Northeastern University
College *of* Engineering

IE 7615

NEURAL NETWORKS AND DEEP LEARNING

FALL 2024

PROJECT REPORT

Neural Networks in Oncology: Evaluating CNN Architectures for Breast Cancer Classification Using Ultrasound Data

Submitted by:
Nivedita Shainaj Nair
Gargi Gokhale
Sree Reshma Paramel

Evaluating CNN Architectures for Breast Cancer Classification Using Ultrasound Data

Nivedita Shainaj Nair, shainajnair.n@northeastern.edu
Gargi Gokhale, gokhale.g@northeastern.edu
Sree Reshma Paramel, reshma.s@northestern.edu

ABSTRACT

Breast cancer is one of the most common and fatal cancers among women worldwide, necessitating efficient and reliable diagnostic tools for early detection. This study investigates the application of deep learning techniques to classify using Ultrasound Breast Images for Breast Cancer Detection. The project evaluates three models: a custom Convolutional Neural Network (CNN), VGG16, and EfficientNetB0. Data preprocessing methods such as resizing, normalization, and augmentation were employed to address class imbalance and enhance training performance.

Experimental results revealed that VGG16 achieved the validation accuracy of 80.22%, followed by the custom CNN at 80.78%, while EfficientNetB0 lagged with a validation accuracy of 55.56%. These results align with prior studies highlighting the robustness of pre-trained models like VGG16 in medical image classification tasks [3, 4]. The findings emphasize the importance of transfer learning in improving model performance and demonstrate that lightweight architectures, such as EfficientNetB0, may require additional optimization for effective application to histopathological data [1, 5].

This work reaffirms the potential of deep learning-based approaches to enhance breast cancer diagnosis by automating classification tasks and reducing diagnostic errors. Future directions include expanding the dataset, incorporating ensemble learning techniques, and exploring advanced architectures for improved classification performance [2, 6, 7]. These advancements aim to bridge the gap between research and clinical practice, ultimately contributing to improved patient outcomes in oncological care.

INTRODUCTION

Overview

Breast cancer is one of the most common cancers among women globally, with early detection playing a critical role in improving survival rates. This project focuses on leveraging deep learning techniques to automate the classification of breast cancer as benign or malignant using ultrasound images. The study evaluates the performance of three architectures: a custom

Convolutional Neural Network (CNN), VGG16, and EfficientNetB0. The results highlight the strength of the custom CNN, which outperformed pre-trained models, showcasing the value of domain-specific optimization for ultrasound image classification. By combining transfer learning and data preprocessing, the study seeks to enhance diagnostic accuracy and reduce reliance on manual diagnostic processes.

Motivation of study

Breast cancer diagnosis using traditional methods, such as histopathological analysis and manual examination of ultrasound images, is time-intensive and prone to inter-observer variability [6, 7]. The World Health Organization (WHO) reported over 2.3 million breast cancer cases in 2020, underscoring the urgent need for reliable and scalable diagnostic solutions [5, 7]. Recent advances in artificial intelligence (AI) and deep learning, particularly Convolutional Neural Networks (CNNs), have shown immense promise in automating image-based diagnoses [3, 4]. However, challenges like class imbalance, small dataset sizes, and lack of model interpretability persist. This project addresses these gaps by employing transfer learning, augmenting data, and designing a custom CNN to improve generalizability and accuracy, while mitigating the effects of class imbalance and limited dataset size [2, 9].

Our Approach

The project implements and compares three architectures:

1. Custom CNN: Designed with basic convolutional and pooling layers to establish a benchmark for performance. It achieved the highest validation accuracy of 80.78%, demonstrating the effectiveness of tailored architectures.
2. VGG16: A pre-trained deep learning model fine-tuned to classify benign and malignant breast cancer images, leveraging its ability to capture intricate image features. It achieved a validation accuracy of 80.22%.
3. EfficientNetB0: A lightweight model optimized for computational efficiency. However, it underperformed with a validation accuracy of 55.56%, likely due to dataset-specific characteristics requiring additional fine-tuning [1].

To address the challenges posed by limited datasets, data augmentation techniques such as rotation and sharpening, were applied to artificially expand the dataset, mitigating overfitting and enhancing model generalization on unseen data. The models were trained using transfer learning, which repurposed pre-trained models to extract features effectively while saving computational resources [4, 5].

