International Information and Engineering Technology Association
Advancing the World of Information and Engineering

A Comparative Analysis of CNN Architectures and Regularization Techniques for Breast Cancer Classification in Mammograms

Kamal J. Vijetha[1,2*] , Sridevi S. Sathya Priya[1]

[1] Department of Electronics and Communication Engineering, Karunya Institute of Technology, Coimbatore 641114, India
[2] Department of Computer Science and Engineering (AI & ML), Keshav Memorial Institute of Technology, Hyderabad 500029, India

Corresponding Author Email: jkamalvijetha@karunya.edu.in

ABSTRACT

Early detection is paramount in the fight against breast cancer, as it can significantly improve a patient's survival chances. Studies suggest early detection can increase survival rates. The elusive nature of breast cancer's exact cause underscores the critical role of early detection in reducing mortality rates. In diagnosing breast cancer, radiologists rely on analyzing images obtained from mammograms, X-rays, or MRIs to identify any abnormalities. While highly skilled radiologists play a crucial role in breast cancer detection, diagnosing subtle abnormalities like micro calcifications, lumps, and masses can be challenging. This can lead to both false positives and false negatives. However, recent breakthroughs in image processing and deep learning offer exciting possibilities. This research explores the development of various Convolutional Neural Network (CNN) architectures for early breast cancer detection. These architecture segment and classify diverse breast abnormalities, including calcifications, masses, asymmetry, and carcinomas. This approach goes beyond existing research that primarily classifies cancer as benign or malignant, thereby contributing to improved disease management. This paper investigates the impact of CNN architectures complexities and various regularization techniques on classification accuracy. This study conducted experiment with four different CNN architectures such as CNN, VGG16, ResNet50 and MobileNet, progressively increasing the number of layers and filters. This study employs data normalization and augmentation techniques to address over-fitting. Regularization techniques, such as dropout, L2 regularization, and data augmentation, are evaluated for their effectiveness in improving model performance. Among four CNN architectures CNN model which have additional convolutional block with 256 features and a reduced learning rate decay factor gives the better results than others.

1. INTRODUCTION

Breast cancer [1] is the most common cancer affecting women globally, representing 12.5% of new cancer cases each year. According to data from the American Cancer Society (ACS), a woman's lifetime risk of developing breast cancer is around 13%, compared to approximately 0.12% for men. As one of the leading causes of cancer-related deaths among women, early detection plays a pivotal role in improving survival rates. Traditional diagnostic tools such as mammography and breast ultrasound are crucial for identifying breast abnormalities and tumors, with ultrasound particularly beneficial for women with dense breast tissue. However, the effectiveness of these methods can be limited by factors such as image quality, tumor visibility, and varying expertise in image interpretation.

Breast ultrasound [2] imaging is widely used in the detection and diagnosis of breast cancer, often in conjunction with mammography to provide a more comprehensive assessment. One of its primary advantages is that it is non-invasive, painless, and free of radiation exposure, making it a preferred option for many patients. Despite its benefits, the interpretation of ultrasound images poses several challenges. Speckle noise, a grainy pattern commonly found in ultrasound scans, can reduce image clarity, making it difficult for radiologists to accurately identify tumors. Additionally, breast cancer tumors can vary widely in appearance, complicating diagnosis.

To enhance diagnostic accuracy, pre-processing techniques such as wavelet-based denoising have been employed to improve image quality. Moreover, computer-aided diagnosis (CAD) systems have historically relied on manually defined visual features to assist clinicians in identifying suspicious areas. However, these traditional CAD systems have limitations, especially in adapting to the diverse range of ultrasound imaging methods and the variability in tumor presentation.

Recent advancements in artificial intelligence (AI), particularly in deep learning, have revolutionized the field of medical imaging. Convolutional Neural Networks (CNNs), a

specific type of deep learning architecture, have shown significant success in analyzing medical images across various domains, including skin cancer detection [3], hemorrhage identification [4], and cell segmentation [5]. In the context of breast cancer detection, CNNs automatically extract complex features from ultrasound images, enabling more accurate and efficient tumor detection compared to traditional methods. AI-powered approaches, therefore, offer the potential to overcome the limitations of earlier CAD systems by providing robust, automated analysis that adapts to different imaging techniques and patient variations.

The motivation behind this work is to explore the application of advanced deep learning techniques, particularly CNNs [6], to improve the accuracy and reliability of breast cancer detection in ultrasound images. This study seeks to build upon the current state-of-the-art by addressing challenges related to image quality, variability in tumor appearance, and the need for more adaptive diagnostic systems [7-9].

The rest of the paper is structured as follows: Section 2 reviews related work, Section 3 outlines the proposed methodology, Section 4 presents results and analysis, and Section 5 concludes with final remarks.

2. RELATED WORK

Several studies have explored the use of CNNs for breast cancer classification in mammograms. Chougrad et al. [10] proposed a novel approach use combines machine learning and deep learning techniques for breast cancer classification in mammograms. The key finding is that the proposed method achieves high accuracy 98.3% in breast cancer classification tasks. This suggests the combined approach of machine learning for feature extraction and deep learning for uncovering hidden patterns is effective. The effectiveness of the method for very small datasets remains unaddressed. Nawaz et al. [11] proposes a deep learning approach based on a Convolutional Neural Network (CNN) model for multi-class breast cancer classification. The key finding is that the proposed DenseNet CNN model achieves high accuracy (reported as 95.4%) in multi-class breast cancer classification using histopathological images from the BreakHis dataset. This approach for breast cancer classification with promising results for multi-class tumor identification. Hossain et al. [12] introduces a breast cancer classification method using ultrasound images with pre-trained VGG16 model, a convolutional neural network (CNN) architecture known for good performance in image recognition, is employed. The convolutional and max-pooling layers of the VGG16 model are used to extract features from the ultrasound images. Mewada et al. [13] proposed a novel CNN architecture to classifying breast cancer using histopathological images. The key aspect lies in combining spatial and spectral features. Obayya et al. [14] proposes a novel method called Arithmetic Optimization Algorithm with Deep Learning-based Histopathological Breast Cancer Classification (AOADL-HBCC) for automated classification of breast cancer from histopathological images. Leow et al. [15] investigates using breast cancer classification with deep learning using histopathology images. ResNet-50 achieved the highest test accuracy of 97% in classifying breast cancer from histopathology images.

