Using CNNs to Detect Breast Cancer From Histological Images

Qi-Rui Edward Chee

Abstract

Breast cancer is one of the most common cancers worldwide, and early, accurate diagnosis is essential for improving patient outcomes. Histopathological image analysis from biopsy remains a key method for identifying malignant and benign breast tissue, but manual review can be time consuming which may delay treatment decisions when the cancer progresses rapidly. In this study, we evaluated several deep learning methods including the CNN and pre-trained models for automated classification of breast cancer between the malignant and benign samples from a curated dataset of histological images. Images were balanced before training, and model performance was evaluated using metrics including accuracy, precision, recall, and confusion matrix analysis. The results showed that the models were able to achieve strong classification performance, with the Keras Sequential having the highest accuracy (85.7%) and precision (88.4%), and the VGG19 the highest recall (94.8%). These findings support the potential of deep learning as a tool for assisting histopathological breast cancer diagnosis, and potential future work exploring ensemble approaches combining the best-performing models to further improve robustness and classification accuracy in challenging cases.

Keywords: breast cancer, histological images, artificial intelligence, deep learning, cancer detection

Introduction

Breast cancer is one of the most common cancers in women and a leading cause of cancer-related deaths worldwide [1]. It is the abnormal replication of the body's cells in breast tissue, either in the gland tissues or supporting tissue [2], and has many subtypes, such as ductal carcinoma, lobular carcinoma, and tubular carcinoma, some of which are invasive [3]. If left untreated, the cancer cells can invade into surrounding tissues, or spread to other parts of the body [2], worsening a patient's health and prognostic outcome. Therefore, early and accurate detection and diagnosis are critical to improve patient outcomes.

The clinical gold standard for diagnosing cancer is to perform a biopsy, during which tissue is removed from the breast cancer area to obtain histological tissue samples, which are then examined under a microscope to examine for the presence of cancerous cells [4]. At this stage, a microscope or digital scanner captures high-resolution images of the histological tissue samples that can be used for further analyses. As histological tissues from breast cancer often exhibit significant heterogeneity across subtypes, patients and imaging conditions, this process of determining the malignancy of the tissue sample, while effective, relies heavily on manual pathologist review and is time-intensive, sometimes taking up to several weeks, that can delay treatment decisions when the cancer progresses rapidly [5].

Artificial intelligence (AI), particularly deep learning, has emerged as a promising approach to address this limitation. Unlike traditional approaches that rely on hand-crafted features that may not fully capture the complexity of histopathological images, machine learning methods have the ability to automatically detect, learn and characterize patterns directly from image data [6]. Deep learning models, especially Convolutional Neural Networks (CNNs) such as VGG16, VGG19, and ResNet50, can learn patterns and hierarchical feature representations directly from histological images, allowing them to characterize complex morphological patterns, including cellular structure and tissue organization. This ability makes these methods well-suited to cancer detection distinguishing between malignant and benign samples.

The CNN is a widely used machine learning algorithm that detects patterns, also called features, in images. It first looks for simple shapes like lines, and then more complex patterns like objects, see Fig. 1. A CNN does this by processing an image through several layers of neurons (Fig. 2). In the first layers, the CNN decomposes the image into small regions using filters or small sections of pixels (also called kernels), which scan across the image and assign numerical values based on the patterns they detect. These kernels help identify simple features such as lines, edges and textures. As the kernels move through more and deeper layers, the model begins to recognize more complex patterns, such as cell shapes and how tissues are organized [7]. To make the model more effective, activation functions are applied to help it learn complex patterns, and pooling layers are used to reduce the size of the data while keeping the most important information. This helps the model focus on the most relevant features and improves efficiency. At the end of the network, the extracted features are combined to make a final prediction determining the class. The CNN model outputs a probability indicating whether the image is cancerous or not.

By learning patterns directly from the images, CNNs can identify important differences between malignant and benign tissue, making them useful and proficient for analyzing histopathological images. The CNN model is highly customizable; almost all fields and parameters can be modified. It also takes less time to train compared to other models [8].