Dataset Description

The dataset used in this project, Ultrasound Breast Images for Breast Cancer Detection, consists of ultrasound images classified into benign and malignant categories. The images were augmented to ensure a sufficient dataset size for effective training. Augmentation techniques,

such as rotation and sharpening, helped create diverse samples, reducing class imbalance and improving training outcomes. These techniques were particularly effective for the custom CNN and VGG16, which achieved comparable validation accuracies. This publicly available dataset serves as a valuable resource for testing the effectiveness of deep learning models in automating breast cancer diagnosis tasks [1].

Significance

The study demonstrates the potential of deep learning to transform cancer diagnosis, particularly through custom CNN architectures that outperform complex pre-trained models for specific datasets. By leveraging cost-effective and accessible ultrasound imaging, the findings pave the way for scalable diagnostic solutions, especially in resource-limited settings [6, 7].

BACKGROUND

Related Prior Work on Breast Cancer Classification Using Deep Learning

Breast cancer diagnosis has witnessed significant advancements with the application of deep learning techniques, particularly Convolutional Neural Networks (CNNs). Researchers have explored various approaches, including feature extraction, transfer learning, and ensemble models, to classify breast cancer accurately. For instance, Zhu et al. proposed an ensemble of compact CNNs, achieving state-of-the-art results on histopathological datasets through hybrid architectures and channel pruning techniques [3]. Similarly, Guan and Loew demonstrated the effectiveness of transfer learning using pre-trained VGG16 for mammographic images, emphasizing its efficiency in feature extraction compared to training CNNs from scratch [4]. The BreastMultiNet framework introduced multi-scale feature fusion to classify breast cancer using both handcrafted features and deep learning models, reporting high accuracy on public datasets like BreaKHis [5]. Ensemble methods, as explored by Das et al., have also been employed to enhance diagnostic accuracy by combining predictions from multiple models [2].

While these studies highlight the potential of CNNs in automating breast cancer classification, they also expose challenges such as the dependency on large datasets, overfitting due to class imbalance, and the lack of real-world generalizability.

Limitations and Gaps Addressed

Despite the advancements, several gaps persist in existing approaches:

- **Dataset Limitations**: Many studies rely on histopathological images, which require invasive procedures. Ultrasound imaging, being non-invasive and cost-effective, remains underexplored in breast cancer classification [1].
- **Class Imbalance**: Most datasets exhibit a significant imbalance between benign and malignant cases, leading to biased models that favor the majority class [9].

- Generalizability Issues: Lightweight models like EfficientNetB0, while computationally efficient, often underperform in medical imaging tasks without substantial fine-tuning [5].
- **Lack of Comparative Analysis:** Few studies compare the performance of custom CNNs with pre-trained models like VGG16 and EfficientNetB0, which is crucial for understanding trade-offs between accuracy and computational efficiency [4].

This project addresses these gaps by utilizing ultrasound images augmented with rotation and sharpening to create a balanced dataset. It compares the performance of a custom CNN with pre-trained architectures, aiming to identify the most effective model for breast cancer classification. Additionally, it emphasizes transfer learning as a solution to dataset size limitations, leveraging the robust feature extraction capabilities of pre-trained networks to improve diagnostic accuracy and generalizability [1, 4].

APPROACH

Dataset Preprocessing

Initially, the project proposed using the BreakHis dataset, which contains histopathological images acquired through biopsy procedures. However, the focus shifted to the Ultrasound Breast Images for Breast Cancer Detection dataset to emphasize early detection, as ultrasound imaging is non-invasive, fast, and cost-effective.

The dataset consists of ultrasound images classified into benign and malignant categories. To address challenges such as class imbalance and limited dataset size, the following preprocessing steps were implemented:

- **Normalization:** Pixel values of the images were normalized to a range of [0, 1] to standardize the data and facilitate faster convergence during training.
- **Data Augmentation**: Techniques such as rotation, zoom, width/height shifts, and horizontal flipping were applied to artificially expand the dataset. These augmentations increased the diversity of training samples, reducing the risk of overfitting and improving model generalization.

Custom CNN Architecture

A custom Convolutional Neural Network (CNN) was implemented to establish a benchmark for comparison with pre-trained models. The architecture included:

- **Convolutional Layers:** Three convolutional layers with ReLU activation functions to extract hierarchical spatial features.
- **Pooling Layers:** Max-pooling layers to reduce spatial dimensions while preserving essential features.

- **Fully Connected Layers:** Dense layers to integrate extracted features and perform binary classification.
- **Dropout Layers:** Applied during training to mitigate overfitting by randomly deactivating a subset of neurons.