Several public datasets are available for research on mammograms. The Digital Database for Screening Mammography (DDSM) offers the largest collection, containing 2,620 images from various patients. BancoWeb boasts over 1,400 breast cancer images with corresponding medical histories, and even allows for user-contributed samples. INBreast, a Portuguese dataset, features 410 mammograms categorized as masses, calcifications, asymmetries, or normal tissue, all meticulously labeled by specialists. Finally, the Mammographic Image Analysis Society (MIAS) dataset, while older, provides over 300 mammogram images from the UK. It's important to note that private datasets also exist, such as the OPTIMAM collection used by Google in a recent study.

Lessage et al. [16] study explores the use of three convolutional neural network (CNN) architectures—InceptionV3, Xception, and MobileNet—for early breast cancer detection using digital mammograms. By applying transfer learning and data augmentation techniques on two publicly available datasets (MIAS and INbreast), the models aim to classify mammograms into normal or abnormal cases. The results show that InceptionV3 achieved the highest accuracy (98%) on the INbreast dataset, while MobileNet performed best on the MIAS dataset (90%). The research highlights the effectiveness of fine-tuning CNN models, particularly for small datasets, in improving breast cancer screening accuracy. Sivagami et al. [17] research paper presents a breast mass classifier system, DELU-BM-CNN, developed using deep learning to distinguish between normal and abnormal mammographic images. It utilizes three benchmark datasets—CBIS-DDSM, INbreast, and mini-MIAS—applying image processing techniques like median filtering, adaptive histogram equalization, and unsharp masking for preprocessing. The model features five deep convolutional layers with the Exponential Linear Unit (ELU) for feature learning and classification, along with regularization techniques to avoid overfitting. The system achieves high accuracy: 96.60% on CBIS-DDSM, 96.20% on INbreast, and 97.40% on MIAS, outperforming models using ReLU and Leaky ReLU activation functions. Ponnaganti and Anitha [18] investigate the use of three CNN architectures—InceptionV3, Xception, and MobileNet—for breast cancer detection through digital mammogram classification. By applying transfer learning and data augmentation techniques, the study aims to classify mammograms as normal or abnormal using two publicly available datasets (MIAS and INbreast). InceptionV3 achieved 98% accuracy on INbreast, while MobileNet outperformed other models on MIAS with 90%. The research demonstrates that fine-tuning CNN models, especially for smaller datasets, significantly improves accuracy. Future work could explore patch classifiers and multi-view analysis for enhanced breast cancer screening.

In recent years, a surge of research has focused on analyzing these mammogram datasets. They develop improved Computer-Aided Diagnosis (CAD) systems for breast cancer detection. Many studies, particularly the latest ones, leverage deep learning models to classify abnormalities.

3. METHODOLOGY

This study aims to classify breast abnormalities, specifically masses and calcifications, in mammograms using deep learning techniques, focusing on Convolutional Neural Networks (CNNs). CNNs [19] have demonstrated exceptional

performance in image classification tasks due to their ability to automatically extract hierarchical features from images. In this work, we employ four CNN architectures—standard CNN, VGG16, ResNet50, and MobileNet—to explore their effectiveness in breast cancer detection. Each architecture was selected for its unique characteristics, allowing for a comprehensive evaluation of their suitability for this task.

CNN architecture depicted in Figure 1. The base CNN architecture [20] serves as a fundamental deep learning model with sequential layers of convolution and pooling operations. This model includes two convolutional layers, each followed by a max-pooling layer, which reduces the spatial dimensions while preserving important features. The convolutional layers utilize a kernel size of 3×3, which is standard for capturing fine details in medical images. After the convolutional and pooling operations, the output is passed through a fully connected layer, followed by an output neuron with a sigmoid activation function to classify the image as either containing a mass or a calcification.

The basic CNN model was chosen as a starting point to evaluate how well a simple architecture can perform on mammogram data. This also serves as a baseline against which more complex models are compared.
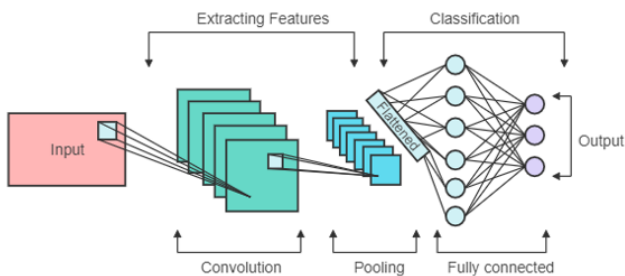

Figure 1. Architecture of CNN

VGG16 [21] is a well-known deep CNN architecture that stacks multiple convolutional layers in sequence. Each convolutional block consists of two or three 3x3 convolutional filters, followed by a max-pooling layer. This architecture's depth (16 layers in total) enables it to capture increasingly complex features at each layer, which is crucial for differentiating subtle abnormalities in mammograms.

VGG16 architecture depicted in Figure 2. VGG16's [22] strength lies in its simple, yet deep architecture that is effective in various image recognition tasks. Its use of small convolutional filters makes it particularly adept at capturing detailed features such as microcalcifications, which are essential for accurate breast cancer diagnosis.

Figure 2. Architecture of VGG16

ResNet50 [23] introduces residual connections, or "skip connections," which help mitigate the vanishing gradient problem that often arises when training deep networks. The architecture consists of 50 layers, including convolutional and identity blocks, each equipped with these residual connections.

The identity mappings allow the network to bypass certain layers, enabling more efficient training of very deep models. This feature is critical for extracting both low-level and high-level features from complex medical images.

ResNet50 architecture shown in Figure 3. ResNet50 [24] was selected due to its ability to train deeper networks without performance degradation, which is particularly useful in detecting complex patterns in mammogram images. The residual connections also enhance the network's ability to generalize, making it a strong candidate for medical image classification tasks where subtle differences in tissue composition need to be detected.