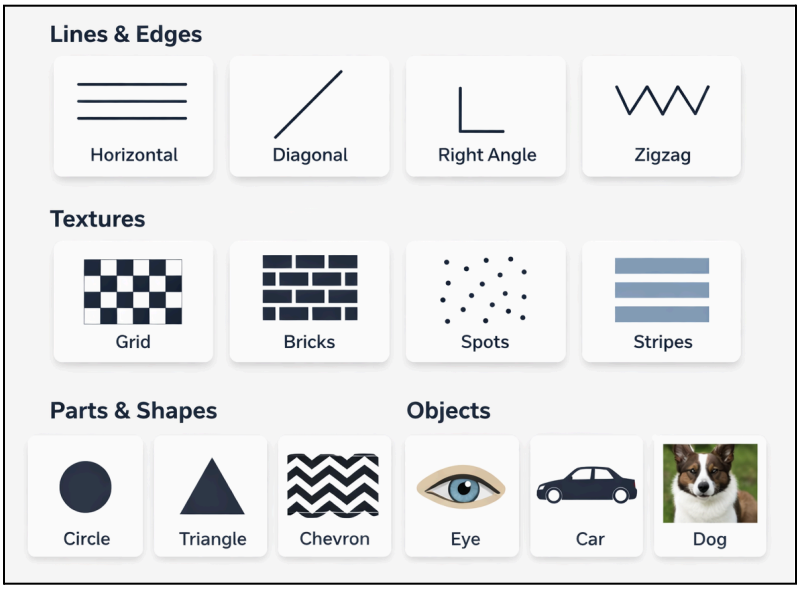


Fig. 1 Example of patterns or features recognized by CNN models

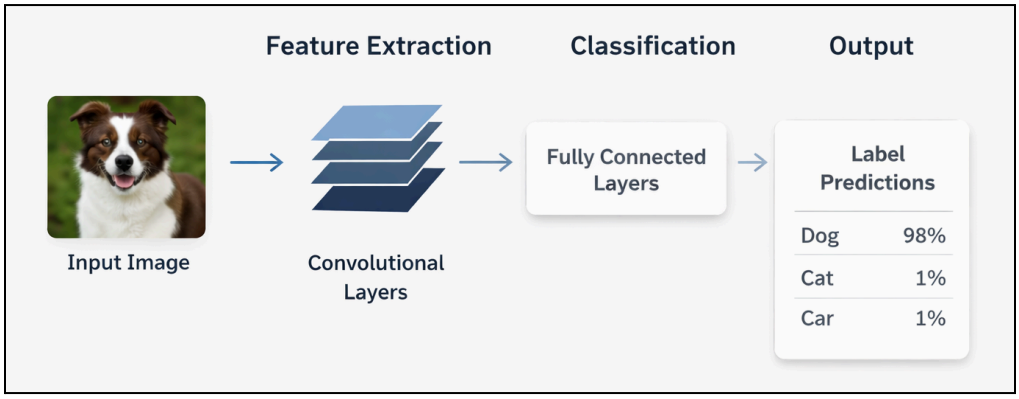


Fig. 2 A simplified illustration of image processing by CNN

The VGG16 and VGG19 models, on the other hand, are deep learning models that had been developed to easily recognize common patterns and features by pre-training them on large datasets. These models use many more parameters and layers respectively, 16 and 19 neuron layers, to process an image. Similar to the sequential model, the VGG16 and VGG19 models run an image through the layers of kernels to eventually produce a prediction probability if a sample is cancerous or not. With the deep network layers, VGG16 and VGG19 models are able to capture detailed and complex features in the image. Being pretrained models, one can use transfer learning to apply and fine-tune these models on new datasets [9, 10, 11].

The ResNet50 model is also a deep learning model that processes an image similar to CNN, VGG16, VGG19 models. However, a key feature of ResNet50 is the use of residual connections, which allow information to bypass certain layers and be passed directly to later layers as an image passes through the kernel layers. helps the model learn more effectively, especially in very deep networks, and improves its ability to capture complex patterns without losing important information [9, 10, 11].

