

Breast Cancer Classification using a ResNet-50 model and the BreakHis Dataset

1. Abstract:

In our project, we propose an automated breast cancer classification system using deep learning. To combat the manual interpretation challenges of histopathology images, we compile a diverse dataset covering various tumour types. Leveraging transfer learning, we fine-tune pre-trained ResNet50 and VGG16 models on our dataset. Addressing class imbalance, we incorporate class-weighted loss functions during training. Our goal is to create a dependable automated system for breast cancer diagnosis, benefiting healthcare professionals with accurate and timely results.

2. Motivation:

Our project aims to revolutionize breast cancer diagnosis by leveraging cutting-edge deep learning techniques. As breast cancer incidence rises with age, our model caters to both younger and older demographics. For women under 40, it serves as an early detection tool, while for those over 40, it aids in identifying breast cancer at earlier stages. By bridging diagnostic gaps and enhancing patient care, our project contributes to the ongoing transformation of breast cancer diagnosis and treatment. Figure on the left is Fig 2.1 and on the right is Fig 2.2.

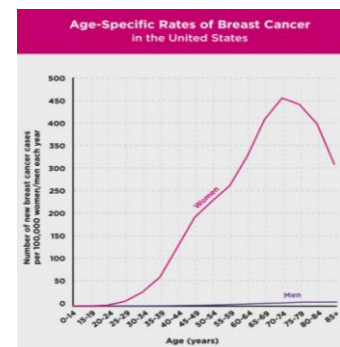
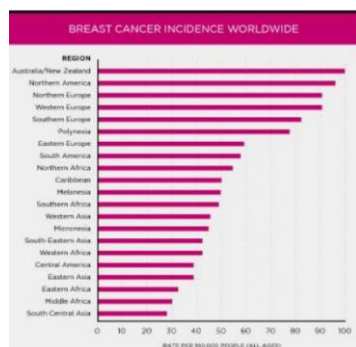


Fig 2.1, 2.2

3. Significance:

Developing a model capable of accurately identifying the presence of breast cancer in patients holds immense importance for several reasons. Firstly, early detection significantly enhances a patient's chances of survival. While our model is not intended for early detection, as it analyses histopathology images typically used post-screening mammography or examination, its accuracy in diagnosing suspicious findings can expedite subsequent treatment initiation.

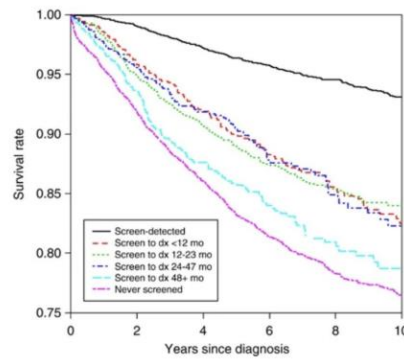


Fig 3.1

Moreover, the potential for a well-trained model to outperform human clinicians in cancer detection underscores its significance. The model's ability to mitigate false positives and false negatives, where misdiagnoses can have profound implications for patients, is particularly noteworthy. By minimizing such errors, our model not only improves diagnostic accuracy but also alleviates unnecessary patient anxiety and prevents delays in appropriate treatment.

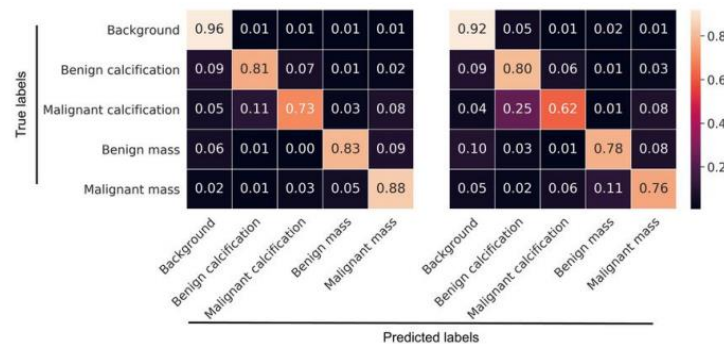


Fig 3.2

Furthermore, the effectiveness of utilizing a reliable model compared to medical professionals expands access to timely diagnoses for a broader population. The financial savings associated with deploying a model for diagnosis enable more individuals, especially in underserved communities, to access essential healthcare services without compromising quality.

