Breast Lesion Classification From Digital Breast Tomosynthesis Images Using Convolutional Neural Networks

Jose Zarate
Department of Biomedical Engineering
Duke University
jhz7@duke.edu

Kevin Xu
Department of Biomedical Engineering
Department of Computer Science
Duke University
kx33@duke.edu

Joshua Tennyson
Department of Biomedical Engineering
Department of Computer Science
Duke University
joshua.tennyson@duke.edu

Abstract

Breast cancer is a leading cause of cancer-related deaths among women worldwide. Early detection through digital breast tomosynthesis (DBT) screening, a 3D mammography technology, is crucial for improving patient outcomes. However, interpreting DBT images can be challenging, even for experienced radiologists. To aid in accurate breast lesion classification, we propose a machine learning approach that explores the impact of various image pre-processing techniques, such as low resolution and low contrast, on the performance of three CNN-based models. The implementation of these techniques serve as physical layers and simulate the effects of using less advanced tomosynthesis equipment by reducing X-ray detectors or lowering X-ray dose. By evaluating our models under these conditions, we assess their robustness and potential for use in resource-limited settings. Our results demonstrate the effectiveness of the proposed approaches in classifying breast lesions, even with image quality challenges. This work contributes to the development of computer-aided diagnosis (CAD) systems that assist radiologists in making accurate and timely diagnoses, ultimately improving patient care.

1 Introduction

Breast cancer is the most commonly diagnosed cancer and a leading cause of cancer-related deaths among women worldwide [1]. Early detection through mammography screening is a critical step in improving patient outcomes and reducing mortality rates; however, the interpretation of mammograms can be challenging, even for experienced radiologists, due to the complexity of breast tissue and the subtle appearance of some lesions [2]. To aid in the accurate interpretation of mammograms, computer-aided diagnosis (CAD) systems have been developed [3]. These systems assist radiologists in identifying and characterizing suspicious regions in mammograms, reducing the risk of misinterpretation and improving detection rates.

Traditional CAD systems often rely on hand-crafted features, such as gray-level co-occurrence matrices (GLCM), Gabor filters, wavelets, scale-invariant feature transform (SIFT), and local binary patterns (LBP); however, these features are manually defined by experts and may not be optimal for all imaging tasks [4]. Moreover, extracting relevant features from raw images can be challenging, highlighting the need for automatic feature extraction techniques. Machine learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful tool for automatic feature extraction and image classification. CNNs have consistently outperformed other methods in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[5]. In the context of breast cancer detection, CNNs have been used as a baseline approach.

To address the limitations of traditional CAD systems, we employ an approach that explores the impact of physical layers modeled by various image pre-processing techniques, such as low resolution and low contrast, on the performance of three different CNN-based models: a CNN inspired by Wichakam *et al.* [6], one custom CNN, and an AlexNet model inspired by El-Shazli *et al.* [7]. These pre-processing techniques are designed to simulate the effects of using less advanced tomosynthesis equipment. Low resolution simulates the use of a machine with fewer X-ray detectors, while low contrast mimics the scenario where a lower X-ray dose is administered, resulting in reduced X-ray beam intensity. By evaluating our models under these conditions, we aim to assess their robustness and potential for use in resource-limited settings.

2 Related Work

The importance of CAD systems in radiology is evident from the increasing number of related research papers presented at major conferences, such as the Annual Meeting of the Radiological Society of North America (RSNA). The number of papers about CAD has increased by approximately 50% per year at RSNA. This can be attributed to it being shown that CAD systems can significantly help radiologists to detect suspicious regions in mammograms, resulting in a better detection rate [8].

Despite the success of traditional CNNs, their prediction accuracy in breast cancer detection remains limited [9]. To improve CNN performance, in a recent study, Du *et al.* proposed a novel method called SIFT-DBT (Self-supervised Initialization and Fine-Tuning for identifying abnormal Digital Breast Tomosynthesis images) to address the data imbalance challenges in DBT imaging [10]. Their framework consists of two main components: a contrastive learning paradigm for learning robust features and a local multi-patch fine-tuning strategy for maintaining image resolution and reducing computational costs. The contrastive learning approach utilizes DBT metadata to build positive pairs, encouraging the model to learn structural and semantic similarities between slices from the same volume and different views of the same breast.