Across these models, while the highly customized CNN model provides a flexible architecture that can be adjusted and trained directly on the dataset, the VGG16 and VGG19 models are deeper models that have already been trained on large datasets such as ImageNet, allowing them to recognize general patterns and extract more detailed features. On the other hand, the ResNet50 model further improves on this by using residual connections, which allow information to pass through the network more effectively and help the model learn complex patterns in very deep layers.

In this paper, we build upon these AI deep learning based models to assess and investigate their accuracy when used to diagnose breast cancer from histological images that are heterogeneous across subtypes and patients. Using a curated histological image dataset and a binary classification task, we evaluate multiple deep learning models with varying parameters to assess

their performance and potential as efficient, reliable AI-assisted tools for binary classification on breast cancer samples as malignant or benign.

Methods

Models. We assessed several CNN models including a Keras Sequential, as well as pretrained models such as VGG16, VGG19, and ResNet50, with epochs ranging from 10 to 20. We trained our models on a pre-processed and balanced BreakHis breast cancer histology image dataset, assessed their performance based on accuracy, precision and recall metrics. For VGG16, VGG19 and ResNet50, we used transfer learning by applying these pretrained models to our dataset and fine-tuning them for binary classification. All models were implemented using the TensorFlow and Keras libraries in Python.

Dataset. We used a dataset consisting of 3,200 histological images curated from the BreakHis breast cancer dataset, where each image is labeled as either malignant or benign [12]. The BreakHis dataset from Kaggle contains over 7,900 histopathological images across eight tissue classes: adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucous carcinoma, and papillary carcinoma [12].

Adenosis, fibroadenoma, phyllodes tumor, tubular adenoma are benign [4]. Adenosis has two main subtypes: sclerosing adenosis and microglandular adenosis. It starts in the lobules and looks like many disordered glands clumped together under a microscope [13, 14]. Fibroadenoma is a benign growth that forms from the fibrous tissue in the breast. It can sometimes grow quickly, and surgical removal is the best option for treatment [15]. Phyllodes tumors are relatively rare benign fibroepithelial growths that present as a rapidly growing lump in the breast. A phyllodes tumor can become cancerous and metastasize, but it has a rare chance of occurring [16]. Tubular adenomas are rare, benign growths from ductal tissues. Tubular adenomas can develop into malignant tumors, but it is very rare [17].

Ductal carcinoma, lobular carcinoma, mucous carcinoma, and papillary carcinoma are malignant [3]. Ductal carcinoma starts in the duct lobular units and can become invasive and metastasize [18]. It can be diagnosed by seeing rod-like calcifications on a mammogram [19]. Lobular carcinoma occurs when cells lose their adhering molecule and become cancerous, invading into other tissues. Under a microscope, the cells sometimes form a characteristic straight line. Lobular carcinoma may not always form a lump in the breast, and using a mammogram or ultrasound might not detect it due to diffuse growths [20]. Mucinous carcinoma happens when the tumor cells make a lot of mucus, giving the cancer its name. They can be detected by mammogram as a small to medium sized visible lump. Papillary carcinoma is a rare type of breast cancer in which the tumor cells form finger-like (“papillary”) projections under the microscope. It is typically detected after an abnormal mammogram, breast ultrasound, MRI, or after symptoms such as a breast lump or bloody nipple discharge prompt imaging. The diagnosis is then confirmed by biopsy, with tissue examined by a pathologist to identify the papillary features of the tumor [21].

As shown in Fig. 3, the histological images displayed significant heterogeneous patterns across the different tissue classes.