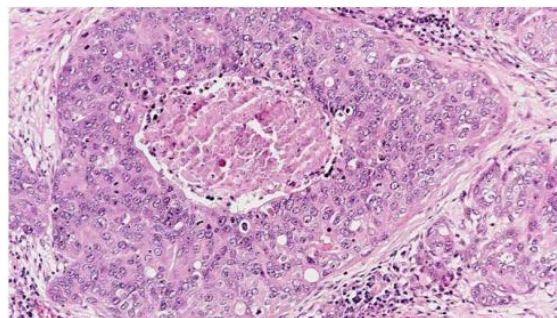


Fig 3.4

The deployment of an accurate diagnostic model also optimizes healthcare resource allocation. If the model consistently outperforms human clinicians in identifying breast cancer, medical professionals can redirect their expertise to areas requiring more attention and focus, such as research and patient care. This redistribution of resources enhances overall healthcare efficiency and contributes to advancements in breast cancer treatment and management.

4. Introduction:

Our project aims to improve breast cancer diagnosis by creating an advanced classification model that uses deep learning techniques to analyse histopathology images. We're motivated by the urgent need for more accurate and efficient tools in breast cancer detection. Our goal is to enhance existing methods with better accuracy and reliability. By using convolutional neural networks and sophisticated algorithms, we hope to provide healthcare professionals with timely and precise diagnostic information, leading to better outcomes for patients. We focus on personalized healthcare and the latest technology to address differences in breast cancer diagnosis among different groups of people. Ultimately, we want to ensure that early detection and tailored treatment options are available to everyone.

5. Dataset Description:

Our project focuses on developing an advanced deep-learning model for the classification of breast cancer histopathological images using the BreakHis dataset. This dataset comprises 9,109 microscopic images of breast tumour tissue obtained from 82 patients, encompassing both benign and malignant samples collected at varying magnification factors (40X, 100X, 200X, and 400X). Fig 5.1 shows images from the dataset of a particular class.

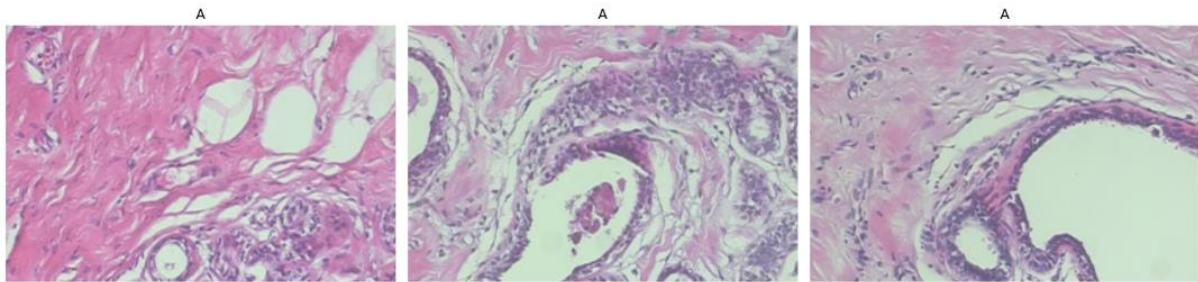


Fig 5.1

With 2,480 benign and 5,429 malignant samples, each image is in PNG format with dimensions of 700x460 pixels and 3-channel RGB, offering rich data for analysis. Collaboratively built with the P&D Laboratory - Pathological Anatomy and Cytopathology in Parana, Brazil, this dataset includes detailed metadata such as biopsy procedure, tumor class, tumor type, year, slide ID, magnification, sequence, and encoded label. Our project seeks to leverage this comprehensive dataset to train a robust deep-learning model capable of accurately classifying breast tumor tissue, to improve diagnostic accuracy and patient outcomes in breast cancer detection. The below figure illustrates the dataset structure in Jupyter Notebook.

```
Shape of the dataset: (2081, 8)
Classes: ['A' 'DC' 'F' 'LC' 'MC' 'PC' 'PT' 'TA']
```