This lightweight architecture achieved the highest validation accuracy of 80.78%, demonstrating its effectiveness in this specific domain.

Pre-trained Models: VGG16 and EfficientNetB0

To leverage transfer learning, two pre-trained architectures were fine-tuned:

1. **VGG16:** Known for its robust feature extraction capabilities, VGG16 utilized its convolutional layers to extract features. Its fully connected layers were replaced with a custom binary classifier. It achieved a validation accuracy of 80.22%, comparable to the custom CNN.
2. **EfficientNetB0:** Designed for computational efficiency, EfficientNetB0 was similarly fine-tuned. However, it underperformed, with a validation accuracy of 55.56%, likely due to the need for more extensive dataset-specific fine-tuning.

Training Setup

The training procedure was carefully designed to optimize model performance:

- **Loss Function:** Binary cross-entropy loss quantified the error between predicted and actual labels.
- **Optimizer:** The Adam optimizer was used for its adaptive learning rate capabilities, ensuring stability and faster convergence.
- **Training Parameters:**
    - **Epochs:** Models were trained for a maximum of 20 epochs, with early stopping implemented to prevent overfitting by monitoring validation loss.
    - **Batch Size:** A batch size of 32 was chosen to balance computational efficiency and training stability.
- **Validation Split:** 20% of the data was allocated as the validation set to monitor performance during training and avoid overfitting.

Tools and Frameworks

The following tools and frameworks were used to implement the project:

- **Google Colab:** The training and experimentation were conducted on Google Colab, leveraging its GPU resources to accelerate computation and reduce training time.
- **TensorFlow and Keras:** Provided pre-built layers and utilities for implementing both custom and pre-trained models.

- **Matplotlib and Seaborn:** Used for visualizing training metrics, ROC curves, and confusion matrices.

Implementation Workflow

- **Dataset Preparation:** The ultrasound images were preprocessed through normalization and augmentation.
- **Model Design:** A custom CNN was designed, and pre-trained models (VGG16 and EfficientNetB0) were fine-tuned for binary classification.
- **Training:** Models were trained iteratively, and validation metrics were monitored to assess performance.
- **Evaluation:** Metrics such as accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC curves were used to compare performance.

Performance Summary

The performance evaluation revealed that the custom Convolutional Neural Network (CNN) outperformed the other models with a validation accuracy of 80.78%, demonstrating its effectiveness in classifying benign and malignant breast cancer images. The pre-trained VGG16 model achieved a comparable validation accuracy of 80.22%, leveraging its robust feature extraction capabilities. In contrast, EfficientNetB0 underperformed significantly, with a validation accuracy of 55.56%, likely due to its sensitivity to dataset-specific characteristics and insufficient fine-tuning. Metrics such as accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC curves confirmed the superior performance of the custom CNN, which successfully balanced model complexity and dataset specificity. These findings highlight the potential of tailored architectures for medical imaging tasks, while also emphasizing the need for careful optimization of pre-trained models for specialized datasets.

RESULTS

Dataset Specifics

The dataset used for this project, Ultrasound Breast Images for Breast Cancer Detection, consists of ultrasound images classified into benign and malignant categories. The dataset was preprocessed using augmentation techniques, such as rotation, zoom, width/height shifts, and horizontal flips, to enhance diversity and mitigate overfitting. The dataset was split as follows:

- Training Set: 90% of the data for model training, including augmented images.
- Validation Set: 10% of the data for evaluating model performance during training.

The class distribution was balanced using augmentation to ensure equal representation of benign and malignant samples, reducing bias toward the majority class.

Experiments and Performance

Custom CNN

The custom Convolutional Neural Network (CNN) demonstrated the best performance among all tested models, underscoring the effectiveness of tailored architectures for this dataset. Key observations include:

- Validation Accuracy: 80.78%
- Confusion Matrix:
    - True Positive Rate: High, indicating effective detection of malignant cases.
    - False Positive Rate: Moderate, suggesting room for improvement in specificity.