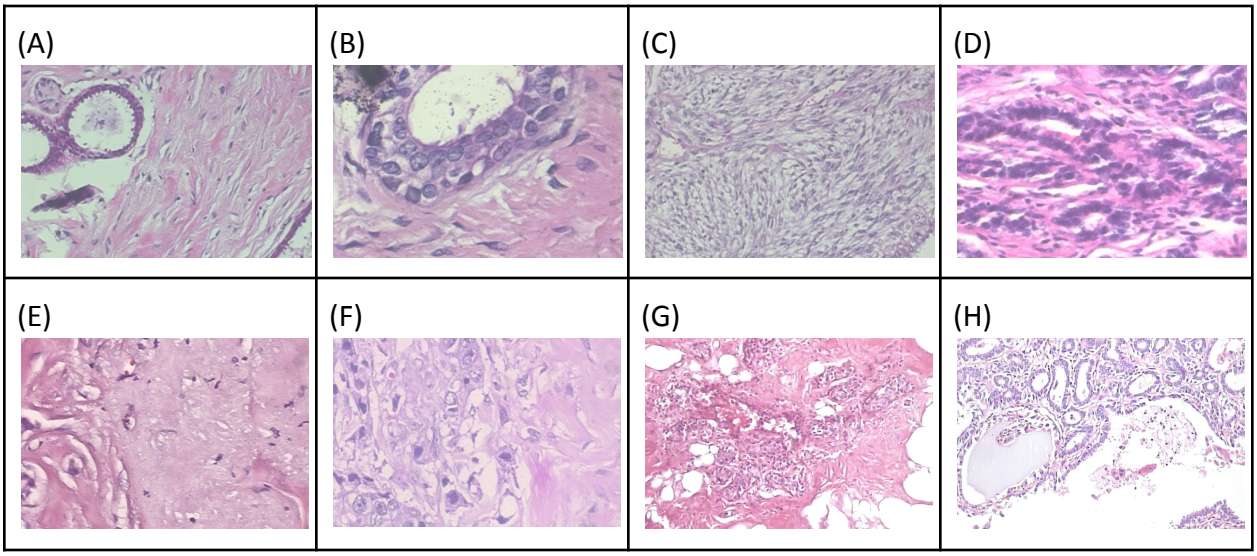


Fig. 3 Histological images of each tissue class from BreakHis dataset: A) adenosis, B) fibroadenoma, C) phyllodes tumor, D) tubular adenoma, E) ductal carcinoma, F) lobular carcinoma, G) mucinous carcinoma, H) papillary carcinoma

Data Pre-processing. It was observed that the BreakHis dataset was dominated by images on ductal carcinoma. To reduce bias toward the majority tissue class, the dataset was pre-processed to ensure equal representation of malignant and benign breast cancer tissues. The dataset was balanced by selecting an equal number of images from each class. After data balancing, we obtained a curated final dataset of 3,200 images, with equal representation from each tissue class [12], to evaluate our models for the binary classification task.

Model Training. To train the models, the final dataset was split into a training set (70%), a validation set (30%), and a testing set consisting of every other image in the final dataset. The training set was used to optimize model parameters, the validation set to monitor performance during training, and the test set for final evaluation. All models were trained using the Adam optimizer with a learning rate of 0.0001 and the sparse_categorical_crossentropy function. Training was conducted for 10 and 20 epochs with a batch size of 16.

Metrics. To evaluate and visualize the final performance of each model, we used a confusion matrix and plotted accuracy across training epochs. In addition to accuracy that measures overall correctness, precision and recall were also computed from the confusion matrix. Precision is the percentage of images predicted to be malignant that actually showed cancer, and recall is the percentage of malignant images that were correctly identified by the model.

Results

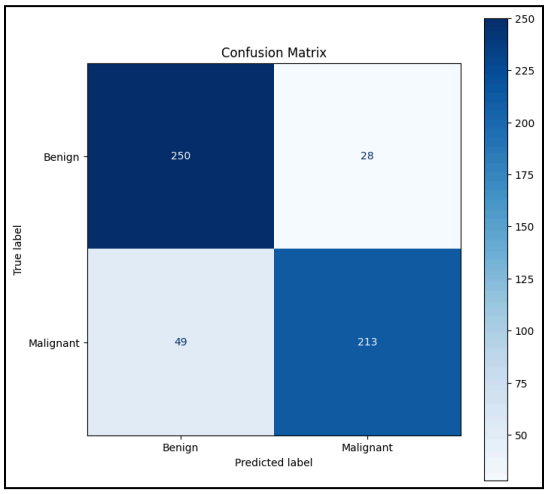
Overall, all the models,the Keras Sequential, VGG16, VGG19, and ResNet50, demonstrated strong performance in classifying histopathological images as malignant or benign.