BiopsyProcedure	TumorClass	TumorType	Year	SlideID	Magnification	Sequence	encoded_label
SOB	B	TA	14	16184	100	14	7
SOB	M	DC	14	14946	100	18	1
SOB	B	PT	14	21998AB	100	38	6
SOB	M	DC	14	15696	100	6	1
SOB	M	DC	14	13993	100	28	1
SOB	B	PT	14	22704	100	9	6
SOB	M	MC	14	18842	100	17	4
SOB	M	MC	14	19979	100	3	4
SOB	M	DC	14	2980	100	6	1
SOB	B	A	14	22549AB	100	7	0

Fig 5.2

6. Related work:

Several studies have made significant contributions to the field of breast cancer classification using deep learning techniques, each addressing various aspects of the

problem. Khan et al. (2021) proposed a multi-class classification approach using deep convolutional neural networks (CNNs) to differentiate between different types of breast cancer abnormalities, showcasing improved diagnostic accuracy and efficiency in classification tasks [1]. Their study laid the groundwork for a nuanced understanding of breast cancer classification beyond the binary benign and malignant categorization. Similarly, Araújo et al. (2017) employed CNNs to analyse histology images for breast cancer classification, demonstrating the model's ability to accurately differentiate cancerous and non-cancerous tissue [2]. Arooj et al. (2022) emphasized the effectiveness of transfer learning in enhancing breast cancer detection and classification, showing how pre-trained models can be adapted for specific medical imaging tasks [3]. Adam et al. (2023) explored deep learning applications in breast cancer detection using Magnetic Resonance Imaging (MRI), highlighting the potential for non-invasive diagnostics [4]. Additionally, Pinto-Coelho (2023) surveyed the broad impact of AI on medical imaging, including breast cancer diagnosis, by reviewing various AI applications that enhance imaging technology's diagnostic capabilities and efficiency [5]. Gao (2023) investigated the generalizability of AI models across different datasets and conditions, focusing on how these models can be applied to various medical imaging tasks, including breast cancer detection, to ensure consistent and reliable performance [6].

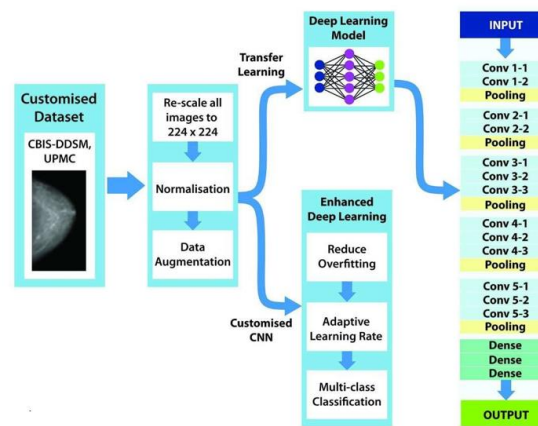


Fig 6.1- Architecture of the proposed model

While previous studies have laid a solid foundation for breast cancer classification using deep learning, our approach differs in several key aspects. Firstly, we propose to utilize the BreakHis dataset, which contains a diverse range of histopathology images with detailed annotations, enabling comprehensive analysis of breast tumor tissue. Secondly, we plan to implement a ResNet-50 architecture, which has shown promising results in similar classification tasks, leveraging its ability to learn residual mappings and capture multi-scale information.