The study experiments with two backbone models, a conventional 7-layer CNN and ResNet50, and compares the proposed method with several baselines, including models trained from scratch and pre-trained with classical contrastive learning methods. The SIFT-DBT method demonstrates promising results in improving abnormal DBT image detection, highlighting the potential of combining contrastive learning and multi-patch fine-tuning techniques with backbone architectures like ResNet50. However, the authors emphasize that further research is needed to optimize these techniques for real-world clinical applications.

3 Methods

3.1 Dataset

DBT images were obtained from the Cancer Imaging Archive [11]. Raw files were in the form of 3D DICOM images, and were split into training, validation, and testing sets. In addition, ground truth classifications were assigned to each image (normal, actionable, benign, cancerous). For images classified as benign or cancerous, the bounding box enclosing the lesion was given as well as the specific slice it was located in. Since bounding boxes were not provided for the normal and actionable classes, they were aggregated as one class for our models. The number of images used for each category are summarized in Table 1.

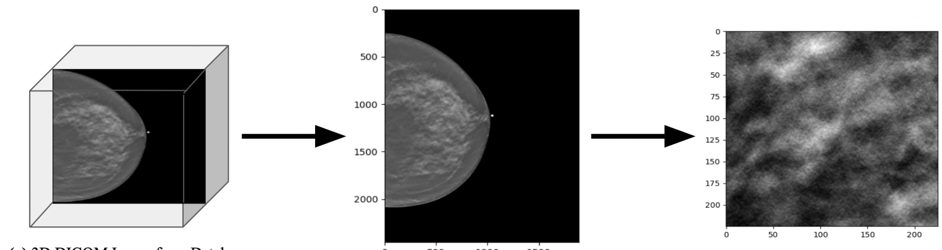
Table 1. Number of Images Used

Image set	Classification	Count
Training	Normal/Actionable	1491
	Benign	115
	Cancerous	74
Validation	Normal/Actionable	299
	Benign	38
	Cancerous	34
Testing	Normal/Actionable	200
	Benign	59
	Cancerous	59

Nearly every benign and cancerous image in the database were used during training, validation, and testing. To retain class balance, only a subset of the normal and actionable images were used.

3.2 Pre-processing

3D DICOM images were uncompressed into 3D numpy arrays. For the benign and cancerous images, the bounding box within a particular slice containing the lesion was extracted and resized to 225 x 225 pixels. For the normal and actionable images, the middle slice was selected and an arbitrary 225 x 225 pixel region containing tissue was extracted.

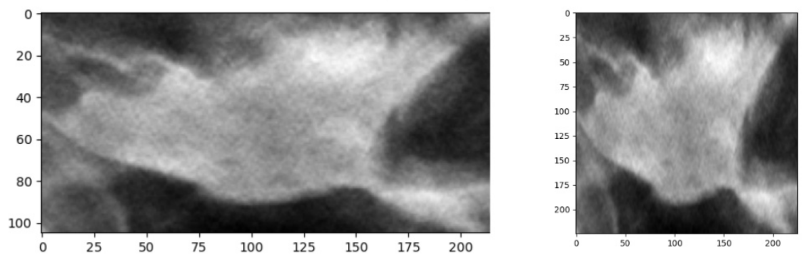


(a) 3D DICOM Image from Database

(b) 2D Slice of Interest

(c) 225px x 225px Region of interest

Figure 1. Data Extraction Pipeline



(a) Given Bounding Box

(b) Rescaled Bounding Box

Figure 2. Bounding Box Adjustment

3.3 Physical layers

Physical layers including low resolution, low contrast, and low resolution + low contrast combined were applied to the images to form 3 additional image sets to simulate less X-ray detector elements and reduced X-ray intensity. Low resolution was achieved by convolving each image with a 4x4 kernel of ones. Low contrast was achieved by applying the sigmoid function $\frac{1}{1+e^{-3(x-0.5)}}$ to each image, where x is the raw pixel intensity. To combine both physical layers, the low contrast sigmoid function was applied to the convolved images.