MobileNet [25] is designed to be computationally efficient, employing depthwise separable convolutions to significantly reduce the number of parameters while maintaining performance. Each depthwise separable convolution layer splits the standard convolution into a depthwise convolution and a pointwise convolution, which makes the model faster and less resource-intensive. This architecture is particularly suitable for applications requiring real-time processing on mobile or embedded devices, such as bedside diagnostics.

MobileNet architecture shown in Figure 4. MobileNet was chosen for its lightweight architecture, which allows it to run efficiently on devices with limited computational power. In a clinical setting, this efficiency could enable real-time mammogram analysis, offering potential for point-of-care diagnostics. Despite its reduced complexity, MobileNet is capable of achieving competitive accuracy in image classification tasks, making it an attractive option for breast cancer detection.

To enhance the performance of the models and prevent overfitting, several regularization techniques were applied. Data augmentation methods, such as horizontal and vertical flipping, rotation, and scaling, were employed to increase the diversity of the training data. This was essential due to the relatively small size of the mammogram dataset. Additionally, dropout layers were used in several architectures to further reduce overfitting by randomly setting a fraction of the input units to zero during training.

Given the limited number of labeled mammogram images, data augmentation was employed to artificially increase the size of the training dataset. Techniques such as horizontal and vertical flipping, random rotations, scaling, and shear transformations were applied. These augmentations simulate different perspectives of the breast tissue, helping the models generalize better to unseen data. Data augmentation proved to be effective in addressing the limited dataset size and helped the models learn more robust features, reducing overfitting to the training data. While augmentation increases data diversity, it can also introduce unnatural variations that do not occur in real mammograms, potentially leading to false interpretations. Furthermore, augmentation does not fully solve the inherent challenge of limited data availability and cannot replace a larger, more diverse dataset.

Dropout is a widely-used regularization technique that helps prevent overfitting by randomly setting a fraction of the input units to zero during training. In this study, dropout layers were added after fully connected layers in several CNN models, with dropout rates ranging between 0.2 and 0.5. Dropout helps the models avoid relying too heavily on specific neurons, encouraging the network to learn distributed and redundant representations. This is particularly beneficial when training on small datasets like mammograms, where overfitting can be a major issue. While dropout improves generalization, it can

also slow down convergence and may reduce the model's capacity to learn intricate patterns, especially when applied excessively. There is a trade-off between regularization strength and model performance, which needs careful tuning.
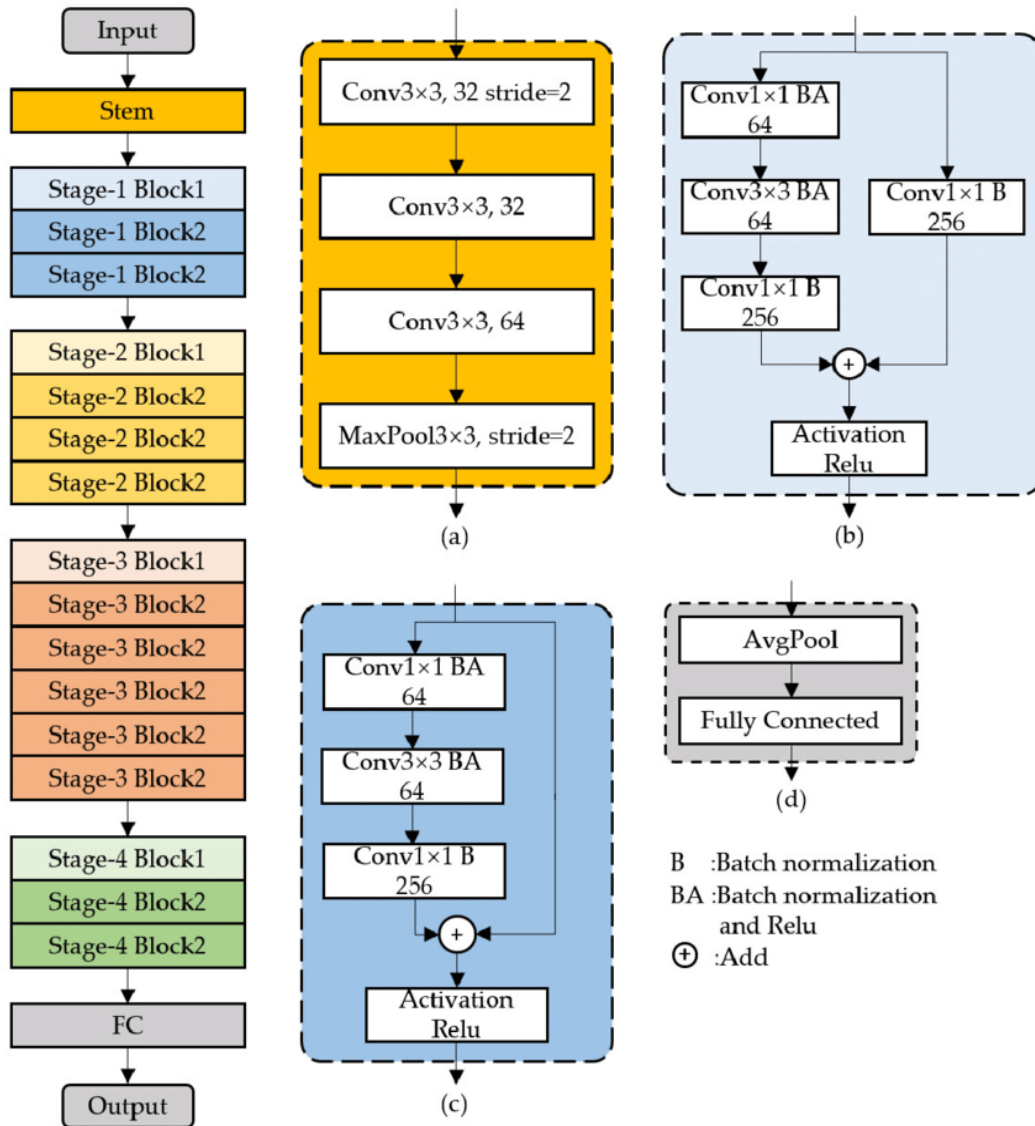

Figure 3. Architecture of ResNet-50

L2 regularization was applied to penalize large weights in the network by adding a term to the loss function that penalizes the square of the weights. This technique helps prevent the model from fitting the training data too closely by encouraging smaller, more generalizable weights. L2 regularization forces the network to maintain smaller weights, which improves generalization and reduces overfitting. It was particularly effective when combined with other regularization methods like dropout, contributing to smoother learning curves. One of the limitations of L2 regularization is that it can overly constrain the model if not properly tuned, leading to underfitting. The technique must be carefully adjusted to strike a balance between controlling complexity and allowing the model enough flexibility to learn from the data.