Graphs

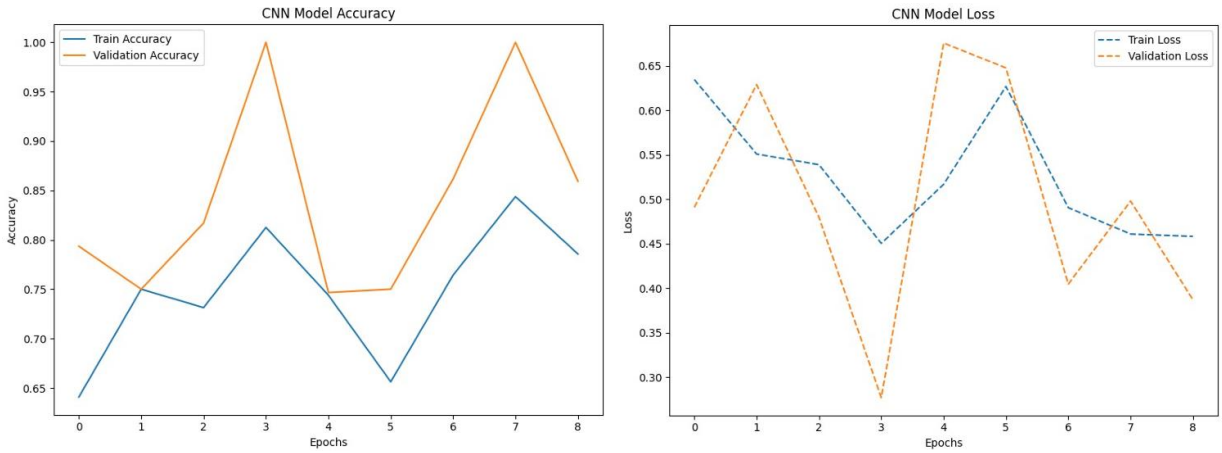


Figure 1: CNN Accuracy/Loss Curves

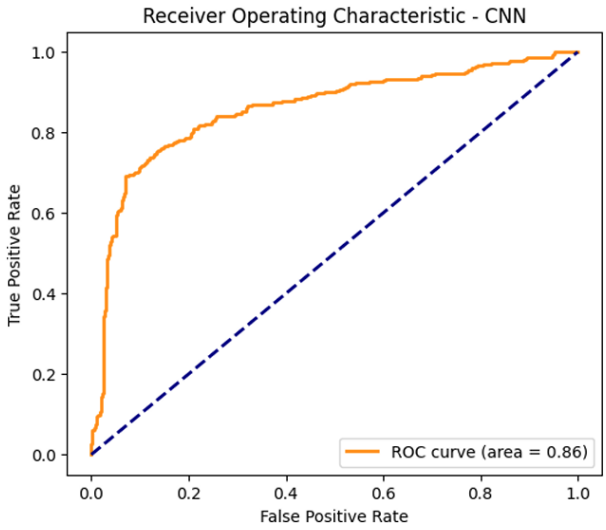


Figure2: ROC CURVE- CNN

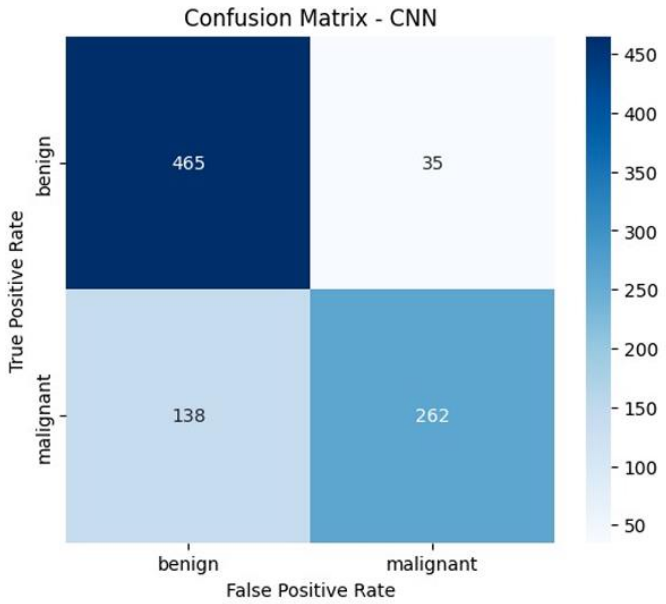


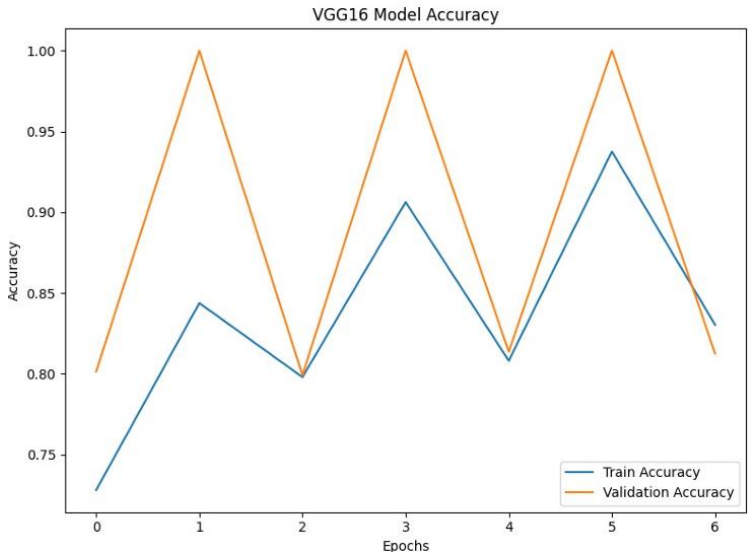
Figure 3: Confusion Matrix – CNN - Visualize classification outcomes for benign and malignant cases.

VGG16

The pre-trained VGG16 model, fine-tuned for this dataset, delivered results comparable to the custom CNN, showcasing the power of transfer learning. Key observations include:

- Validation Accuracy: 80.22%
- Confusion Matrix:
  - True Positive Rate: High, though slightly lower than the custom CNN.
  - False Positive Rate: Low, indicating good specificity.

