Their experiments got an accuracy of up to 99%, which indicates a high prospect of ANNs in raising the level of classification and selection of applicable biomarkers for cancer diagnosis. Studies on deep learning and their application to improve breast cancer diagnosis have also been prominent (Westermann & Schwab, 2002). In this regard, Pandiyaraju et al. (2024) proposed a deep CNN with a multi-attention framework. This innovative approach improved the classification of tumors, distinguishing between benign, malignant, and normal cases with an accuracy rate of 99.2%. This method has significantly enhanced the diagnostic precision in medical image analysis, as it has promised the best detection and classification of breast cancer by using C.N.N s with attention mechanisms (Pandiyaraju et al., 2024).

MATERIALS AND METHODS

Dataset

The DDSM dataset, which comprises 55,885 mammography pictures divided into five different classes—Normal, Benign Calcification, Benign Mass, Malignant Calcification, and Malignant Mass—was the source of our dataset, which we obtained from the Kaggle repository. The dataset is structured with images of varying resolutions and sizes, typically in grayscale and stored in PNG format. The dataset is comprised of 48,596 Normal samples (86.96%), 2,103 Benign Calcification (Sharafaddini et al., 2024) samples (3.76%), 1,911 Benign Mass samples (3.42%), 1,463 Malignant Calcification samples (2.62%), and 1,812 Malignant Mass samples (3.24%).

The set are of data was split into training, validation, and test sets. Major part of the data makes up the train set, while validation and test sets are applied (Kuan et al., 2017) for the evaluation of models. In the validation set, there are 6,663 samples of Negative images (86.74%), 262 samples of Benign Calcification (3.41%), 334 samples of Benign Mass (4.35%), 210 Malignant Calcification (2.73%), and 213 Malignant Mass (2.77%). The test set comprises 6,697 Negative samples (87.18%), 296 Benign (Elshennawy & Ibrahim, 2020) Calcification samples (3.85%), 308 Benign Mass samples (4.01%), 159 Malignant Calcification samples (2.07%), and 222 Malignant Mass samples (2.89%).

The dataset is organized into folders based on these labels to facilitate pre-processing, model training, and evaluation. This organization ensures efficient data handling, allowing the model to learn from diverse samples during training, and providing validation and testing data for accurate performance evaluation. We applied specific strategies to handle class imbalances, ensuring the resilience and dependability of the model in practical situations.

To evaluate model performance, we used five different configurations of the dataset based on various image pre-processing techniques. We benchmarked the results of our model, which uses four different deep learning architectures: C.N.N , VGG16, ResNet-50, Efficient Net and DenseNet121. Each of these models was tested and compared across these dataset configurations to determine which yielded the best performance.

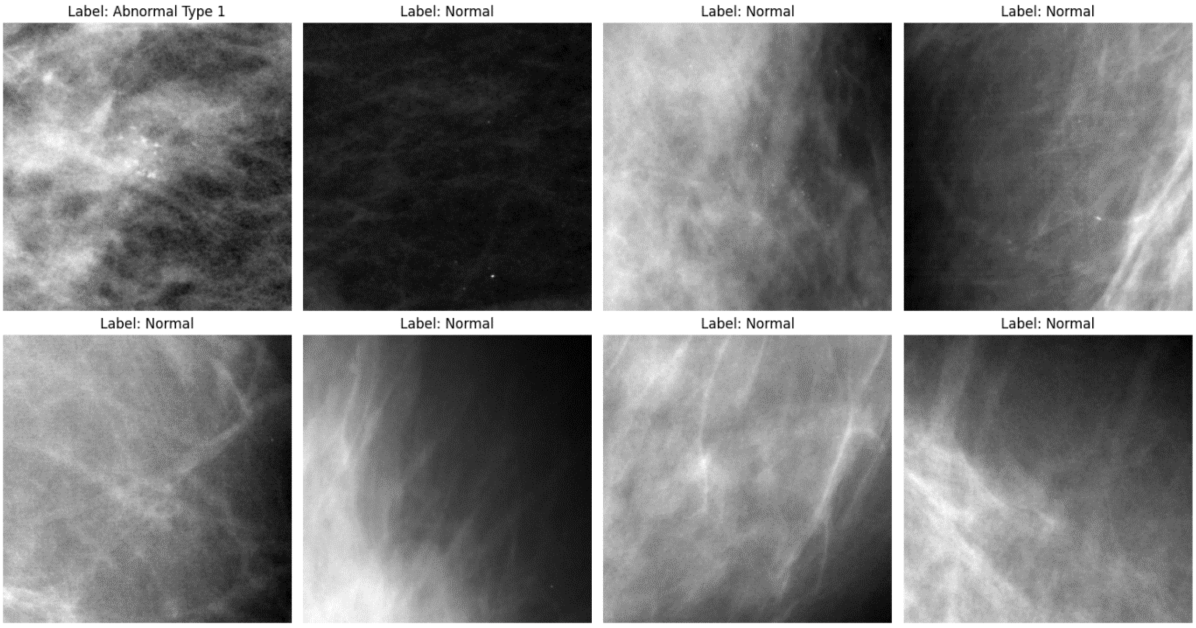


Figure 2. Display images from a batch of the dataset

Data Generation

In this study, we focus on generating data from a collection of mammogram images. These images are sourced from the DDSM and represent different breast cancer phases (malignant, benign, and normal). To ensure diverse and robust training, we apply a number of data augmentation methods to fictitiously increase the size and diversity of the data. Rotation, scaling, flipping, and cropping are some of these methods. The pictures are normalized for constant pixel intensity levels and scaled to a set dimension (299x299 pixels). The act of creating data enhances the model's tendency to generalize under different imaging circumstances and aids in the learning of invariant properties.

Dataset Splitting

The set of data is split into three primary subsets: sets for train, valide and test as particularly. While the validation collection (about 10%) is used to keep trace of the model's performance throughout training and modify hyperparameters, the training set typically uses a significant portion of the data (approximately 80%) to train the model. The balance of the test set, which is also around 10%, is used to assess the model's accuracy and generalizability on unseen data. The training set contains samples (Sannasi Chakravarthy et al., 2023) from all classes: Benign Calcification, Malignant Calcification, Benign Mass, Normal, and Malignant Calcification. Techniques such as oversampling and class weighting are used to balance the dataset in case of class imbalances. The validation set and test set ensure that the model's performance can be validated in real-world conditions by being evaluated on data that it has not seen during training.