The confusion matrix and learning curves across epochs for each model are illustrated in Fig. 4 and 5, showing high accurate classification rates with most malignant and benign samples being correctly classified, with relatively few false positives and false negatives.

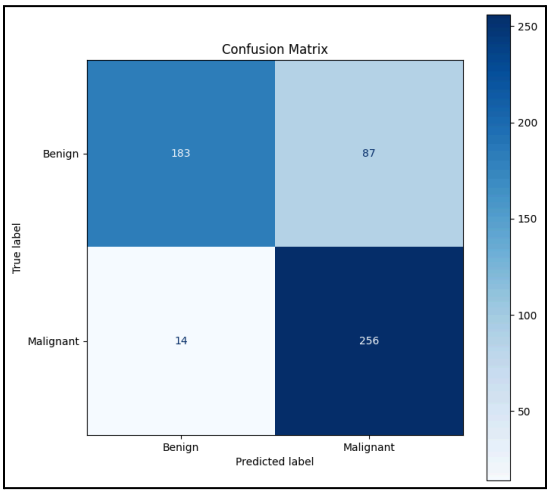
The performance metrics of each model are summarized in Table 1. The Keras Sequential model that trained for 20 epochs had the highest accuracy of 85.7% and highest precision of 88.4%; the VGG 19 model that trained for 10 epochs had the highest recall of 94.8%. The sequential model that trained for 20 epochs best correctly predicted if an image was malignant, making it a rule in model, which means that if it predicts malignant, then it being malignant can be ruled in with high probability. The VGG19 model has the highest recall, identifying the most malignant images, making it a rule out model, so if it predicts that an image is malignant, it being benign can be ruled out.

Out of these models, both the Keras Sequential with 10 and 20 epochs, and VGG16 perform well on the final curated dataset achieving more than 80% on all three performance metrics. The Keras Sequential with 20 epochs has 85.7% accuracy, 88.4% precision and 81.3% recall; Keras Sequential with 10 epochs has 81.5% accuracy, 83.7% precision and 80.1% recall; the VGG16 has 84.3% accuracy, 81.4% precision and 88.4% recall.

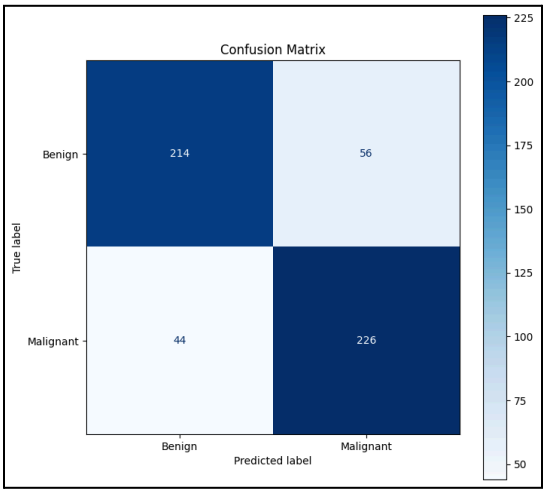
(a)


Confusion Matrix
Benign
Malignant
True label
250
28
49
213
Benign
Malignant
Predicted label
250
225
200
175
150
125
100
75
50