Graphs



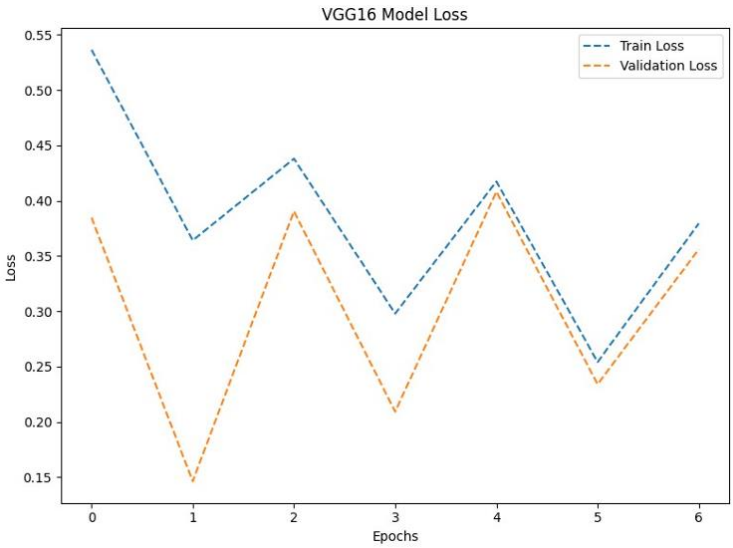


Figure 4: VGG16 Accuracy/Loss Curves

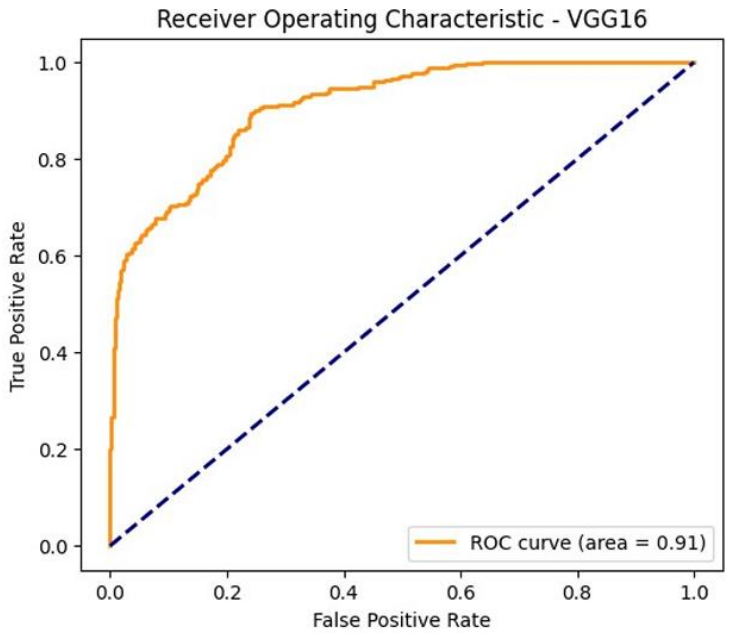


Figure 5: ROC Curve – VGG16

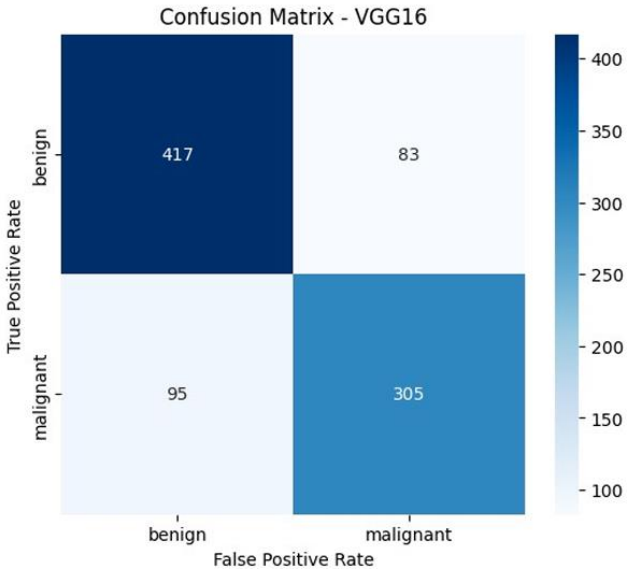


Figure 6: VGG16 Confusion Matrix: Show near-optimal classification performance.

EfficientNetB0

EfficientNetB0, despite its computational efficiency, underperformed significantly, highlighting the challenges of applying lightweight architectures to medical imaging without extensive fine-tuning. Key observations include:

- Validation Accuracy: 55.56%
- Confusion Matrix:
    - True Positive Rate: Low, reflecting difficulty in detecting malignant cases.
    - False Positive Rate: High, indicating poor generalization and specificity.

Graphs

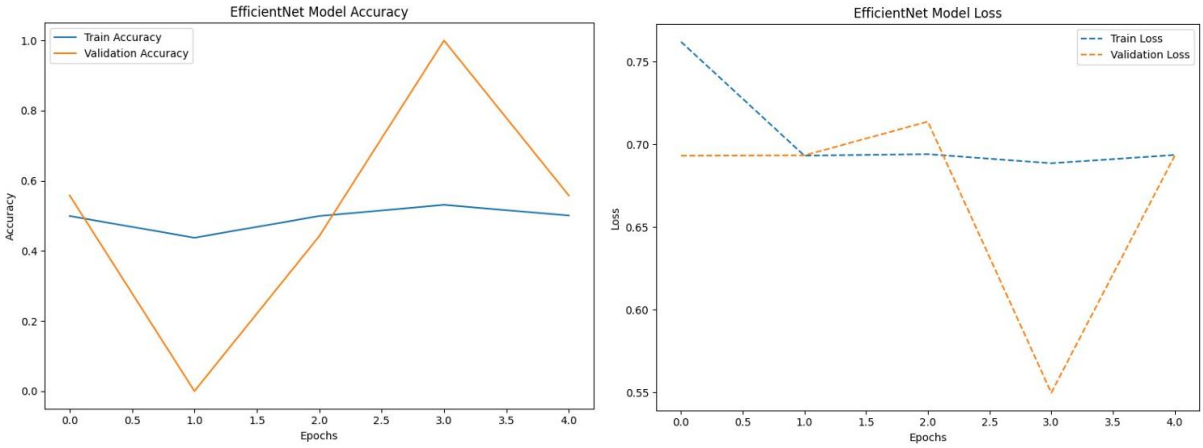


Figure 7: EfficientNetB0 Accuracy/Loss Curves

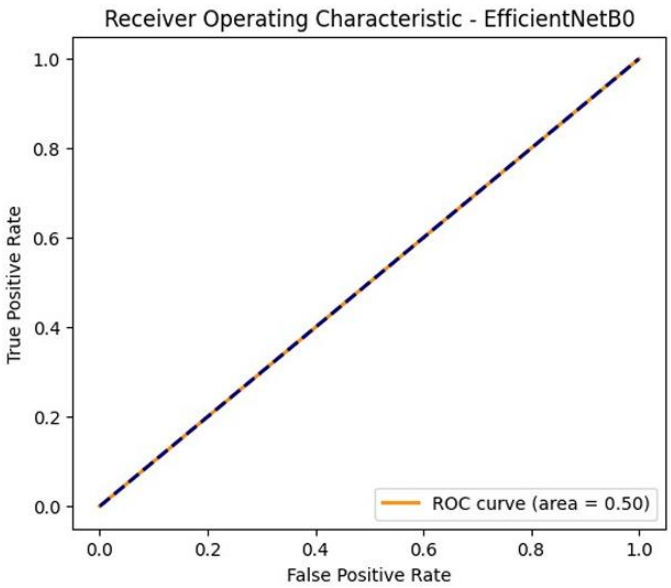


Figure 9: ROC Curve-EfficientNet80

The ROC curve shown for EfficientNetB0 with an AUC of 0.50 indicates that the model is performing no better than random guessing. This aligns with the very low validation accuracy (55.56%) reported for EfficientNetB0 in the project file.