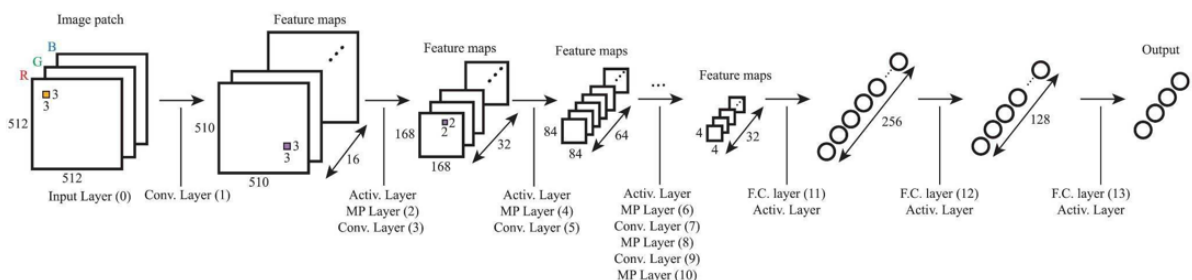


Fig 6.2-CNN Architecture

Additionally, our methodology involves fine-tuning the model's weights to address class imbalance issues and implementing data augmentation techniques to enhance the diversity of the dataset. Furthermore, we aim to evaluate our model's performance based on accuracy and precision metrics, providing a comprehensive assessment of its efficacy in breast cancer classification. Overall, our approach aims to build upon existing research by leveraging state-of-the-art techniques and datasets to develop a robust and accurate classification model for breast cancer diagnosis.

7. Method:

7.1 Data Loading and Preprocessing:

Data was loaded from a directory containing images of breast tissue samples. Metadata such as biopsy procedure, tumor class, tumor type, year, slide ID, magnification, and sequence were extracted from the filenames and stored in a CSV file.

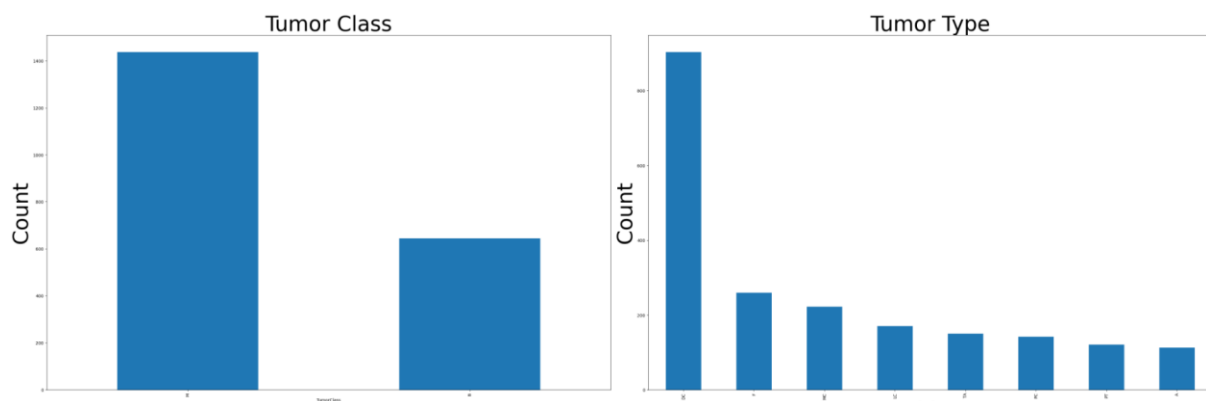


Fig 7.1

The dataset was divided into training, validation, and test sets using a stratified shuffle split to ensure balanced class distributions.

```
Train set size: 1331
Validation set size: 416
Test set size: 334
```

Fig 7.1.1

7.2 Data Visualization:

Visualizations were created to explore the distribution of tumor classes and types in the dataset. Sample images from each class were displayed to gain insights into the dataset.