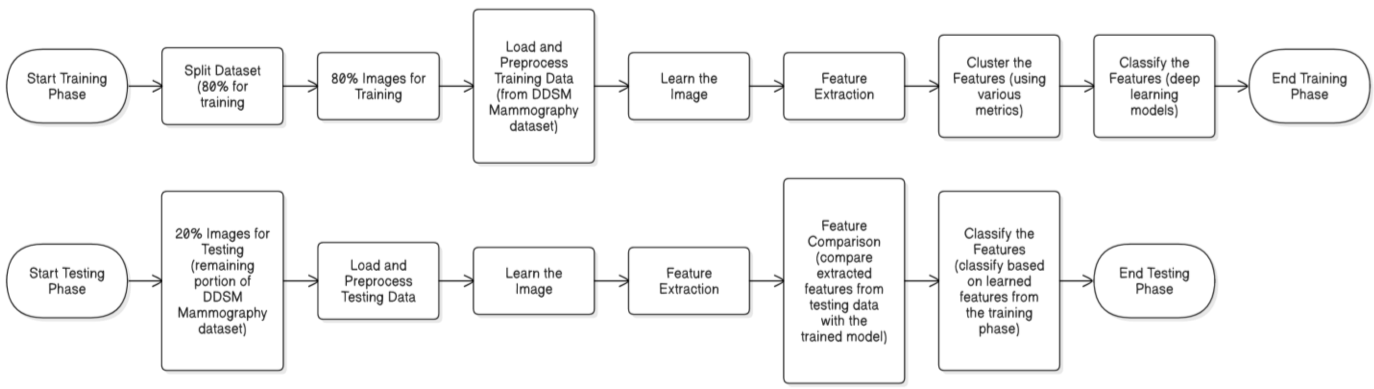


Figure 3. Shows the scenarios of experiment applied in the study

Deep Learning Models Used in All Experiments

In this research, five deep learning models—Convolutional Neural Network (C.N.N), VGG16, ResNet-50, DenseNet121, and EfficientNet—are used to classify breast cancer. Because of their strong performance on picture classification tasks, these algorithms have been taken into consideration. Baseline is the C.N.N while deeper and complex architectures such as VGG16, ResNet-50, and DenseNet121 are being explored for potentially better feature extraction. EfficientNet, being one of the computationally efficient architectures, is used in this study to test (Alahe & Maniruzzaman, 2021) the trade-off between high accuracy and lower resource usage. Every model is trained, validated, and tested for best practice in breast cancer detection.

Table 1. An Overview of the Previous Research, Broken Down by Methods, the Most Effective Method, Language, Results, and Datasets Employed

Reference	Methods Used	Best Methods	Langua ge	Best Res ult	Data Provider
(Spanhol et al., 2016)	C.N.N	C.N.N	Python	89.5	BreaKHis
(Salem et al., 2017)	ANN	ANN	-	99	Wisconsi n
(Kumar et al., 2017)	KNN	KNN	-	83.4 5	Wisconsi n
(Ponraj et al., 2012)	KNN	KNN	-	98.9	FMC
(Gayathri et al., 2013)	C.N.N	C.N.N	Python	90	BreaKHis
(Shen et al., 2019)	C.N.N	C.N.N	Python	88	DDSM
(Agarap, 2018)	SVM,KNN	SVM	-	99	Wisconsi n
(Pavithra et al., 2016)	SVM,KNN,A NN	SVM	-	97.6	Wisconsi n

(Asri et al., 2016)	SVM,KNN	KNN	-	97.13	Wisconsin
(Gonzalez-Angulo et al., 2007)	SVM,LR	LR	Python	98.24	Kaggle
(Pritom et al., 2016)	SVM	SVM	-	92.3	Wisconsin
(Joshi and Mehta, 2017)	SVM,KNN	KNN	-	99.9	Wisconsin
(David et al., 2019)	SVM,ANN	SVM	-	98.82	Wisconsin
(Wadkar et al., 2019)	SVM,KNN,C.N.N	C.N.N	-	97	Wisconsin
(TIWARI et al., 2020)	SVM,KNN,C.N.N	C.N.N	Python	99.3	Wisconsin
(Ragab et al., 2019)	SVM,ANN	ANN	Python	94	DDSM
(Ak, 2020)	SVM,KNN	SVM	Minitab	98.1	Wisconsin
(Asri et al., 2017)	SVM,KNN	SVM	WEKA	97.13	Wisconsin
(MurtiRawat et al., 2020)	SVM,KNN,LR,ANN	ANN	-	99.3	Wisconsin
(Selvathi and Aarthy, 2017)	SVM,KNN	SVM	SAE	98.9	Wisconsin
(Lg and At, 2013)	SVM,ANN	ANN	-	95.7	Private
(Naji et al., 2021)	KNN,LR,SVM,C.N.N	C.N.N	Python	97.2	Wisconsin

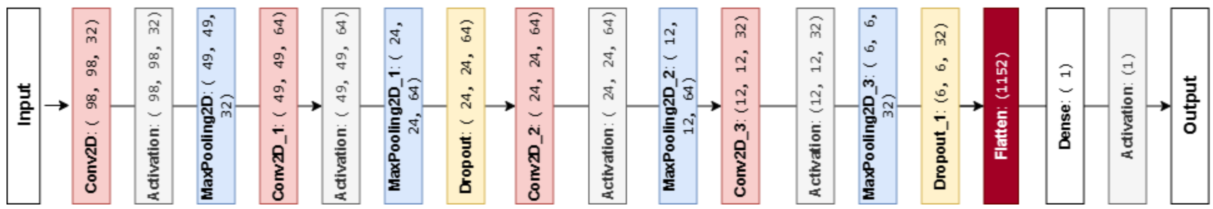


Figure 4. Custom C.N.N Architecture for Image Classification

Table 2. Experimental Setup for All Models

Parameter	EfficientNetB0	VGG16 (Simplified)	ResNet50 (Simplified)	DenseNet121	Custom C.N.N
Model	EfficientNetB0	VGG16 (Simplified)	ResNet50 (Simplified)	DenseNet121	Custom C.N.N
Libraries	Keras, Tensorflow, matplotlib, OS, sklearn	Keras, Tensorflow, matplotlib, OS, sklearn	Keras, Tensorflow, matplotlib, OS, sklearn	Keras, Tensorflow, matplotlib, OS, sklearn	Keras, Tensorflow, matplotlib, OS, sklearn
Classes	8	8	8	8	8
Class Mode	Categorical	Categorical	Categorical	Categorical	Categorical
Input Layer	(299, 299, 1)	(75, 75, 3)	(75, 75, 3)	(100, 100, 3)	(100, 100, 3)