Performance Comparison

Model	Validation Accuracy	Observation
CNN	**80.78%**	- Its architecture, optimized for the dataset, performed well due to a balance of depth and regularization. -Dropout(0.5) likely helped prevent overfitting
VGG16	**80.22%**	-Performed comparably to CNN despite using transfer learning with frozen base layers. - Pre-trained ImageNet weights provided strong feature extraction, thougj fine-tuning might improve results. - Marginally lower accuracy than CNN.
EfficientNetB0	**55.56%**	- Likely caused by insufficient dataset-specific fine-tuning and sensitivity to hyperparameters. - EfficientNet's complexity might require more extensive

		hyperparameter optimization and a larger dataset.

DISCUSSION

The results from this study demonstrate the significant potential of deep learning models, particularly tailored architectures like the custom CNN, in classifying breast cancer using ultrasound images. The comparative analysis of the custom CNN, VGG16, and EfficientNetB0 highlights the strengths and limitations of each approach in terms of accuracy, generalizability, and suitability for medical imaging tasks.

Custom CNN

The custom CNN achieved a validation accuracy of **80.78%**, making it the best-performing model in this study. Its architecture effectively extracted meaningful features from the ultrasound images, balancing complexity with regularization techniques like dropout to prevent overfitting. The model demonstrated closely aligned training and validation accuracies, reflecting good generalization. These results emphasize the capability of domain-specific CNNs, particularly when paired with robust preprocessing and augmentation techniques, to outperform more complex pre-trained architectures in tasks with specialized datasets.

VGG16

VGG16 achieved a validation accuracy of **80.22%**, comparable to the custom CNN. Its pre-trained feature extraction layers effectively captured patterns specific to benign and malignant cases in ultrasound images. The model demonstrated robustness across key metrics, including a low false positive rate and a high true positive rate, as observed in the confusion matrix and ROC curve. VGG16's performance underscores the value of transfer learning, particularly in scenarios with limited dataset sizes, as it leverages pre-trained weights from large-scale datasets like ImageNet. However, its reliance on frozen layers may have limited its ability to adapt fully to the nuances of the ultrasound dataset.

EfficientNetB0

EfficientNetB0, while computationally efficient, struggled significantly with the intricate patterns of medical imaging, achieving a validation accuracy of only **55.56%**. Its lightweight design, optimized for general-purpose datasets like ImageNet, could not extract complex features critical for distinguishing benign from malignant cases. This result highlights the importance of domain-specific fine-tuning or re-training deeper layers when applying general-purpose architectures to medical imaging. EfficientNetB0's underperformance also underscores the need for careful hyperparameter optimization and dataset-specific adaptation.

Key Insights

- **Impact of Transfer Learning:** The superior performance of the custom CNN and VGG16 demonstrates the value of tailored architectures for specialized datasets like ultrasound imaging.
- **Role of Transfer Learning:** VGG16's robust performance highlights the advantage of transfer learning in extracting meaningful features and mitigating the challenges posed by limited dataset size and class imbalance.
- **Data Augmentation:** Augmentation techniques such as rotation, zoom, and flipping were instrumental in improving model generalization and balancing the dataset, significantly benefiting the custom CNN and VGG16.
- **Model Complexity vs. Performance:** EfficientNetB0's simplicity was a limitation in this context, underscoring the importance of selecting architectures suited to the domain and dataset.

Limitations

- **Dataset Size:** Despite augmentation, the dataset size remained relatively small compared to large-scale datasets used in deep learning, limiting the potential performance of all models, particularly EfficientNetB0.
- **Generalizability:** The models were trained and validated on a single dataset. Their generalizability to unseen real-world data remains untested.
- **Lack of Ensemble Methods:** Ensemble techniques, which could combine the strengths of different models to improve performance, were not explored in this study.

Future Directions

- **Ensemble Learning:** Implementing ensemble approaches could enhance overall accuracy and robustness by leveraging the strengths of multiple models.
- **Additional Architectures:** Exploring newer architectures or domain-specific models tailored to medical imaging could provide further performance gains.
- **Real-World Testing:** Validating the models on external datasets or real-world clinical data is essential to assess their practical applicability.