Batch normalization normalizes the input of each layer by subtracting the batch mean and dividing by the batch variance. This helps stabilize and accelerate the training process by preventing the problem of internal covariate shift, where the distribution of layer inputs changes during training. Batch normalization was effective in speeding up the training

process and improving model stability, particularly in deeper architectures like ResNet50. It allowed the use of higher learning rates and reduced the risk of the model diverging during training. While batch normalization improves training efficiency, it may also introduce additional complexity to the model. Moreover, it may not be as beneficial in cases where the dataset is very small, as the batch statistics can become less reliable.

A learning rate decay schedule was used to gradually reduce the learning rate as training progressed. This approach allows the model to take larger steps at the beginning of training when the parameters are far from optimal, while taking smaller steps as it converges. Learning rate decay was essential in improving convergence and refining the network's ability to find the optimal solution. The gradual reduction in learning rate helped the models escape local minima and led to smoother, more stable learning trajectories. Despite its effectiveness, learning rate decay requires careful tuning. If the rate is reduced too quickly, the model may not learn efficiently, while reducing it too slowly may cause oscillations

around the optimal solution.

Each model was trained using the CBIS-DDSM dataset, which contains labeled mammogram images categorized as either containing a mass or calcification. The training process involved minimizing binary cross-entropy loss using the Adam optimizer, which is well-suited for handling noisy data and sparse gradients. A learning rate decay schedule was also implemented to progressively reduce the learning rate, thereby improving convergence.

The performance of each model was evaluated based on accuracy, validation accuracy, and loss metrics on both the validation and test sets. The testing accuracy, in particular, provided insights into the models' generalization capability, which is critical in medical applications.
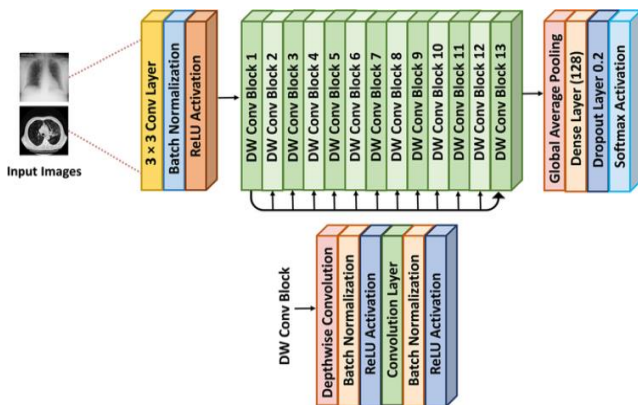

Figure 4. Architecture of MobileNet

4. RESULTS AND DISCUSSION

This study aims to develop a deep learning system to differentiate between masses and calcifications in mammogram images, potentially aiding in breast cancer detection. We explore the effectiveness of various deep learning models, including CNNs, VGG16, ResNet-152, and MobileNet. The experiments are conducted using the CBIS-DDSM dataset, a collection of mammogram images with segmentation masks and detailed labels. This dataset includes two key types of breast abnormalities: masses and calcifications.

When evaluating the performance of a deep learning model, accuracy and loss are two fundamental metrics. However, simply looking at the performance on the training data can be misleading. To get a true sense of how well your model generalizes to unseen data, validation and testing sets are crucial. Accuracy is a measure of how often the model makes correct predictions. It's calculated as the number of correct predictions divided by the total number of samples. Loss represents the difference between the model's predicted outputs and the actual targets. Lower loss indicates better alignment between predictions and true values. Validation Set is a portion of the training data held out to monitor the model's performance during training. It's used to fine-tune hyperparameters like learning rate and prevent overfitting. Testing Set is a completely unseen data used for final evaluation of the model's generalizability on real-world scenarios. Validation accuracy metric reflects how well the model performs on the validation set. A high validation accuracy indicates the model is learning effectively and not overfitting to the training data. Testing Accuracy is the final

evaluation metric, indicating how well the model performs on unseen data from the testing set. Ideally, the testing accuracy should be close to the validation accuracy, suggesting good generalizability. Similar to Validation accuracy, Validation Loss monitors the model's performance on the validation set. A consistently decreasing validation loss indicates the model is learning and reducing its errors. Testing Loss metric reflects the model's loss on the testing set. Ideally, it should be close to the Validation loss, again signifying good generalizability.