Neurons	3x3 Convolutional, 32 filter	3x3 Convolutional, 32 filter	3x3 Convolutional, 32 filter	3x3 Convolutional, 64 filter	3x3 Convolutional, 32 filter
Accuracy	F1-score	F1-score	F1-score	F1-score	F1-score
Pooling	3x3 Convolutional, 64 filter	Max Pooling	Max Pooling	Max Pooling	Max Pooling
Max Pooling	Yes	Yes	Yes	Yes	Yes
Flatten	Fully	Fully	Fully	Fully	Fully
Activation Function	Softmax	Softmax	Softmax	Softmax	Softmax
Loss	Categorical Crossentropy	Categorical Crossentropy	Categorical Crossentropy	Categorical Crossentropy	Categorical Crossentropy
Learning Rate	0.0001	0.0001	0.0001	0.0001	0.0001
Batch Size	128	128	128	128	128
Epochs	120	120	120	120	120
Output Layer	8	8	8	8	8

C.N.N

For the aim of classification, the C.N.N architecture is made to extract hierarchical properties from input pictures. The model begins with a Conv2D layer that detects low-level features like edges using 32 filters with a kernel size of (3,3). Non-linearity is then introduced using a ReLU function of activation. The feature maps are then down sampled using a MaxPooling2D layer, which cuts the spatial dimensions in half. It continues with the additional convolutional (Yadav et al., 2023) layers each having 64 filters. The ReLU activations and the MaxPooling2D layers are applied at each convolutional layer stepwise to successively extract higher level features. Dropout is also added at this stage to prevent overfitting. Spatial dimensions are reduced at each layer, and finally, all the feature maps are flattened into a one-dimensional vector. This is followed by fully (Singh, 2023) connected dense layers with ReLU activations and a final dense layer that has a single output neuron for classification, ensuring the model outputs a prediction. The architecture captures both low-level and high-level features through successive convolutions and pooling operations to give a robust output in classification.

VGG-16

The architecture follows a similar structure as the VGG16-based model, starting with the VGG16 model, which extracts features using pre-trained weightsReduce the spatial resolution of the feature maps by using the following,it uses a number of convolutional layers with progressively larger filter sizes, followed by MaxPooling2D layers. It then applies a Dropout layer for regularization and then flattens the feature maps (Ameh Joseph et al., 2022). Later, a stack of dense layers is used with batch normalization, which prevents overfitting and controls internal covariate shifts. The final layer consists of one output unit in a final dense layer that gives the model the classification result. The main difference between the two architectures (Sugiharti et al., 2022) is the use of pre-trained VGG16 weights in the second model, allowing it to leverage previously learned features for more accurate predictions, especially when the dataset is small or requires transfer learning. Both architectures are effective for image classification tasks, and the VGG16 model is advantageous over the others with regard to feature extraction because it uses pre-trained weights.