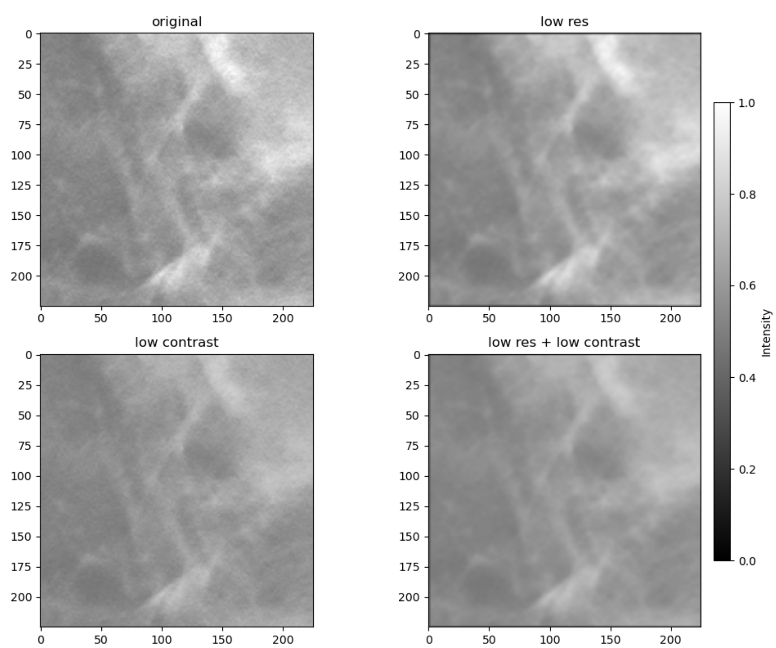


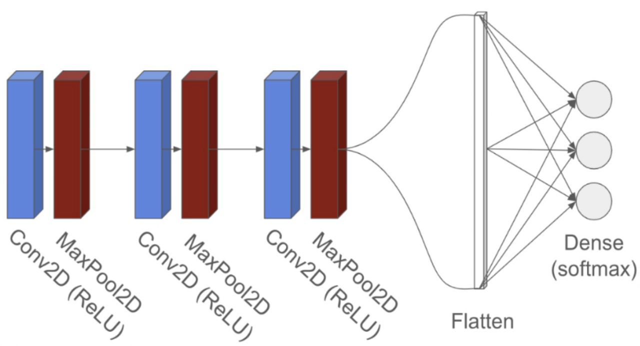
Figure 3. Physical Layers

3.4 Models

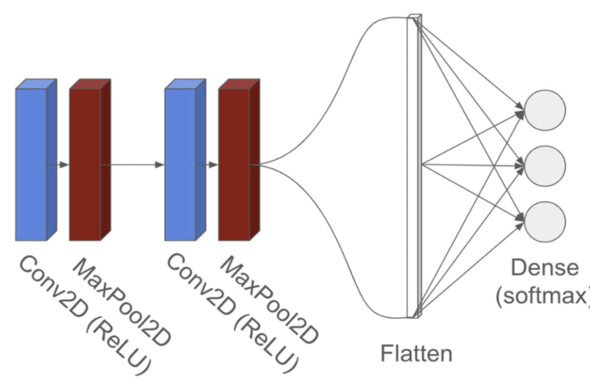
In this study, three distinct CNN-based model architectures were explored to address the limitations of traditional CAD systems and assess their performance under various image pre-processing techniques, which serve as physical layers simulating less advanced tomosynthesis equipment. The first model, inspired by Wichakam *et al.* [6], is referred to as the “original CNN.” This architecture consists of a stack of three Conv2D and MaxPool2D layers, followed by a flattening operation and a fully connected dense layer corresponding to the three classes. The original CNN serves as a baseline model, capturing hierarchical features from the input images.

To investigate the potential impact of model complexity on performance, we introduced a second model called the “Simple CNN.” This model is derived from the original CNN by removing one stack of Conv2D and MaxPool2D layers. The motivation behind this modification is to reduce the risk of overfitting, which can occur when a model becomes too complex relative to the size and diversity of the training dataset. By simplifying the architecture, the Simple CNN aims to strike a balance between capturing relevant features and generalizing well to unseen data.