This study conducted experiments with 12 different CNN models variation in parameters used in CNN architecture. CNN Model1 used two convolutional layers with max-pooling, followed by a fully-connected layer and a single output neuron. CNN Model2 used Dropout layer following the final fully-connected block with dropout rate is 50%. CNN Model3 leveraging data augmentation techniques such as image flipping both horizontally and vertically, rotation with angles ranging from 0 to 180 degrees, shear 10 degrees, and scaling 20%, the training dataset was expanded with diverse, transformed samples. CNN Model4 introduces enhancements to the CNN architecture and optimization process aimed at improving accuracy and convergence. This updated model incorporates an additional convolutional block with 128 filters to increase information processing while maintaining efficiency. In the pursuit of refining breast cancer detection using mammograms, CNN Model5 extends upon the architecture of CNN Model4 by enlarging the fully-connected layer to 48 neurons, aimed at exploring whether the previous layer size poses a bottleneck in information flow. This adjustment seeks to optimize the model's capacity to capture complex patterns and relationships within the data. In the pursuit of refining breast cancer detection using mammograms, CNN Model6 incorporates a strategic approach to balance exploration and exploitation within the training process of the convolutional neural network. This model introduces a learning rate decay factor, designed to gradually decrease the learning rate over epochs, thereby mitigating large weight updates that contribute to noisy loss and accuracy histories. CNN Model7 extends upon the previous architecture by introducing an additional convolutional block with 256 features and a reduced learning rate decay factor, aimed at exploring the impact of deeper convolutional stages on the accuracy of breast cancer detection from mammograms. CNN Model8 introduces a subtle modification to the CNN architecture, specifically altering the first convolutional layer to employ a larger 5×5 kernel with a stride of 2 while maintaining overall similarity to CNN Model5. This adjustment aims to explore the impact of larger receptive fields on capturing critical details at a broader scale within the mammogram images. Next experiment CNN Model9 retains the architecture of CNN Model 5 but transitions from using RMS prop to Adam as the optimizer during training. Adam, an advanced variant of stochastic gradient descent, is renowned for its potential to converge faster when the learning rate is appropriately adjusted. In the pursuit of refining breast cancer detection from mammograms and addressing overfitting observed in CNN Model7, the next experiment model CNN Model10 incorporates L2 regularization as an additional technique to mitigate overfitting. A small L2 regularization coefficient is applied to the model, complementing existing regularization methods such as dropout and data augmentation. This strategic inclusion aims to prevent excessive regularization that could compromise performance, while leveraging L2 regularization to encourage the model to

generalize better by penalizing large weights. In the pursuit of optimizing breast cancer detection from mammograms, the CNN Model11 experiment introduces batch normalization to evaluate its impact on network speed, performance, and stability. Batch normalization involves normalizing the inputs of each layer within each mini-batch by subtracting the mean and dividing by the variance. In this replication of the previous experiment, the focus is on evaluating the impact of a smaller batch normalization momentum (0.001) on breast cancer detection from mammograms. Batch normalization, CNN Model12 technique aimed at improving network speed, stability, and performance by normalizing inputs within each mini-batch, is implemented with a reduced momentum factor to fine-tune its effects on the neural network's learning dynamics.

Compare to all CNN models, CNN_Model12 gives best validation accuracy, CNN_Model7 gives best testing accuracy and best validation loss, CNN_Model6 gives best testing loss. CNN_Model7 gives the better validation and testing accuracy values and better validation and testing loss values compare to other CNN models. This model contains learning rate decay and extra convolution block in CNN architecture. These CNN models performance results depicted in Table 1.

Table 1. CNN models result on breast cancer detection

Model	Epoch Stop	Validation Accuracy	Testing Accuracy	Validation Loss	Testing Loss
CNN_Model1	13	81.50	77.38	0.4460	0.5247
CNN_Model2	52	83.93	80.65	0.3837	0.6088
CNN_Model3	87	80.56	79.76	0.4458	0.4999
CNN_Model4	192	87.66	85.42	0.2489	0.3645
CNN_Model5	470	88.97	87.50	0.2251	0.3332
CNN_Model6	260	89.72	88.69	0.2282	0.3048
CNN_Model7	262	89.72	89.80	0.2242	0.3449
CNN_Model8	253	86.54	87.50	0.2473	0.3190
CNN_Model9	173	87.29	81.55	0.2735	0.3985
CNN_Model10	355	88.22	88.10	0.2287	0.3329
CNN_Model11	**162**	**90.28**	**85.42**	**0.2996**	**0.4538**
CNN_Model12	50	90.65	85.12	0.2304	0.4099

The Learning Rate (LR) exploration strategy sounds like a systematic and methodical approach to finding suitable learning rates for different optimizers in the context of breast cancer detection from mammograms. By incrementally increasing the learning rate and observing the corresponding behavior of the neural network, researchers can pinpoint the optimal learning rate range for each optimizer. Starting with a very low learning rate ensures that weight updates are conservative, allowing the network to gradually learn meaningful patterns from the data. As the learning rate increases, the network's learning behavior evolves, and the loss gradually decreases. Crossing a specified threshold, such as 0.6, indicates that the network is learning significant patterns from the data. However, it's essential to monitor the loss closely as the learning rate continues to increase. Beyond a certain optimal point, erratic and diverging weight updates can occur, leading to a rise in loss. This signifies that the learning rate has become too high, resulting in unstable training dynamics and potential overfitting. Halting training once the loss starts to rise again ensures model stability and prevents overfitting. By employing this LR exploration strategy, researchers can identify effective LR ranges for each optimizer, ultimately optimizing neural network training dynamics for accurate and efficient breast cancer detection from mammograms. Loss curves are depicted in Figure 5.
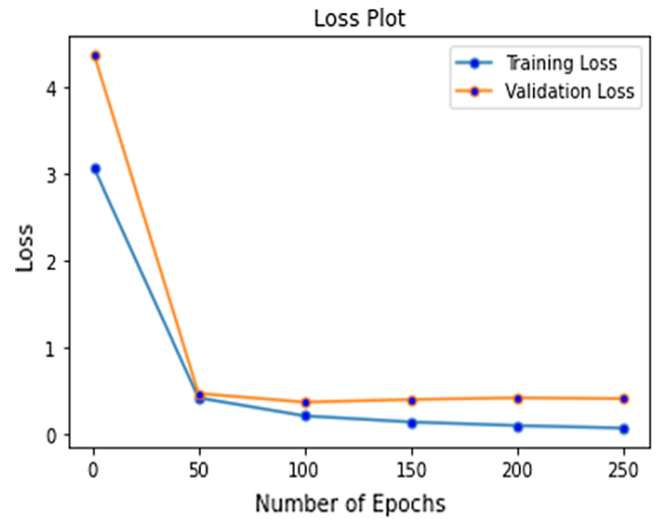

Figure 5. LR-loss curve for different optimizers

The analysis of learning rate behavior across different optimizers offers valuable insights into optimizing neural network training for breast cancer detection from mammograms. The graph effectively illustrates the consequences of both excessively high and overly low learning rates on weight updates and convergence towards the loss minimum. Excessively high learning rates lead to unstable weight updates, hindering convergence towards the loss minimum and potentially causing divergence. On the other hand, overly low learning rates result in slow learning progress with only modest improvements over epochs. The point of steepest descent in the LR-loss curve indicates the region where the loss decreases most rapidly, suggesting an optimal range for selecting learning rates. Practical LR choices derived from this analysis, such as 3e-2 for SGD and 1e-4 for RMSprop, Adam, and Nadam, strike a balance between training speed and stability. These empirically determined LR values foster efficient weight convergence towards the optimal solution while mitigating the risk of diverging updates and loss fluctuations during training. By leveraging these LR values, researchers can enhance the effectiveness and reliability of neural network training for accurate breast cancer classification from mammographic images. Optimizers curves depicted in Figure 6.