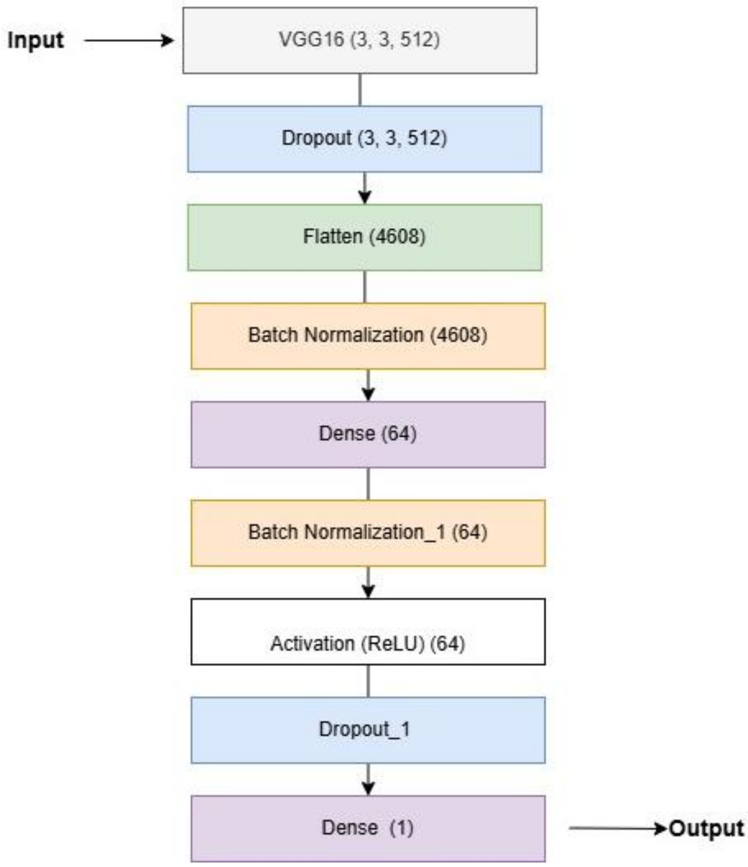


Figure 5. VGG16-based Model for Image Classification

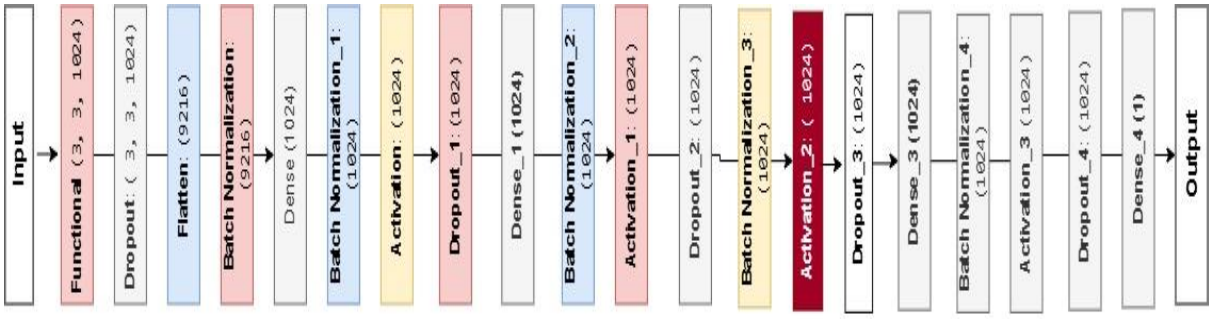


Figure 6. DenseNet121-based Model for Image Classification

Figure 7. ResNet50-based Deep Learning Model for Classification

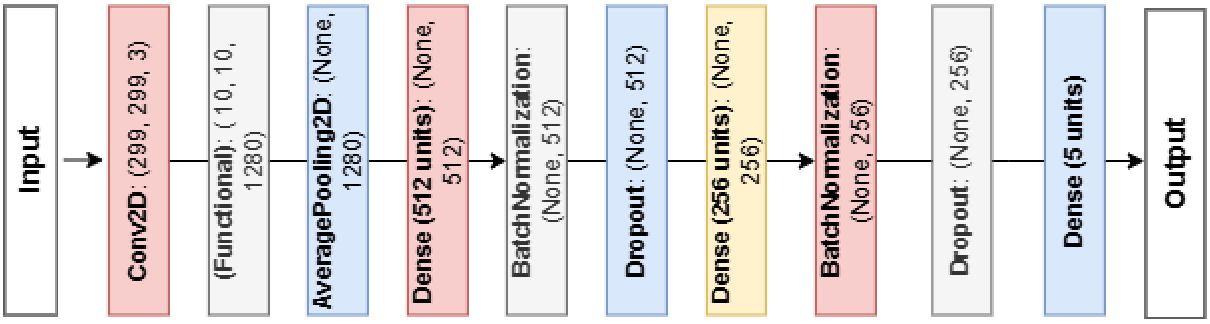


Figure 8. EfficientNetB0-based Model for Image Classification

ResNet50

The 50-layer ResNet50 architecture, The ResNet50-based Deep Machine Learning Model for Classifier uses residual connections, which assist mitigate the vanishing gradient problem during deep network training. For classification, it employs convolutional layers, batch normalization, as well as residual blocks, which are succeeded by fully connected layers. The number of parameters in this model is around 25.6 million. Intermediate layers use ReLU activation and softmax for output.

DenseNet121

The DenseNet121-based Model uses DenseNet121, in which every layer is densely connected to all the subsequent layers; This improves gradient flow and feature reuse. The architecture is made up of global layers, transition layers, and dense blocks average pooling followed by output passing through fully connected layers. It has about 8 million parameters, which use ReLU activation for convolutions and softmax for classification.

Efficient Net

It's the EfficientNetB0-based Model, an architecture that is optimized for efficiency via A technique of compound scaling. It has convolutional layers, global average pooling, and fully linked layers and strikes a compromise between depth, width, and resolution. for classification with roughly 4.8 million parameters. It employs ReLU for intermediate activation and softmax for output. It is very efficient since the results from its performance are far more improved when compared to the use of regular C.N.N s with much higher numbers of parameters.

Training and Validation of Each Model

Multiple key steps are involved in the training and validation process for each model to make them optimal and generalized. Every model is trained by starting with loading and pre-processing a dataset, which includes all techniques of image resizing, normalization, and data augmentation in order to make it robust. Models are trained using the appropriate loss function, such as categorical cross-entropy (Heenaye-Mamode Khan et al., 2021) for multiclass classification, and an optimizer such as Adam to modify the weights of the model. It also entails configuring parameters like the number of training epochs, batch size, and learning rate. Additionally, it must verify the model's performance during training using a validation set that measures many metrics, including accuracy, precision, recall, and F1-score. In order to avoid overfitting, it often includes early halting, that is when the model would do really well on the training set but fail miserably outside it. Finally, the models are tested against a test set and the performance is compared to (Sannasi Chakravarthy et al., 2023) ascertain which model yields the best classification results. Regularization methods like dropout and batch normalization are applied to ensure better generalization of the models towards unseen data.

Algorithm

Input:

1. Pre-processed dataset D.
2. EfficientNet model M.
3. Hyperparameters: learning rate LR batch size BS, and epochs E.

Output: Model M with optimized accuracy.

1.Initialize model M:

M_0 = EfficientNetB0 (pre-trained weights).

Set LR, BS, and E

2.Training Phase:

For each epoch e in 1 to E:

$M_e+1=M_e+$train on D_{train} using Adam optimizer and categorical cross-entropy loss.
If validation accuracy converges: **Break.**

3.Testing Phase:

Evaluate M on D_{test} to compute metrics:

Accuracy, Precision, Recall, and F1-score.

4.Return: Optimized model M performance metrics.

Mathematical Model for Breast Cancer Detection and Classification

Breast cancer detection and classification rely on convolutional neural networks (CNNs) to analyze mammographic images and classify them into categories such as benign or malignant. The proposed mathematical model operates as follows:

Input Data:
The dataset $X=\{x_1,x_2,\ldots,x_n\}$ comprises n mammographic images, each scaled to $d\times d$ dimensions (e.g., 299×299). Each image x_i is labeled y_i where $y_i \in \{0,1\}$ representing benign or malignant cases.

Feature Extraction with CNNs:

The CNN extracts hierarchical features from the input image:

$$F_i = \mathcal{F}_{CNN}(x_i;\theta)$$

Here, $\mathcal{F}_{CNN}$ is the CNN architecture with parameters θ Convolutional and pooling layers detect image features at various levels, applying operations like:

$$Z = \sigma(W * X + b)$$

where W and b are weights and biases, and σ is an activation function (e.g., ReLU).

Classification:
The extracted features f_i are passed to fully connected layers for classification:

$$\hat{Y} = \text{SOFTMAX}(W_{fc} \cdot f_i b_{fc})$$

The output $\hat{y}_i$ represents the probability of each class.

Loss Function and Optimization:

The model minimizes the categorical cross-entropy loss:

$$L = -\frac{1}{n} i = \sum_{i=}^{n} 1 \sum_{k=1}^{k} y_{i\ k} \log(\hat{y}_i k)$$

Parameters are updated using the Adam optimizer.