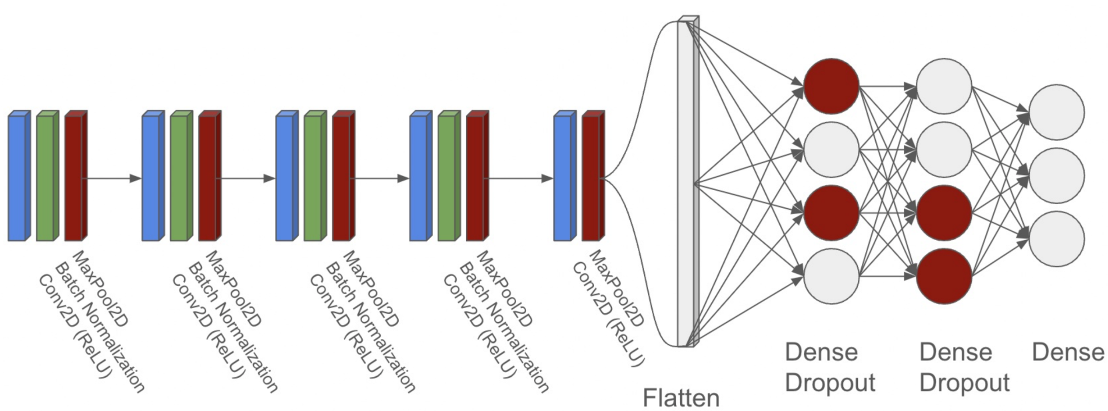
The third model, inspired by El-Shazli *et al.* [7], is based on the AlexNet architecture. AlexNet is a deeper and more sophisticated CNN compared to the original and Simple CNNs. It incorporates additional components such as batch normalization layers between the Conv2D and MaxPool2D layers in each stack. Batch normalization helps to stabilize the learning process and accelerate convergence by normalizing the activations of each layer. Furthermore, AlexNet includes two more dense layers and dropout layers before the final dense layer. The dropout layers introduce regularization by randomly setting a fraction of the input units to zero during training, which helps to prevent overfitting. The increased depth and complexity of AlexNet are intended to capture more intricate patterns in the images and potentially achieve higher accuracy in detecting cancerous lesions.