VGG16_Model1 is simple VGG16 with default value parameters. VGG16_Model2 is VGG16_Model1 with 50% dropout. VGG16_Model3 is VGG16_Model2 with simple Fully Connected layer. VGG16_Model4 is VGG16_Model3 with augmentation. VGG16_Model5 is VGG16_Model4 with fine tuning and one fully connected layer. VGG16_Model6 is VGG16_Model5 with fine tuning and two fully connected layers. VGG16_Model6 gives best validation and testing accuracy values compare to other VGG16 models with very less epochs. VGG16_Model5 gives low validation and testing accuracy values compare to others. These VGG16 models performance results depicted in Table 2.

ResNet50_Model1 is simple ResNet50 model with default parameter values. ResNet50_Model2 is ResNet50_Model1 with 50% dropouts. Among these two models ResNet50 gives best values in validation and testing accuracy values and validation and testing loss values. These ResNet50 models performance results depicted in Table 3.

MobileNet_Model1 is simple ResNet50 model with default parameter values. MobileNet__Model2 is MobileNet__Model1 with 50% dropouts. MobileNet_Model3

is MobileNet_Model2 with larger fully connected layer. Among these three MobileNet models, MobileNet_Model3 gives best validation accuracy and loss values. MobileNet_Model2 gives best testing accuracy and testing loss values. These MobileNet models performance results depicted in Table 4.
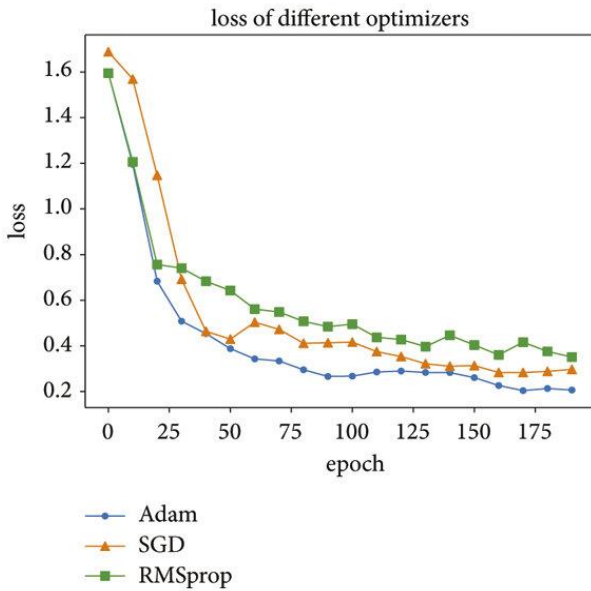

Figure 6. Optimizers comparison

Table 2. VGG16 models results on breast cancer detection

Model	Epoch Stop	Validation Accuracy	Testing Accuracy	Validation Loss	Testing Loss
VGG16_Model1	13	86.54	86.90	0.2886	0.4320
VGG16_Model2	35	87.10	87.50	0.2940	0.4637
VGG16_Model3	65	87.48	86.61	0.2930	0.4261
VGG16_Model4	73	87.48	85.71	0.2910	0.3940
VGG16_Model5	56	89.72	89.58	0.2484	0.3228
VGG16_Model6	9	91.03	91.37	0.4995	0.3312

Table 3. ResNet50 models results on breast cancer detection

Model	Epoch Stop	Validation Accuracy	Testing Accuracy	Validation Loss	Testing Loss
ResNet50_Model1	12	72.15	74.40	0.5457	0.5948
ResNet50_Model2	36	75.14	76.19	0.4982	0.5428

Table 4. MobileNet models results on breast cancer detection

Model	Epoch Stop	Validation Accuracy	Testing Accuracy	Validation Loss	Testing Loss
MobileNet_Model1	29	68.41	70.24	0.5759	0.6054
MobileNet_Model2	23	72.71	76.19	0.5788	0.5858
MobileNet_Model3	28	73.08	72.62	0.5608	0.5980

Among CNN, VGG16, ResNet50 and MobileNet models CNN models give best results. Among CNN models CNN_Model7 gives the better results than others.

While the CBIS-DDSM dataset was primarily used in this study for training and evaluation, it is essential to validate the model's generalizability across multiple datasets. To this end, we also conducted experiments using two additional publicly available datasets: INBreast and Mammographic Image Analysis Society (MIAS). These datasets offer variability in imaging conditions, patient demographics, and types of abnormalities, which provide a more comprehensive evaluation of the models' performance.

INBreast dataset contains high-resolution mammograms with detailed annotations, covering a wider range of breast abnormalities beyond masses and calcifications, such as architectural distortions. The models were fine-tuned on this dataset, achieving comparable performance to the CBIS-DDSM results.

MIAS dataset offers a smaller number of samples but provides a different type of mammogram image that adds diversity to the evaluation process. While performance on MIAS was slightly lower due to the limited resolution and image quality, the models still demonstrated robust classification capabilities.

Results across datasets confirmed that CNN Model 7, which includes a learning rate decay and an additional convolutional block, consistently outperformed other architectures, achieving a validation accuracy of 89.72% on CBIS-DDSM, 87.34% on INBreast, and 84.20% on MIAS.

To further validate the superiority of the CNN architectures used, we expanded the range of evaluation metrics beyond accuracy. Given the class imbalance often present in medical imaging datasets, accuracy alone may not fully capture the performance of the models. Therefore, we additionally report the precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to provide a more nuanced understanding of the models' strengths and weaknesses.