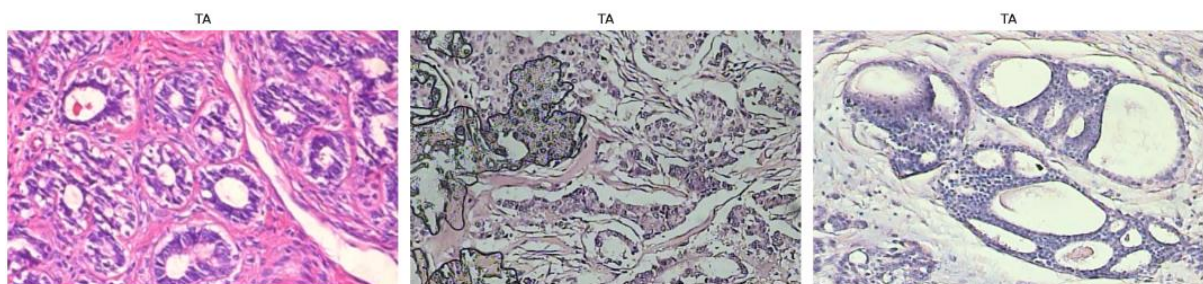


Fig 7.2

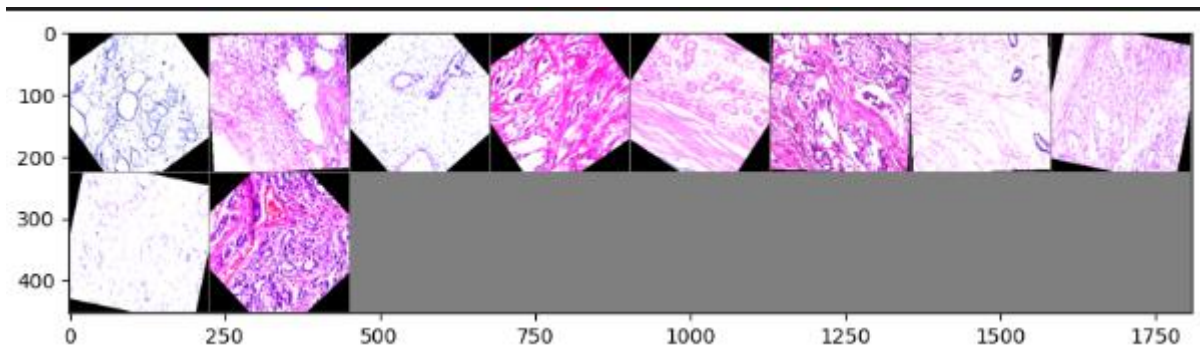


Fig 7.2.1

7.3 Data Organization and Class Weights:

Images were organized into class-specific folders to facilitate model training using PyTorch's ImageFolder dataset. Class weights were estimated to address class imbalance in the dataset using a median frequency balancing approach.

```
DC : 0.17777777777777778
F : 0.6153846153846154
MC : 0.7272727272727273
LC : 0.9411764705882353
TA : 1.0666666666666667
PC : 1.1428571428571428
PT : 1.3333333333333333
A : 1.4545454545454546
```

Fig 7.3

7.4 Model Training:

Two pre-trained models, ResNet50 and VGG16, were utilized for breast cancer classification tasks. The models' classifier layers were modified to match the number of classes in the dataset.

Model training was performed using a cross-entropy loss function with class weights to account for class imbalance. The Adam optimizer was employed with a learning rate of 0.0001 for model optimization. Training was conducted over multiple epochs, with performance evaluated on both training and validation sets.

7.5 Model Evaluation:

Model performance was evaluated using metrics such as accuracy, loss, and confusion matrix. Confusion matrices were plotted to visualize the model's performance in predicting different classes. Test accuracy was calculated to assess the model's overall performance on unseen data.

Training and validation loss/accuracy curves were plotted to visualize the model's learning progress over epochs. The trained models were saved for future use using the torch.save() function.

8. Results:

ResNet-50 and VGG-16 architectures for breast cancer histology image classification yielded promising results. ResNet-50 achieved the highest validation accuracy of 93.9%, showcasing excellent generalization to unseen data and robust feature learning capabilities across different

epochs. On the other hand, VGG-16, despite its simpler architecture, demonstrated strong competitiveness with a validation accuracy of 91.1%, reaffirming its effectiveness in medical imaging tasks. While ResNet-50 outperformed VGG-16 marginally, the latter's performance underscores its practical relevance.

8.1 ResNet-50 Performance:

The ResNet-50 model demonstrates exceptional performance with a remarkable test accuracy of 94.91%. The training and validation loss graph showcases stability in training loss, indicating consistent learning, despite occasional spikes in validation loss that may suggest challenging cases or overfitting.