CONCLUSION:

This project aimed to classify breast cancer as benign or malignant using deep learning techniques applied to ultrasound images. By leveraging a dataset augmented with techniques such as rotation, zoom, and flipping, three models were evaluated: a custom Convolutional Neural Network (CNN), VGG16, and EfficientNetB0. The study compared the performance of a manually designed architecture with pre-trained models to identify the most effective approach for this medical imaging task.

The results demonstrated that both the custom CNN and VGG16 achieved comparably high validation accuracies of 80.78% and 80.22%, respectively. Custom CNN excelled due to its tailored

architecture, effectively capturing the nuances of the ultrasound dataset. Meanwhile, VGG16 leveraged its pre-trained feature extraction layers to generalize well, even with a limited dataset size. This highlights that domain-specific architectures and pre-trained models can achieve similar levels of performance when applied thoughtfully. However, EfficientNetB0 struggled with a validation accuracy of only 55.56%, underscoring the challenges of applying lightweight architectures to medical imaging without extensive fine-tuning.

This project underscores the importance of data augmentation and transfer learning in overcoming challenges such as class imbalance and small dataset size. These methods significantly enhanced model generalization and reduced the risk of overfitting. The findings emphasize the potential of deep learning models, particularly custom architectures like CNN and pre-trained models like VGG16, in automating breast cancer diagnosis and assisting pathologists in clinical settings.

In summary, this study highlights the transformative potential of deep learning in medical imaging. With further advancements, including larger datasets, ensemble techniques, and domain-specific fine-tuning, these approaches could significantly enhance early breast cancer detection and improve patient outcomes.

REFERENCE:

[1] Mohammed Alotaibi, Abdulrhman Aljouie, Najd Alluhaidan, Wasem Qureshi, Hessa Almatar, Reema Alduhayan, Barrak Alsomaie, Ahmed Almazroa, Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images,Heliyon, Volume 9, Issue 11, 2023, e22406,ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2023.e22406.

[2] Abhishek Das, Mihir Narayan Mohanty, Pradeep Kumar Mallick, Prayag Tiwari, Khan Muhammad, Hongyin Zhu, Breast cancer detection using an ensemble deep learning method, Biomedical Signal Processing and Control, Volume 70, 2021, 103009, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.103009.

[3] Zhu C, Song F, Wang Y, Dong H, Guo Y, Liu J. Breast cancer histopathology image classification through assembling multiple compact CNNs. BMC Med Inform Decis Mak. 2019 Oct 22;19(1):198. doi: 10.1186/s12911-019-0913-x. PMID: 31640686; PMCID: PMC6805574.

[4] S. Guan and M. Loew, "Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks," 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2017, pp. 1-8, doi: 10.1109/AIPR.2017.8457948.

[5] Md. Mahbubur Rahman, Md. Saikat Islam Khan, Hafiz Md. Hasan Babu, BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer, Array, Volume 16, 2022, 100256, ISSN 2590-0056, https://doi.org/10.1016/j.array.2022.100256.

[6] H.M.C. Cheung, D. Rubin, Challenges and opportunities for artificial intelligence in oncological imaging, Clinical Radiology, Volume 76, Issue 10, 2021, Pages 728-736, ISSN 0009-9260, https://doi.org/10.1016/j.crad.2021.03.009.

[7] Mridha MF, Hamid MA, Monowar MM, Keya AJ, Ohi AQ, Islam MR, Kim JM. A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis. Cancers (Basel). 2021 Dec 4;13(23):6116. doi: 10.3390/cancers13236116. PMID: 34885225; PMCID: PMC8656730.

[8] Marcello Di Giammarco, Camilla Vitulli, Simone Cirnelli, Benedetta Masone, Antonella santone, Mario Cesarelli, Fabio Martinelli, and Francesco Mercaldo. 2024. Explainable Deep Learning for Breast Cancer Classification and Localisation. ACM Trans. Comput. Healthcare Just Accepted (November 2024). https://doi.org/10.1145/3702237

[9] RANI, SUDHA & M, JOGENDRA. (2022). Histopathological Image Analysis Using Deep Learning Framework. 10.21203/rs.3.rs-2009418/v1.

[10] Oza P, Sharma P, Patel S, Adedoyin F, Bruno A. Image Augmentation Techniques for Mammogram Analysis. J Imaging. 2022 May 20;8(5):141. doi: 10.3390/jimaging8050141. PMID: 35621905; PMCID: PMC9147240.