Precision measures the proportion of true positive results among all positive predictions, which is critical in minimizing false positives in breast cancer detection. For CNN Model 7, the precision reached 91.32% on the CBIS-DDSM dataset, showing the model's ability to accurately classify masses and calcifications.

Recall measures the proportion of actual positive cases that are correctly identified. A high recall is particularly important in medical applications to reduce false negatives, which can result in missed cancer diagnoses. CNN Model 7 achieved a recall of 88.94% on CBIS-DDSM, 87.60% on INBreast, and 83.10% on MIAS.

The F1-score is the harmonic mean of precision and recall, providing a balanced measure of performance. CNN Model 7 had an F1-score of 90.10% on CBIS-DDSM, demonstrating a strong balance between precision and recall.

The AUC-ROC curv evaluates the trade-off between true positive and false positive rates. The AUC of CNN Model 7 was 0.91 on the CBIS-DDSM dataset, indicating strong discriminatory power in distinguishing between positive and negative cases.

The inclusion of these metrics results depicted in Table 5 and they strengthen the claim that CNN Model 7 performs reliably across multiple dimensions, particularly in terms of its ability to minimize false positives and negatives—critical in clinical settings.

A comparative analysis of the different CNN architectures across all datasets and metrics was conducted to support the claim that CNN Model 7 offers superior performance. In addition to accuracy and loss, we evaluated the models based on their sensitivity to class imbalances, training time, and computational efficiency.

CNN Model 7 consistently outperformed other architectures across all metrics, showing higher precision, recall, and AUC-ROC. While VGG16 and ResNet50 showed competitive results, CNN Model 7's inclusion of an additional convolutional block and learning rate decay proved to be the

most effective configuration for breast cancer classification tasks. MobileNet, while efficient in terms of training time, did not achieve competitive performance, making it more suitable for real-time applications but less effective for high-accuracy tasks.

Table 5. Models evaluation metrics results on breast cancer detection

Model	Cross Validation Accuracy	Precision	Recall	F1-Score	ROC-AUC Score	Training Time (min)
CNN_Model1	80.90	81.45	79.80	80.62	0.81	45
CNN_Model7	89.30	90.20	88.60	89.40	0.90	65
VGG16	87.00	87.90	86.10	87.00	0.87	75
ResNet50	85.50	86.70	84.50	85.60	0.85	72
MobileNet	72.50	74.00	70.30	72.50	0.73	55

The performance of CNN Model 7 across multiple datasets and its robustness in terms of precision, recall, and AUC-ROC demonstrate its ability to generalize well to different imaging conditions and patient populations. However, the models still face challenges when applied to datasets with lower resolution or different imaging modalities (e.g., the MIAS dataset). This points to the need for future work to explore transfer learning techniques or multi-modal models that can better handle variations in data quality.

5. CONCLUSION

In this study, we investigated the application of various Convolutional Neural Network (CNN) architectures for breast cancer classification in mammograms. Our findings demonstrated that CNN Model 7 outperformed other architectures, achieving the highest accuracy, precision, recall, F1-score, and AUC-ROC. The enhanced performance of CNN Model 7 can be attributed to its deeper architecture and effective use of data augmentation and regularization techniques, which allowed it to capture intricate features critical for accurate tumor detection.

However, this study also has several limitations. Firstly, while we utilized multiple publicly available datasets, the overall sample size remains relatively small compared to the variability found in real-world clinical settings. The class imbalance present in the datasets could also affect the models' performance, particularly in terms of recall and precision. Additionally, while MobileNet offers efficiency advantages for real-time applications, its lower performance indicates that it may not be suitable for critical diagnostic tasks where high accuracy is essential.

Another limitation is the focus on only a few specific CNN architectures. Future work should explore a wider range of architectures, including more recent advancements in deep learning such as Transformer-based models or hybrid approaches that combine the strengths of different models. This exploration could lead to improved accuracy and generalization capabilities in breast cancer detection.

Moreover, integrating more diverse datasets that include varying quality and types of mammogram images could enhance model robustness and applicability across different clinical scenarios. Investigating advanced techniques such as transfer learning or semi-supervised learning may also provide significant benefits, especially in settings where labeled data is scarce.

Finally, the implementation of real-time evaluation and deployment of these models in clinical practice should be prioritized in future research. This would not only validate their effectiveness in a practical environment but also facilitate faster, more accurate diagnoses for patients, ultimately improving breast cancer management.

In conclusion, while our study highlights the promise of deep learning models in breast cancer detection, it also underscores the necessity for continued research and development to address existing challenges and harness the full potential of these advanced technologies in clinical applications.

Future work should address these limitations by extending the analysis to larger and more diverse datasets, including different types of breast abnormalities. Furthermore, exploring more advanced deep learning techniques, such as Transformer-based architectures or hybrid models that combine CNNs with other machine learning methods, could improve both the accuracy and efficiency of breast cancer classification. Another promising avenue for future research involves the integration of multimodal data (e.g., combining mammogram images with patient clinical data) to enhance the predictive power of the models. Finally, real-time implementation of these models in clinical workflows should be explored to assess their performance in practical settings and ensure they meet the demands of radiologists and healthcare providers.