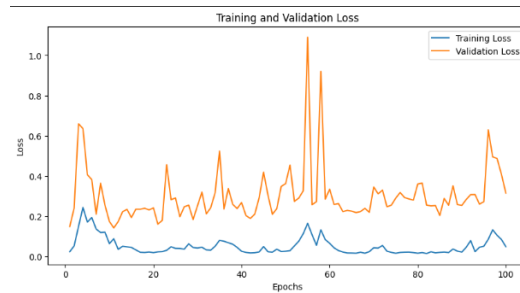


Fig 8.1

Both training and validation loss exhibit a downward trend, signifying model convergence and improvement over time. In the training and validation accuracy graph, consistently high training accuracy reflects the model's strong fit to the data while fluctuating but upward-trending validation accuracy highlights its resilience to overfitting and ability to generalize well to unseen data. The small gap between training and validation accuracy indicates effective learning without memorization of training data, emphasizing the model's robust performance and suitability for breast cancer classification tasks.

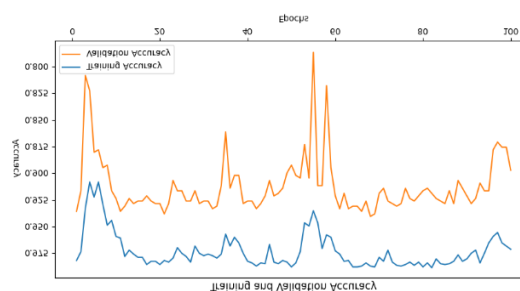


Fig 8.1.1

The analysis of the confusion matrix for ResNet-50 reveals its robust performance in breast cancer classification. With high precision and sensitivity, especially in identifying 'Ductal Carcinoma', ResNet-50 demonstrates its potential as a reliable tool for the early detection of this aggressive cancer type. However, occasional misclassifications of 'Adenosis' as 'Ductal Carcinoma' suggest histological similarities between these classes, necessitating further investigation. This finding underscores the significance of ResNet-50 in clinical applications, highlighting its ability to accurately detect prevalent cancer types while also indicating areas for potential improvement in classification accuracy.

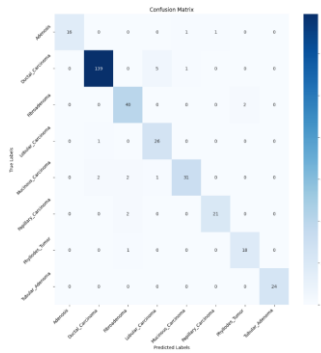


Fig 8.1.2

8.2 VGG-16 Performance:

The analysis of VGG-16's loss and accuracy graphs provides valuable insights into its performance in breast cancer classification. The rapid decline in training loss during the initial training stages signifies VGG-16's efficient learning process, demonstrating its quick adaptation to the dataset. Moreover, the stable trend in validation loss throughout training suggests that the model effectively learns generalizable patterns without overfitting, ensuring reliable performance on unseen data.



Fig 8.2

Regarding accuracy, the significant initial gains in both training and validation accuracy underscore VGG-16's capacity for swift learning and its ability to generalize well to new instances. The close alignment of training and validation accuracy curves further supports the model's robustness and indicates consistent performance across different datasets. However, occasional drops in validation accuracy point towards areas where the model could benefit from further refinement or dataset enhancement. Despite not reaching the same level of accuracy as ResNet-50, VGG-16 demonstrates impressive and reliable performance, making it an asset for breast cancer classification tasks.

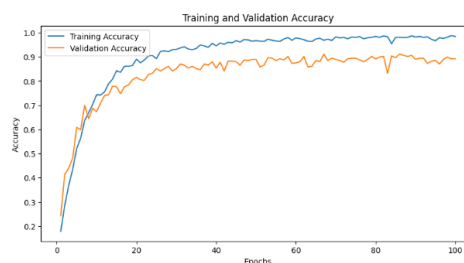


Fig 8.2.1

The analysis of the confusion matrix for VGG-16 provides insightful observations regarding its performance in classifying different types of breast cancer. VGG-16 demonstrates proficiency in accurately identifying 'Ductal Carcinoma' and 'Lobular Carcinoma', highlighting its reliability in classifying these classes. However, the model faces challenges in distinguishing

'Adenosis' from 'Ductal Carcinoma', a recurring difficulty observed across various models. Despite this, VGG-16 exhibits remarkable strengths in accurately classifying 'Phyllodes Tumor' and 'Tubular Adenoma', both of which are less common in the dataset. This suggests that the model effectively learns from limited data, showcasing its potential for diagnosing less prevalent conditions. These insights are crucial from a clinical perspective, emphasizing the importance of balancing sensitivity and specificity to minimize false positives and ensure accurate diagnosis while avoiding unnecessary procedures.