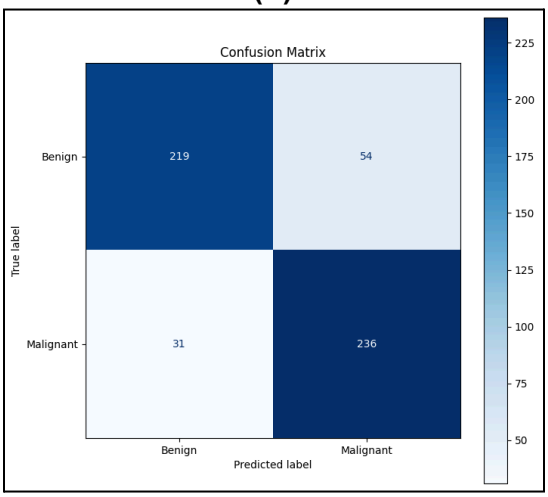
(b)


Confusion Matrix
Benign
Malignant
True label
183
87
14
256
Benign
Malignant
Predicted label
250
200
150
100
50

(c)


Confusion Matrix
Benign
Malignant
True label
214
56
44
226
Benign
Malignant
Predicted label
225
200
175
150
125
100
75
50

(d)


Confusion Matrix
Benign
Malignant
True label
219
54
31
236
Benign
Malignant
Predicted label
225
200
175
150
125
100
75
50

(e)

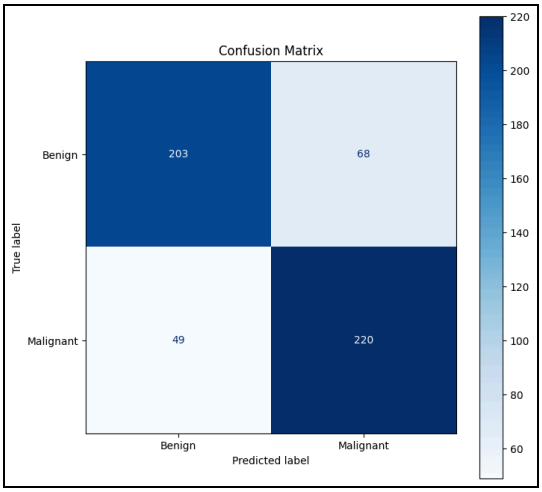
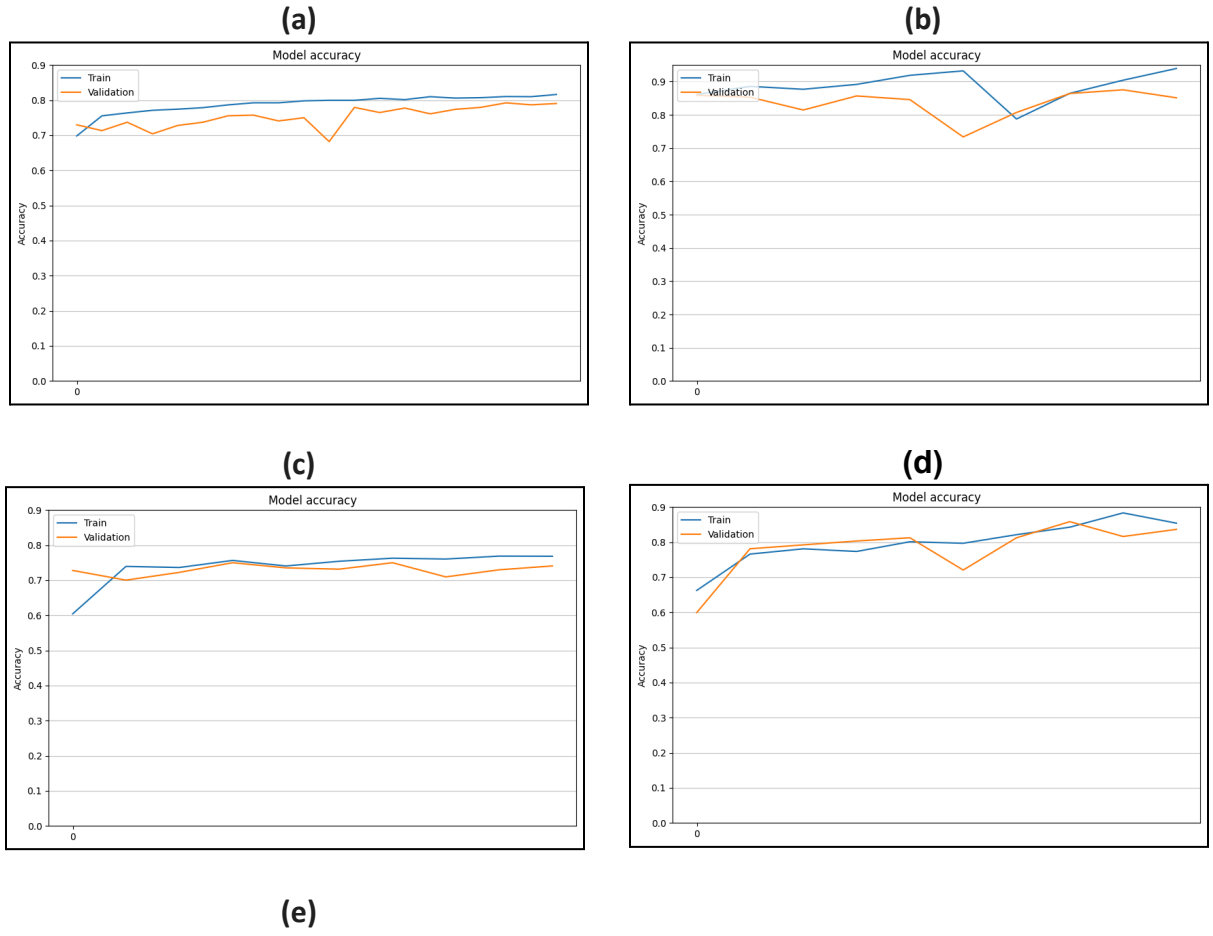