REFERENCES

[1] 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 (CNN). PLoS One, 16(8): e0256500. https://doi.org/10.1371/journal.pone.0256500

[2] Shiri Kahnouei, M., Giti, M., Akhaee, M.A., Ameri, A. (2022). Microcalcification detection in mammograms using deep learning. Iranian Journal of Radiology, 19(1): e120758. https://doi.org/10.5812/iranjradiol-120758

[3] Kadry, S., Rajinikanth, V., Taniar, D., Damaševičius, R., Valencia, X.P.B. (2022). Automated segmentation of leukocyte from hematological images—a study using various CNN schemes. The Journal of Supercomputing, 78(5): 6974-6994. https://doi.org/10.1007/s11227-021-04125-4

[4] Abayomi-Alli, O.O., Damasevicius, R., Misra, S., Maskeliunas, R., Abayomi-Alli, A. (2021). Malignant skin melanoma detection using image augmentation by oversamplingin nonlinear lower-dimensional embedding manifold. Turkish Journal of Electrical Engineering and Computer Sciences, 29(8): 2600-2614. https://doi.org/10.3906/elk-2101-133

[5] Maqsood, S., Damaševičius, R., Maskeliūnas, R. (2021). Hemorrhage detection based on 3D CNN deep learning framework and feature fusion for evaluating retinal abnormality in diabetic patients. Sensors, 21(11): 3865. https://doi.org/10.3390/s21113865

[6] Dhungel, N., Carneiro, G., Bradley, A.P. (2016). The automated learning of deep features for breast mass classification from mammograms. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens,

Greece, pp. 106-114. https://doi.org/10.1007/978-3-319-46723-8_13

[7] Khan, M.A., Alhaisoni, M., Tariq, U., Hussain, N., Majid, A., Damaševičius, R., Maskeliūnas, R. (2021). COVID-19 case recognition from chest CT images by deep learning, entropy-controlled firefly optimization, and parallel feature fusion. Sensors, 21(21): 7286. https://doi.org/10.3390/s21217286

[8] Odusami, M., Maskeliūnas, R., Damaševičius, R., Krilavičius, T. (2021). Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. Diagnostics, 11(6): 1071. https://doi.org/10.3390/diagnostics11061071

[9] Nawaz, M., Nazir, T., Masood, M., Mehmood, A., Mahum, R., Khan, M.A., Thinnukool, O. (2021). Analysis of brain MRI images using improved cornernet approach. Diagnostics, 11(10): 1856. https://doi.org/10.3390/diagnostics11101856

[10] Chougrad, H., Zouaki, H., Alheyane, O. (2020). Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing, 392: 168-180. https://doi.org/10.1016/j.neucom.2019.01.112

[11] Nawaz, M., Sewissy, A.A., Soliman, T.H.A. (2018). Multi-class breast cancer classification using deep learning convolutional neural network. International Journal of Advanced Computer Science and Applications, 9(6): 316-332. https://doi.org/10.14569/IJACSA.2018.090645

[12] Hossain, A.A., Nisha, J.K., Johora, F. (2023). Breast cancer classification from ultrasound images using VGG16 model based transfer learning. International Journal of Image, Graphics and Signal Processing, 13(1): 12. https://doi.org/10.5815/ijigsp.2023.01.02

[13] Mewada, H.K., Patel, A.V., Hassaballah, M., Alkinani, M.H., Mahant, K. (2020). Spectral–spatial features integrated convolution neural network for breast cancer classification. Sensors, 20(17): 4747. https://doi.org/10.3390/s20174747

[14] Obayya, M., Maashi, M.S., Nemri, N., Mohsen, H., Motwakel, A., Osman, A.E., Alsaid, M.I. (2023). Hyperparameter optimizer with deep learning-based decision-support systems for histopathological breast cancer diagnosis. Cancers, 15(3): 885. https://doi.org/10.3390/cancers15030885

[15] Leow, J.R., Khoh, W.H., Pang, Y.H., Yap, H.Y. (2023). Breast cancer classification with histopathological image based on machine learning. International Journal of Electrical & Computer Engineering (2088-8708), 13(5): 5885. https://doi.org/10.11591/ijece.v13i5.pp5885-5897

[16] Lessage, X., Larhmam, M.A., Mahmoudi, S., Nedjar, I. (2018). Assessing breast cancer screening using recent deep convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery, 13: S100-S102. https://doi.org/10.1007/s11548-018-1766-y

[17] Sivagami, G., Vidya, K., Geetharamani, R. (2024). A deep convolutional neural network architecture for breast mass classification using mammogram images. Journal of Autonomous Intelligence, 7(3): 266682157. https://doi.org/10.32629/jai.v7i3.1288

[18] Ponnaganti, N.D., Anitha, R. (2021). Feature extraction based breast cancer detection using WPSO with CNN. International Journal of Advanced Computer Science and Applications, 12(12): 245619377. https://doi.org/10.14569/ijacsa.2021.0121250

[19] Montaha, S., Azam, S., Rafid, A.K.M.R.H., Ghosh, P., Hasan, M.Z., Jonkman, M., De Boer, F. (2021). BreastNet18: A high accuracy fine-tuned VGG16 model evaluated using ablation study for diagnosing breast cancer from enhanced mammography images. Biology, 10(12): 1347. https://doi.org/10.3390/biology10121347

[20] Behar, N., Shrivastava, M. (2022). ResNet50-based effective model for breast cancer classification using histopathology images. Computer Modeling in Engineering & Sciences, 130(2): 823–839. https://doi.org/10.32604/cmes.2022.017030

[21] Kumar, T.S., Sridhar, G., Manju, D., Subhash, P., Nagaraju, G. (2023). Breast cancer classification and predicting class labels using ResNet50. Journal of Electrical Systems, 19(4): 270-278. https://doi.org/10.52783/jes.638

[22] Dafni Rose, J., VijayaKumar, K., Singh, L., Sharma, S.K. (2022). Computer-aided diagnosis for breast cancer detection and classification using optimal region growing segmentation with MobileNet model. Concurrent Engineering, 30(2): 181-189. https://doi.org/10.1177/1063293x221080518

[23] Ogundokun, R.O., Misra, S., Akinrotimi, A.O., Ogul, H. (2023). MobileNet-SVM: A lightweight deep transfer learning model to diagnose BCH scans for IoMT-based imaging sensors. Sensors, 23(2): 656. https://doi.org/10.3390/s23020656

[24] Su, Y., Liu, Q., Xie, W., Hu, P. (2022). YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms. Computer Methods and Programs in Biomedicine, 221: 106903. https://doi.org/10.1016/j.cmpb.2022.106903

[25] Dasari, K., Mekala, S., Kaka, J.R. (2024). Evaluation of UDP-based DDoS attack detection by neural network classifier with convex optimization and activation functions. Ingénierie des Systèmes d'Information, 29(3): 1031-1042. https://doi.org/10.18280/isi.290321