(a) Model 1 - Original CNN



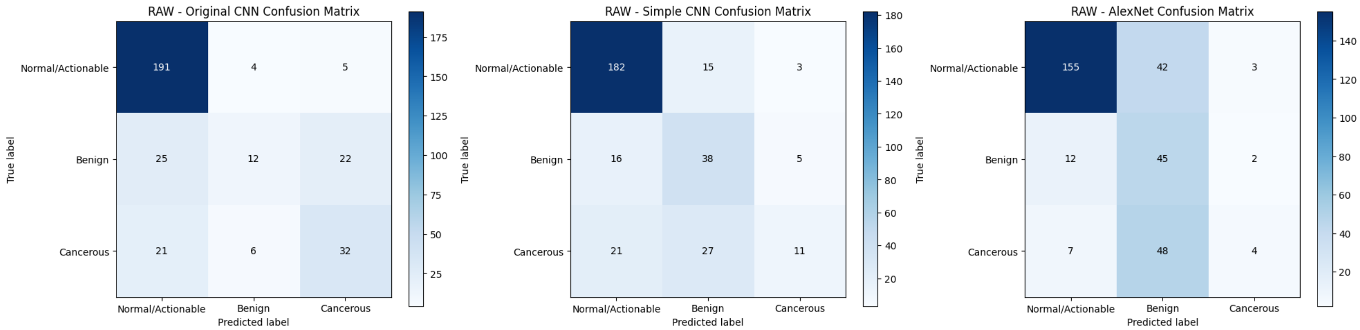
(b) Model 2 - Simple CNN



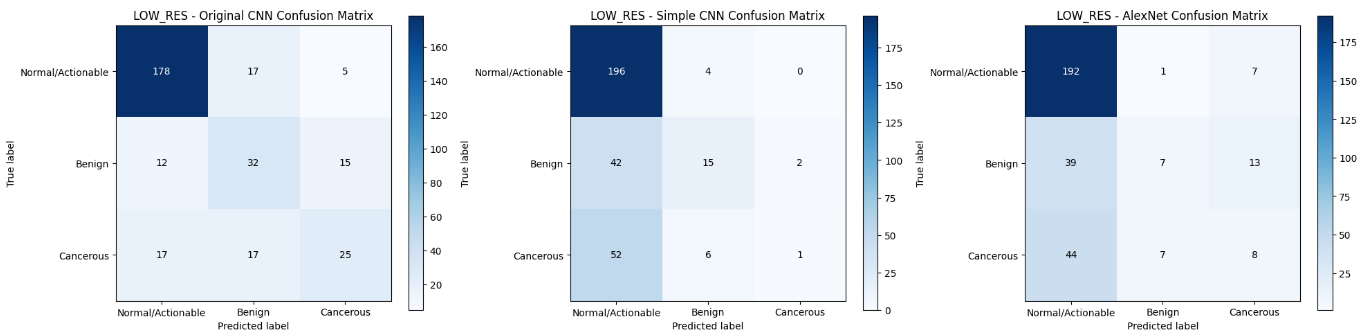
(c) Model 3 - AlexNet

Figure 4. Model Architectures

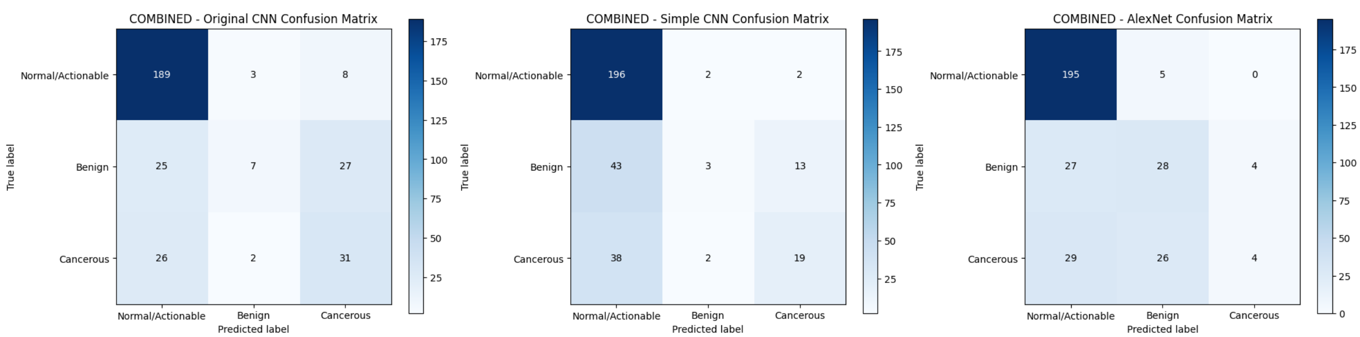
4 Results



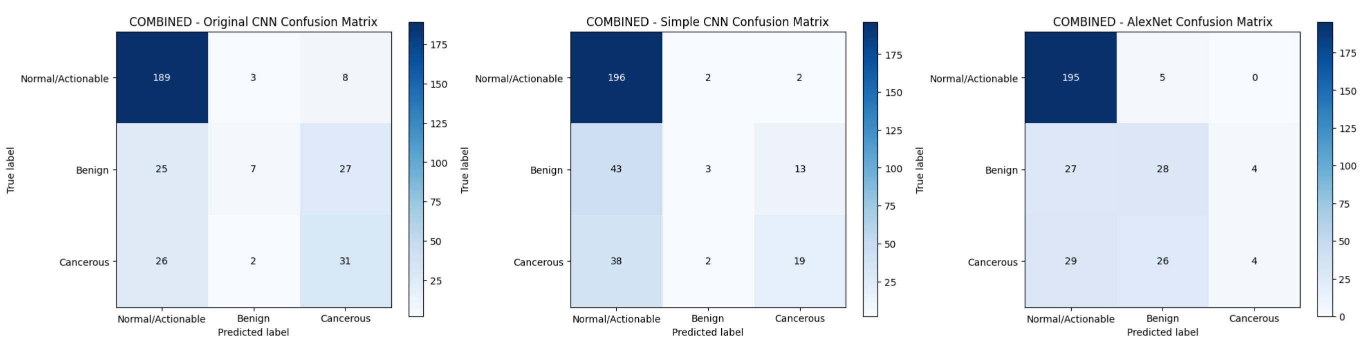
(a) Raw Images



(b) Low Resolution Images



(c) Low Contrast Images



(d) Combined Images

Figure 5. Confusion Matrices

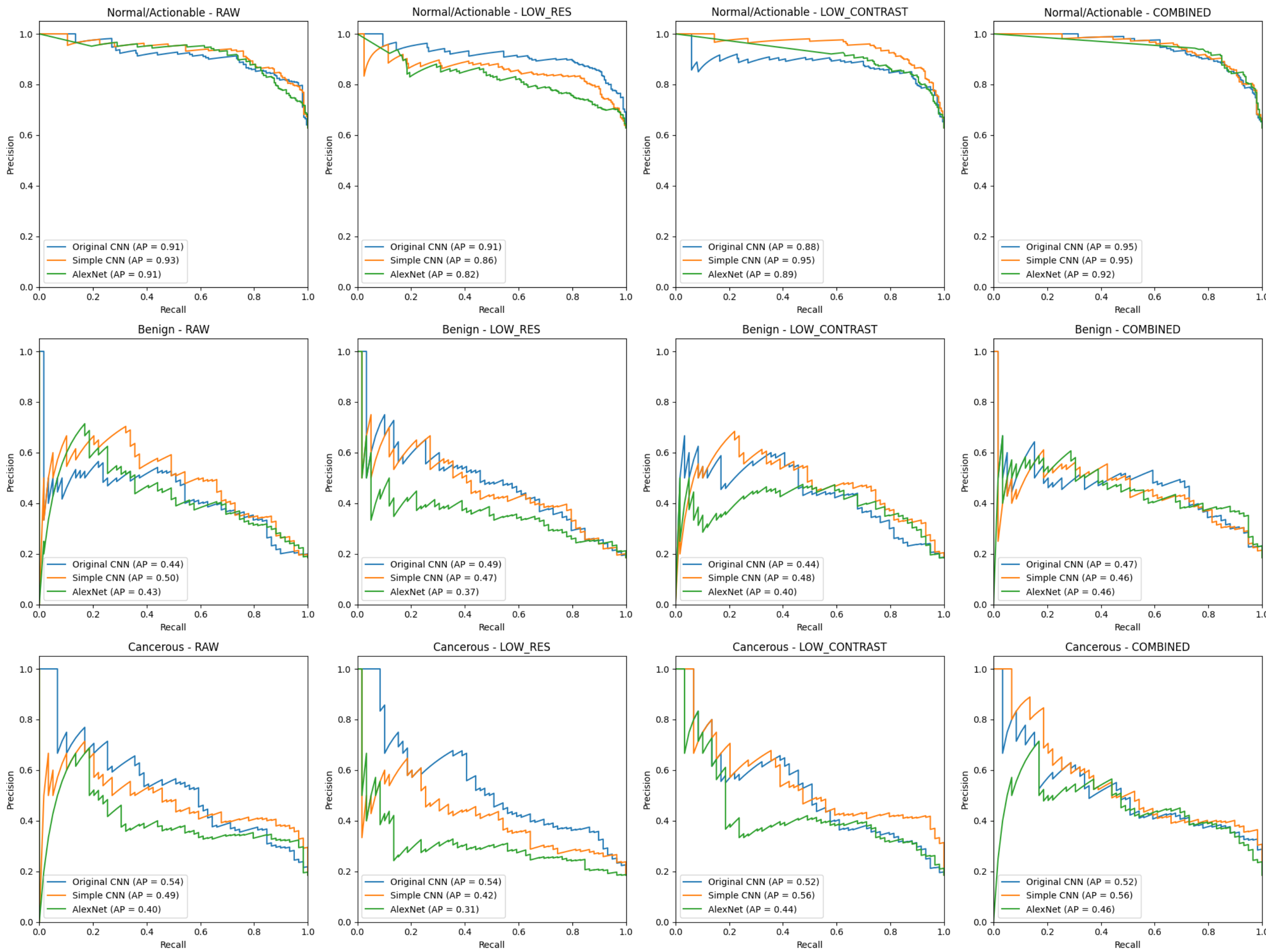


Figure 6. Precision-Recall Curves

The following model performance comparisons are measured by their average precision.

4.1 Normal/Actionable class

For the low resolution images, the original CNN performed the best. For low contrast images, the simple CNN performed the best. For raw images and combined images, all 3 models performed similarly.

4.2 Benign class

For the raw images, the simple CNN performed the best. For the low resolution and low contrast images, AlexNet performed relatively poorly. AlexNet also performed the worst for the combined images but to a lesser degree.

4.3 Cancerous class

For the raw and low resolution images, the original CNN performed the best. For the low contrast and combined images, the original CNN and simple CNN had similar performance but AlexNet performed poorly.

4.4 Summary

For most class/physical layer combinations, the original CNN performed the best or was comparable to the highest performing model, especially for the cancerous images. AlexNet had the most trouble matching performance with the the other 2 models.

5 Discussion

The original CNN model demonstrated the highest performance across the majority of class and physical layer combinations, making it the most suitable choice for this specific image set. The model's complexity allowed it to capture deeper patterns within the images, while its robustness ensured that it was not significantly affected by the diminishing image quality resulting from the application of physical layers. This highlights the importance of striking a balance between model complexity and robustness when dealing with varying image quality conditions.