Fig. 4 Confusion matrix for (a) Keras Sequential with 20 epochs, (b), VGG19, (c) Keras Sequential with 10 epochs, (d) VGG16, (e) ResNet50



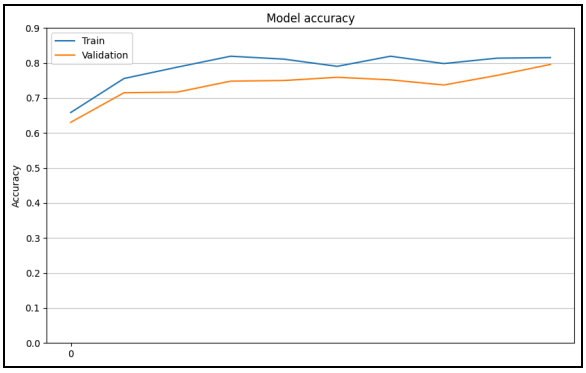


Fig. 5 Learning curves showing training and validation accuracy across training epochs for (a) Keras Sequential with 20 epochs, (b), VGG19, (c) Keras Sequential with 10 epochs, (d) VGG16, (e) ResNet50

Model	Parameters	Accuracy	Precision	Recall
Keras Sequential	Epochs=20 Kernel size=3 Pooling (2,2)	**85.7%**	**88.4%**	81.3%
Keras Sequential	Epochs=10, Kernel size=3 Pooling (2,2)	81.5%	83.7%	80.1%
VGG16	Epochs=10	84.3%	81.4%	88.4%
VGG19	Epochs=10	81.3%	74.6%	**94.8%**
ResNet50	Epochs=10	78.3%	81.8%	76.4%

Table 1 Performance metrics for each model

Discussion

The results of this study demonstrate that deep learning models can effectively classify breast cancer histopathological images as malignant or benign, achieving high performance across several performance metrics on the final curated dataset. Among the models tested, both the Keras Sequential with 10 and 20 epochs, and VGG16 perform well with strong overall classification ability with accuracy, precision and recall more than 80%. The Keras Sequential with 20 epochs has 85.7% accuracy, 88.4% precision and 81.3% recall; Keras Sequential with 10 epochs has 81.5% accuracy, 83.7% precision and 80.1% recall; the VGG16 has 84.3% accuracy, 81.4% precision and 88.4% recall.