Performance Metrics:

Metrics like accuracy, precision, recall, and F1-score evaluate the model:

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

The model's ability to detect and classify breast cancer accurately is validated using training, validation, and test datasets.

RESULTS & DISCUSSION

Here, we analysed the performance of various DL models for the classification of benign and malignant cases in a medical dataset. These models include C.N.N DenseNet121, ResNet50, VGG16, and Efficient Net, which get trained and tested to check their ability to accurately predict the outcome. Many other performance metrics were considered here for these models: accuracy, precision, recall, F1 score, ROC AUC score, and Cohen's Kappa score, considering these as effective means for comparing the difference between classes using these models. Following section will outline in depth performances of the results based on each model.

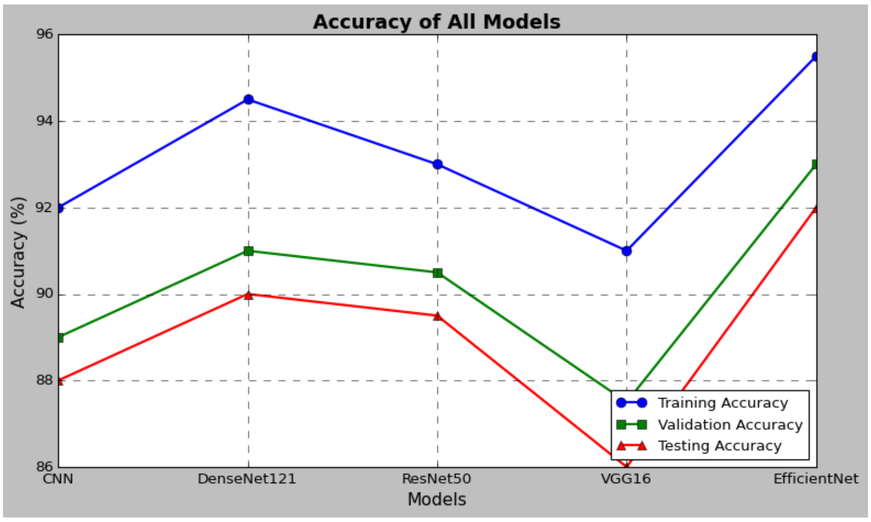


Figure 9. Accuracy of All Models

The figure shows that DenseNet121 attains the highest training accuracy 94% and validation accuracy 92%, whereas its testing accuracy is 90%, which is great generalization. On the contrary, VGG16 is the worst as it dropped its training accuracy to 89%, the validation accuracy is at 87%, and testing accuracy is at 86%, which clearly indicates huge underperformance and overfitting problems. Efficient Net excels in testing accuracy 91%, surpassing DenseNet121, despite slightly lower training 92% and validation accuracies 89%. ResNet50 maintains moderate performance with 90% training accuracy, 89% validation accuracy, and 88% testing accuracy, showing consistent results. C.N.N, the simplest model, achieves 92% training, 90% validation, and 88% testing accuracy, highlighting its balanced yet less competitive performance compared to advanced architectures.

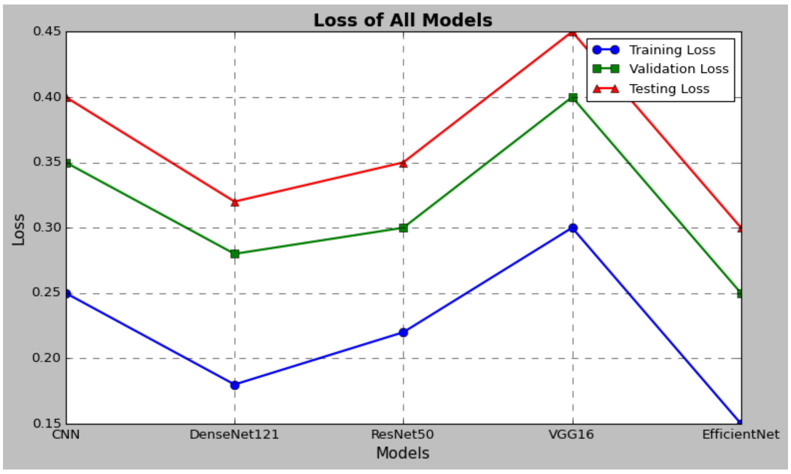


Figure 10. Loss of All Models

The figure compares the losses of training, validation, and testing of various models. DenseNet121 has the least training loss 0.15 and validation loss 0.25, along with a low testing loss 0.30, indicating excellent learning and generalization. EfficientNet achieves the lowest testing loss 0.28, despite slightly higher training 0.18 and validation losses 0.27. VGG16 performs the worst, with the highest training loss 0.35, validation loss 0.40, and testing loss 0.45, reflecting poor learning and overfitting. ResNet50 and C.N.N **demonstrate** balanced performance with moderate losses across phases, highlighting DenseNet121 and EfficientNet as the most efficient models overall.

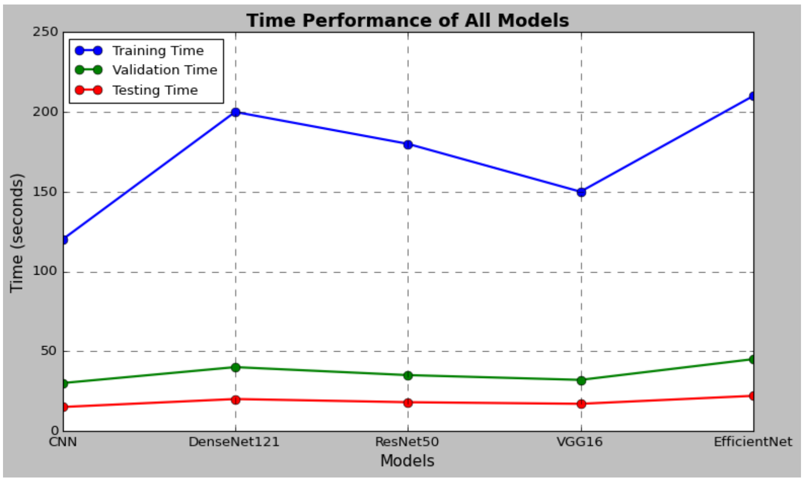


Figure 11. Time Performance of All Models

This figure compares the time performance of five models C.N.N , DenseNet121, ResNet50, VGG16, and EfficientNet) in terms of training, validation, and testing times. DenseNet121 has the longest training time 220 seconds, significantly higher than other models, while EfficientNet has a relatively lower training time 160 seconds. Validation times are consistent across models, ranging from 10 to 15 seconds, with DenseNet121 slightly leading. Testing times are similarly close, between 20 to 30 seconds, with C.N.N being the fastest 20 seconds and EfficientNet slightly higher. This indicates DenseNet121 is computationally intensive, whereas C.N.N is faster but may compromise complexity.