In contrast, the AlexNet model exhibited poor performance, likely due to overfitting the training data. The limited number of training images available for the benign and cancerous classes may have hindered AlexNet's ability to leverage its more intricate architecture effectively. This emphasizes the significance of having a sufficient and balanced dataset to train complex models like AlexNet, especially when dealing with medical images where certain classes may have limited representation. The simple CNN model, on the other hand, showed promising performance, but mainly for the low contrast images in the normal/actionable class and the raw images in the benign class. The reduced complexity of the simple CNN architecture may have helped prevent overfitting and improved its generalization capabilities in these specific scenarios.

To further enhance the performance and reliability of the models, additional research is necessary to explore various architectures and hyperparameter tuning. Moreover, techniques like transfer learning, data augmentation, and regularization could be employed to mitigate overfitting and improve the models' generalization capabilities.

In its current state, the original CNN model serves as the best candidate for integration into a CAD system. However, before deploying the model in a clinical setting, extensive validation and testing on larger, diverse datasets are crucial to ensure its robustness and reliability across different patient populations and imaging conditions. Furthermore, the model's performance should be evaluated in collaboration with experienced radiologists to assess its potential for improving diagnostic accuracy and efficiency in real-world scenarios.

To facilitate the adoption of the original CNN model as a CAD tool, future research should also focus on developing user-friendly interfaces and visualization techniques that enable radiologists to interpret the model's predictions easily.

In conclusion, while the original CNN model demonstrates promising performance as a potential CAD tool for this specific image set, further research is necessary to explore alternative architectures, validate the model's performance on larger datasets, and develop user-friendly interfaces to facilitate its integration into clinical workflows. By addressing these aspects, the model's potential to assist radiologists in accurately diagnosing breast lesions and improving patient outcomes can be fully realized.

References

[1] I. Moreira, I. B. C. Amaral, I. Domingues, A. N. L. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast," Academic radiology, Feb. 01, 2012. https://pubmed.ncbi.nlm.nih.gov/22078258/

[2] E. Ekpo, M. Alakhras, and P. C. Brennan, "Errors in Mammography Cannot be Solved Through Technology Alone," PubMed, Feb. 26, 2018. https://pubmed.ncbi.nlm.nih.gov/29479948/

[3] R. K. Samala, H. P. Chan, L. M. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," Medical physics on CD-ROM/Medical physics, Nov. 29, 2016. https://pubmed.ncbi.nlm.nih.gov/27908154/

[4] J. Torrents-Barrena, D. Puig, J. Melendez, and A. Valls, "Computer-aided diagnosis of breast cancer via Gabor wavelet bank and binary-class SVM in mammographic images," Journal of experimental and theoretical artificial intelligence, Apr. 15, 2015. https://ui.adsabs.harvard.edu/abs/2016JETAI..28..295T/abstract

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, May 24, 2017. https://dl.acm.org/doi/10.1145/3065386

[6] I. Wichakam and P. Vateekul, "Combining deep convolutional networks and SVMs for mass detection on digital mammograms," IEEE Conference Publication | IEEE Xplore, Feb. 01, 2016. https://ieeexplore.ieee.org/document/7440527

[7] A. M. A. El-Shazli, S. M. Youssef, and A.-H. Soliman, "Intelligent Computer-Aided Model for Efficient Diagnosis of Digital Breast Tomosynthesis 3D Imaging Using Deep Learning," Applied sciences, Jun. 05, 2022. https://www.mdpi.com/2076-3417/12/11/5736

[8] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," Computerized medical imaging and graphics, Jun. 01, 2007. https://pubmed.ncbi.nlm.nih.gov/17349778/

[9] Y. Tang, "Deep Learning using Linear Support Vector Machines," arXiv.org, Jun. 02, 2013. https://arxiv.org/abs/1306.0239

[10] Y. Du, R. J. Hooley, J. Lewin, and N. C. Dvornek, "SIFT-DBT: Self-supervised Initialization and Fine-Tuning for Imbalanced Digital Breast Tomosynthesis Image Classification," arXiv.org, Mar. 19, 2024. https://arxiv.org/abs/2403.13148

[11] Buda, M., Saha, A., Walsh, R., Ghate, S., Li, N., Swiecicki, A., Lo, J. Y., Yang, J., Mazurowski, M. (2020). Breast Cancer Screening – Digital Breast Tomosynthesis (BCS-DBT) (Version 5) [dataset]. The Cancer Imaging Archive. https://doi.org/10.7937/E4WT-CD02