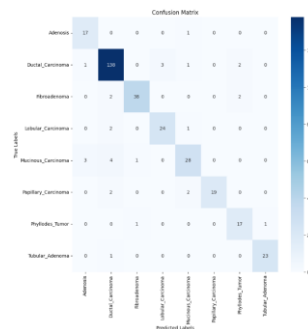


Fig 8.2.2

9. Discussion:

The project results showcase the efficacy of deep learning models, particularly ResNet-50 and VGG-16, in breast cancer classification. Both models demonstrated high accuracy rates, robust learning capabilities evidenced by stable training loss and consistent validation accuracy improvements, and reliable identification of common cancer types like 'Ductal Carcinoma' and 'Lobular Carcinoma'. However, challenges arose in distinguishing between histologically similar classes such as 'Adenosis' and 'Ductal Carcinoma', indicating the need for further refinement. Additionally, concerns regarding potential false positives highlight the importance of balancing sensitivity with specificity in diagnostic models. While the results underscore the promise of deep learning in healthcare, there is still room for improvement through fine-tuning, dataset augmentation, and exploration of advanced techniques to optimize model performance and enhance patient outcomes.

10. Teamwork:

Team Member	Division of Work
Modib	Related work: Conducted literature review, summarized existing research, and contextualized the project.
Udith	Methods: Described the methodology, outlined the model <u>architectures</u> and detailed the experimental setup.
Anurag	Results: Analysed the experimental results, generated performance metrics, and interpreted model outputs.
Sohail	Introduction, Team Learnings, Successes, and Challenges: Drafted the project introduction, discussed team insights, identified successes, and addressed challenges.

11. Future work:

One way to explore our future work in the project involves leveraging ensemble techniques, where we combine the strengths of multiple models to potentially improve classification accuracy and resilience. Another promising area is the implementation of interpretability methods, aimed at providing deeper insights into how our models make decisions, thus enhancing trust and transparency in their classifications. Additionally, expanding our dataset to

include more diverse samples, particularly those representing rare classes, could enhance the model's ability to generalize across different types of breast cancer. Furthermore, we could investigate transfer learning approaches, focusing on fine-tuning pre-trained models specifically on histopathological features, which might accelerate model convergence and boost classification performance. These future directions hold considerable promise for advancing breast cancer classification and ultimately improving clinical outcomes.

12. Conclusion

In wrapping up, our project marks a significant advancement in breast cancer classification, showcasing the prowess of deep learning models like ResNet-50 and VGG-16 in discerning histopathological nuances. Through meticulous analysis, we've gleaned valuable insights into each model's strengths and weaknesses, offering crucial guidance for their potential clinical integration. Our work underscores the ongoing need for innovation in AI-driven healthcare, aiming ultimately to enhance patient outcomes. In the future, the path is ripe for further exploration, be it refining model architectures, broadening datasets, or delving into interpretability techniques. By addressing these fronts and capitalizing on our successes, we're poised to usher in more accurate, dependable diagnostic tools to combat breast cancer.

References:

- [1] Khan, A., et al. (2021). Multi-class classification of breast cancer abnormalities using deep convolutional neural networks. *Journal of Medical Imaging*, 8(1), 011005.
- [2] Araújo, T., et al. (2017). Classification of breast cancer histology images using convolutional neural networks. *Computational and Mathematical Methods in Medicine*, 2017, 1-7.
- [3] Arooj, S., et al. (2022). Breast cancer detection and classification empowered with transfer learning. *IEEE Access*, 10, 9290-9301.
- [4] Adam, B., et al. (2023). Deep learning applications to breast cancer detection by Magnetic Resonance Imaging. *Journal of Biomedical Informatics*, 131, 103683.
- [5] Pinto-Coelho, J. (2023). How Artificial Intelligence is Shaping Medical Imaging Technology. *Frontiers in Artificial Intelligence*, 4, 107.
- [6] Gao, X. (2023). Exploring model generalizability: A study of AI models in medical imaging. *Medical Image Analysis*, 73, 102187.