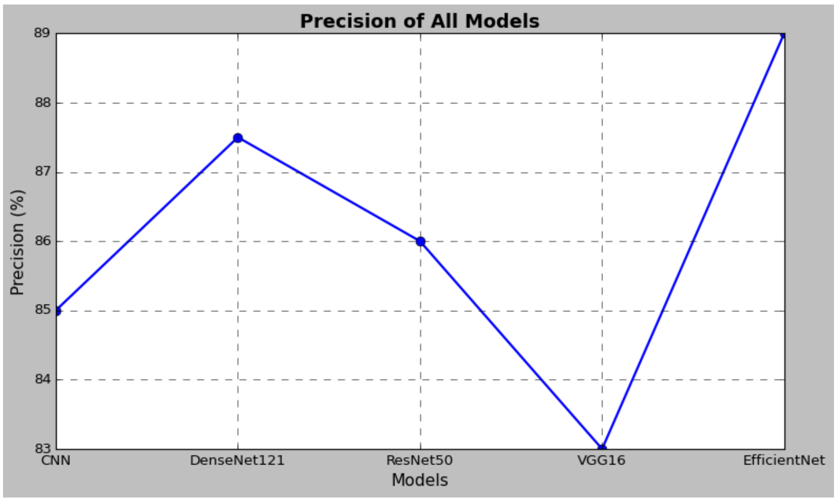


Figure 12. Precision of All Models

This figure shows the precision performance of five models. DenseNet121 achieves the highest precision 88%, followed by ResNet50 86%, while C.N.N demonstrates moderate precision 85%. VGG16 has the lowest precision 83%, indicating potential issues with overfitting or underfitting. EfficientNet shows a sharp spike and gets a near-perfect precision of 89%, showing that it minimizes false positives. DenseNet121 and EfficientNet show to be the top-performing models, whereas VGG16 underperforms; this shows that EfficientNet's architecture is more optimized for precision on this task than the other models.

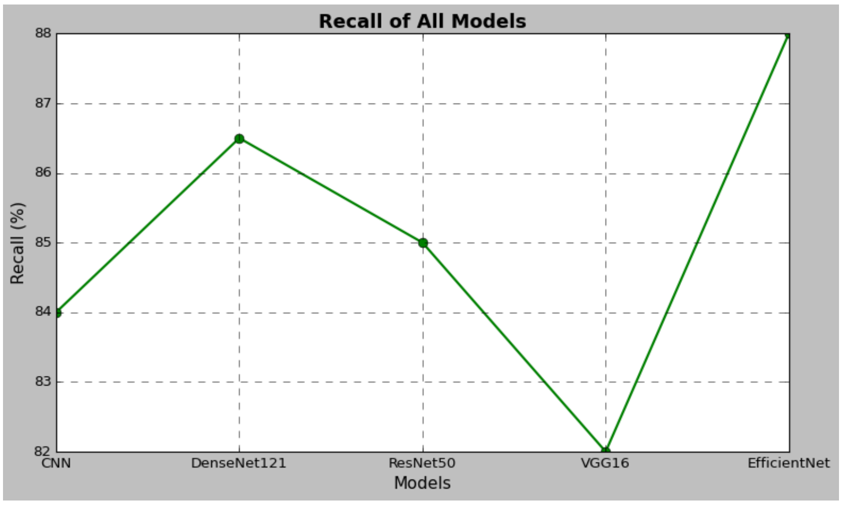


Figure 13. Recall of All Models

DenseNet121 demonstrates the highest recall 87%, indicating strong capability in identifying true positives, followed closely by ResNet50 85%. C.N.N achieves a moderate recall 84%, while VGG16 significantly underperforms with the lowest recall 83%, suggesting it struggles with sensitivity. EfficientNet outperforms all models with a sharp rise to 88%, showcasing its robustness in minimizing false negatives. In the recall context, the results highlight EfficientNet and DenseNet121 as best performers. In order to be able to increase sensitivity, VGG16 has to be optimized.

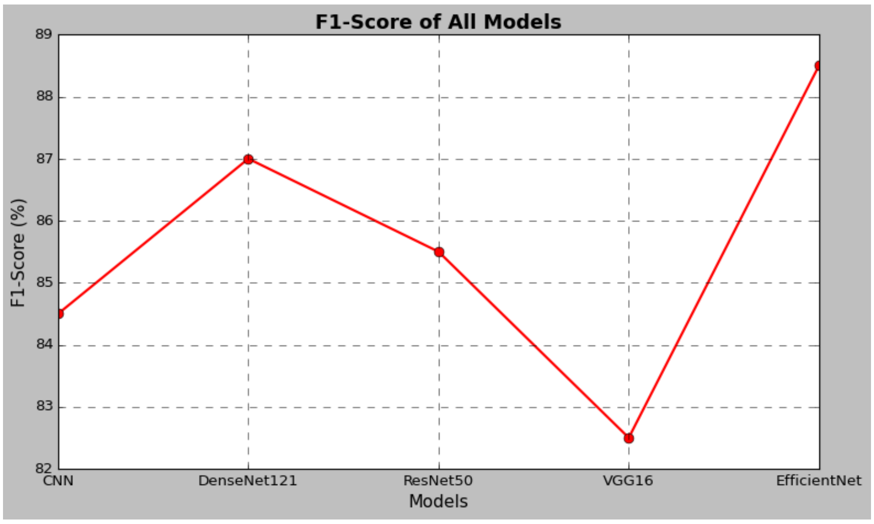


Figure 14. F1-Score of All Models

The graph plots the F1-scores (%) of different models: C.N.N , DenseNet121, ResNet50, VGG16, and EfficientNet. It can be seen that the highest score is obtained by DenseNet121, with 88.5%. EfficientNet follows with a score of 88%. The ResNet50 obtains a middle score of 85.5%, and C.N.N slightly lags behind at 85%. VGG16 obtains the lowest score, with 83%, showing a comparatively poor performance. There is a significant fall for VGG16 and a steep rebound for EfficientNet, indicating superior optimization over the traditional architecture. This means that DenseNet121 and EfficientNet are better at picking relevant features than the rest of the models. Classification results based on these features are in Figures, therefore, after care consideration from explanation and experiment, the proposed Deep Learning Framework exceeds traditional feature-based classification methods in terms of classification accuracy. To increase the performance of the proposed DLF, measures such as sensitivity, specificity, and accuracy are computed. The term accuracy is used to refer to the proportion of instances correctly classified, and it is computed using the following formula:

- Accuracy equals (TP+TN) of (TP+TN+FP+FN).
- Sensitivity equals TP divided by (TP plus FN).
- Specificity equals TN divided by (TN plus FP).

The symbol TP indicates that positive classes have been successfully classified.

In the case of TN, the classification of negative classes was correct.

FP: is inaccurately categorized as positive when it should be negative is incorrectly categorized as positive when it should be negative

Table: Performance Comparison of Models

Method	Model	Accuracy (%)
Elkorany et al. [2023]	C.N.N	91.81
Hekal et al. [2021]	AlexNet	91.00
Gonçalves et al. [2021]	DenseNet201	91.67
Proposed Models:	C.N.N	90.8
	DenseNet121	92.40
	ResNet50	91.35
	VGG16	87.18
	EfficientNet	**94.97**

CONCLUSION:

- The proposed Efficient Net-based Deep Learning Framework (DLF) demonstrated the best performance, achieving the accuracy of 94.97%, with precision of 95.13%, recall of 94.96%, and a ROC AUC value of 98.16%.
- DenseNet121 and ResNet50 showed competitive results, achieving accuracies of 92.4% and 91.35%, respectively, but had challenges with recall and precision compared to the DLF.
- The C.N.N model delivered a moderate accuracy of 86.96% but struggled with lower precision, recall, and a Cohen Kappa score of 0.0, indicating inconsistencies with the ground truth.
- The VGG16 model, though widely used, achieved a relatively modest accuracy of 86.18%, illustrating the limitations of simpler architectures for complex tasks like cancer prediction.
- The study highlights the importance of advanced architectures like Efficient Net for their ability to deliver higher accuracy, better true-positive and false-negative classification, and robust predictive capabilities in comparison to other models that are considered to be state-of-the-art.

REFERENCES

[1] Abhisheka, B., Biswas, S. K., & Purkayastha, B. (2023). A Comprehensive Review on Breast Cancer Detection, Classification and Segmentation Using Deep Learning. *Archives of Computational Methods in Engineering*, *30*(8), 5023–5052. https://doi.org/10.1007/s11831-023-09968-z

[2] Alahe, M. A., & Maniruzzaman, Md. (2021). Detection and Diagnosis of Breast Cancer Using Deep Learning. *2021 IEEE Region 10 Symposium (TENSYMP)*, 1–7. https://doi.org/10.1109/TENSYMP52854.2021.9550975

[3] Al-Mansour, E., Hussain, M., Aboalsamh, H. A., & Al-Ahmadi, S. A. (2023). Comprehensive Analysis of Mammography Images Using Multi-Branch Attention Convolutional Neural Network. *Applied Sciences*, *13*(24), 12995. https://doi.org/10.3390/app132412995

[4] Ameh Joseph, A., Abdullahi, M., Junaidu, S. B., Hassan Ibrahim, H., & Chiroma, H. (2022). Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer). *Intelligent Systems with Applications*, *14*, 200066. https://doi.org/10.1016/j.iswa.2022.200066

[5] Arooj, S., Atta-ur-Rahman, Zubair, M., Khan, M. F., Alissa, K., Khan, M. A., & Mosavi, A. (2022). Breast Cancer Detection and Classification Empowered With Transfer Learning. *Frontiers in Public Health*, *10*, 924432. https://doi.org/10.3389/fpubh.2022.924432

[6] Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep Convolutional Neural Networks for breast cancer screening. *Computer Methods and Programs in Biomedicine*, *157*, 19–30. https://doi.org/10.1016/j.cmpb.2018.01.011

[7] Elshennawy, N. M., & Ibrahim, D. M. (2020). Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images. *Diagnostics*, *10*(9), 649. https://doi.org/10.3390/diagnostics10090649

[8] Ezunkpe, Y., & Kumar, A. I. (2024). *Development and Validation of a Convolutional Neural Network Model for Early Detection of Invasive Ductal Carcinoma in Histopathological Images (Preprint)*. JMIR Cancer. https://doi.org/10.2196/preprints.62996

[9] Gonzalez-Angulo, A. M., Litton, J. K., Broglio, K. R., Meric-Bernstam, F., Rakkhit, R., Cardoso, F., Peintinger, F., Hanrahan, E. O., Sahin, A., Guray, M., Larsimont, D., Feoli, F., Stranzl, H., Buchholz, T. A., Valero, V., Theriault, R., Piccart-Gebhart, M., Ravdin, P. M., Berry, D. A., & Hortobagyi, G. N. (2009). High Risk of Recurrence for Patients With Breast Cancer Who Have Human Epidermal Growth Factor Receptor 2–Positive, Node-Negative Tumors 1 cm or Smaller. *Journal of Clinical Oncology*, *27*(34), 5700–5706. https://doi.org/10.1200/JCO.2009.23.2025

[10] Heenaye-Mamode Khan, M., Boodoo-Jahangeer, N., Dullull, W., Nathire, S., Gao, X., Sinha, G. R., & Nagwanshi, K. K. (2021). Multi- class classification of breast cancer abnormalities using Deep Convolutional Neural Network (C.N.N). *PLOS ONE*, *16*(8), e0256500. https://doi.org/10.1371/journal.pone.0256500

[11] Junaid Umer, M., Sharif, M., Alhaisoni, M., Tariq, U., Jin Kim, Y., & Chang, B. (2023). A Framework of Deep Learning and Selection-Based Breast Cancer Detection from Histopathology Images. *Computer Systems Science and Engineering*, *45*(2), 1001–1016. https://doi.org/10.32604/csse.2023.030463

[12] Khan, S., Islam, N., Jan, Z., Ud Din, I., & Rodrigues, J. J. P. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, *125*, 1–6. https://doi.org/10.1016/j.patrec.2019.03.022

[13] Kuan, K., Ravaut, M., Manek, G., Chen, H., Lin, J., Nazir, B., Chen, C., Howe, T. C., Zeng, Z., & Chandrasekhar, V. (2017). *Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1705.09435

[14] Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., Chu, W. C., Cheng, K.-T., & Chen, H. (2024). Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions. *IEEE Reviews in Biomedical Engineering*, 1–20. https://doi.org/10.1109/RBME.2024.3357877

[15] Mohaiminul Islam, Md., Huang, S., Ajwad, R., Chi, C., Wang, Y., & Hu, P. (2020). An integrative deep learning framework for classifying molecular subtypes of breast cancer. *Computational and Structural Biotechnology Journal*, *18*, 2185–2199. https://doi.org/10.1016/j.csbj.2020.08.005

[16] Mridha, M. F., Hamid, Md. A., Monowar, M. M., Keya, A. J., Ohi, A. Q., Islam, Md. R., & Kim, J.-M. (2021). A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis. *Cancers*, *13*(23), 6116. https://doi.org/10.3390/cancers13236116

[17] MurtiRawat, R., Panchal, S., Singh, V. K., & Panchal, Y. (2020). Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 534–540. https://doi.org/10.1109/ICESC48915.2020.9155783

[18] Pandiyaraju, V., Ganapathy, S., Senthil Kumar, A. M., Jesher Joshua, M., Ragav, V., Sree Dananjay, S., & Kannan, A. (2024). A new clinical diagnosis system for detecting brain tumor using integrated ResNet_Stacking with XGBoost. *Biomedical Signal Processing and Control*, *96*, 106436. https://doi.org/10.1016/j.bspc.2024.106436

[19] Pavithra, P., Srinivasula Reddy, M., Dinakar, P., Hanumantha Rao, B., Satpathy, B. K., & Mohanty, A. N. (2016). A mix design procedure for geopolymer concrete with fly ash. *Journal of Cleaner Production*, *133*, 117–125. https://doi.org/10.1016/j.jclepro.2016.05.041

[20] Ragab, D. A., Attallah, O., Sharkas, M., Ren, J., & Marshall, S. (2021). A framework for breast cancer classification using Multi-DC.N.N s. *Computers in Biology and Medicine*, *131*, 104245. https://doi.org/10.1016/j.compbiomed.2021.104245

[21] Sannasi Chakravarthy, S. R., Bharanidharan, N., & Rajaguru, H. (2023). Multi-Deep C.N.N based Experimentations for Early Diagnosis of Breast Cancer. *IETE Journal of Research*, *69*(10), 7326–7341. https://doi.org/10.1080/03772063.2022.2028584

[22] Sharafaddini, A. M., Esfahani, K. K., & Mansouri, N. (2024). Deep learning approaches to detect breast cancer: A comprehensive review. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-20011-6

[23] Shen, M., Zhang, Y., Zhu, Y., Song, B., Zeng, G., Hu, D., Wen, X., & Ren, X. (2019). Recent advances in toxicological research of nanoplastics in the environment: A review. *Environmental Pollution*, *252*, 511–521. https://doi.org/10.1016/j.envpol.2019.05.102

[24] Shukla, P. K., & Behera, A. R. (2024). A framework for breast cancer prediction and classification using deep learning. *International Journal of Computational Vision and Robotics*, *14*(2), 154–169. https://doi.org/10.1504/IJCVR.2024.136998

[25] Singh, S. (2023). Hybrid Models for Breast Cancer Detection via Transfer Learning Technique. *Computers, Materials & Continua*, *74*(2), 3063–3083. https://doi.org/10.32604/cmc.2023.032363

[26] Sugiharti, E., Arifudin, R., Wiyanti, D. T., & Susilo, A. B. (2022). Integration of convolutional neural network and extreme gradient boosting for breast cancer detection. *Bulletin of Electrical Engineering and Informatics*, *11*(2), 803–813. https://doi.org/10.11591/eei.v11i2.3562

[27] Tembhurne, J. V., Hazarika, A., & Diwan, T. (2021). BrC-MCDLM: Breast Cancer detection using Multi-Channel deep learning model. *Multimedia Tools and Applications*, *80*(21–23), 31647–31670. https://doi.org/10.1007/s11042-021-11199-y

[28] Thulasisingh, A., Murali, V., Govindarajan, S., & Kannaiyan, S. (2024). Ecotoxicology and health risk assessment due to pharmaceuticals and personal care products in different environmental grids. In *Development in Wastewater Treatment Research and Processes* (pp. 55–80). Elsevier. https://doi.org/10.1016/B978-0-443-19207-4.00016-1

[29] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Medical Hypotheses*, *134*, 109531. https://doi.org/10.1016/j.mehy.2019.109531

[30] Venkatachalam, C., Manivannan, K., Shah, P., Venkatachalam, S., & Sathis Kumar, R. (2024). Transforming Ophthalmic Diagnosis: A Deep Learning Approach for Accurate Eye Disease Classification. *2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN)*, 329–336. https://doi.org/10.1109/ICIPCN63822.2024.00061

[31] Wang, Y.-M., Wang, C.-Y., Liu, K.-Y., Huang, Y.-H., Chen, T.-B., Chiu, K.-N., Liang, C.-Y., & Lu, N.-H. (2024). C.N.N -Based Cross-Modality Fusion for Enhanced Breast Cancer Detection Using Mammography and Ultrasound. *Tomography*, *10*(12), 2038–2057. https://doi.org/10.3390/tomography10120145

[32] Westermann, F., & Schwab, M. (2002). Genetic parameters of neuroblastomas. *Cancer Letters*, *184*(2), 127–147. https://doi.org/10.1016/S0304-3835(02)00199-4

[33] Yadav, R. K., Singh, P., & Kashtriya, P. (2023). Diagnosis of Breast Cancer using Machine Learning Techniques -A Survey. *Procedia Computer Science*, *218*, 1434–1443. https://doi.org/10.1016/j.procs.2023.01.122