While pretrained architectures such as VGG16, VGG19, and ResNet50 generally outperformed the Keras Sequential models, the Keras Sequential models in this case, performed comparably similarly to the VGG16 across accuracy, precision and recall on the curated dataset. The Sequential model with 20 epochs has the highest accuracy (85.7%) and highest precision (88.4%). These suggest that when the Sequential model uses well-selected parameters such as the number of epochs, it can perform well against the pretrained models.

From these observations, future work may investigate whether combining the strongest individual models can improve accuracy and other performance metrics while providing a more stable framework for breast cancer image classification. For example, one can investigate an ensemble approach using the two best-performing models from this study. Because each model may capture different image features and make different classification errors, combining their predictions could improve overall robustness and predictive performance beyond that of either model alone by reducing model-specific weaknesses and increasing reliability, particularly for difficult or borderline cases.

Despite these promising results, several considerations may limit the generalization of these findings. First, the dataset was curated into a binary classification task, which simplifies real-world diagnostic scenarios that often require distinguishing between multiple cancer subtypes. Additionally, the curated dataset size (3,200 images) may limit the model's generalizability to external datasets. Another limitation is the reliance on transfer learning from ImageNet, which may introduce bias due to differences between images from the ImageNet and the curated dataset. Future work could also focus on expanding the dataset with more diverse and differently stained samples, extending the model to multi-class classification, and exploring advanced architectures and multimodal approaches that may enhance classification performance.

Overall, these results showed that deep learning models can achieve high accuracy and balanced performance in classifying breast cancer histopathological images, highlighting their potential as reliable tools for AI-assisted diagnosis.

Conclusion

In this study, we evaluated several deep learning–based models for the classification of breast cancer histopathological images as malignant or benign on a curated dataset of breast cancer histological images from BreakHis. The Keras Sequential and pretrained architectures such as VGG16, VGG19, and ResNet50 achieve high accuracy and effectively distinguish between cancerous and non-cancerous tissue. Among the models tested on the curated dataset, the Keras Sequential models, when used with well-selected parameters, can perform similarly to the pretrained models.

While observations from this study on the curated dataset are encouraging, further research work to validate these models on larger and more diverse datasets may help further improve generalizability and enhance interpretability for clinical use. Expanding to multi-class

classification and developing models using ensemble approach may further increase the reliability and applicability of these systems.

Overall, this study showed that AI-based methods have strong potential to assist in breast cancer detection by providing faster and more consistent image analysis. Such tools could help support clinical decision-making, reduce diagnostic workload, and improve early detection outcomes.

Acknowledgements

I would like to thank Rami Abi-Akl for his review and support for this project.

References

1. I. Zaqout. Breast cancer detection and classification using deep learning Xception algorithm. *International Journal of Advanced Computer Science and Applications.* Vol. **13**, pg. 223-228, 2022, https://dx.doi.org/10.14569/IJACSA.2022.0130729.

2. G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, K. K. Sharma. Various types and management of breast cancer: an overview. *Journal of Advanced Pharmaceutical Technology & Research*. Vol. **1**, no. 2, pg. 109–126, 2010, https://doi.org/10.4103/2231-4040.72251.

3. B. Weigelt, F. C. Geyer, J. S. Reis-Filho. Histological types of breast cancer: how special are they? *Molecular Oncology*. Vol. **4**, no. 3, pg. 192–208, 2010, https://doi.org/10.1016/j.molonc.2010.04.004.

4. M. Guray, A. A. Sahin. Benign breast diseases: classification, diagnosis, and management. *The Oncologist*. Vol. **11**, no. 5, pg. 435–449, 2006, https://doi.org/10.1634/theoncologist.11-5-435.

5. S. Lebel, G. Jakubovits, Z. Rosberger, C. Loiselle, C. Seguin, C. Cornaz, J. Ingram, L. August, A. Lisbona. Waiting for a breast biopsy: Psychosocial consequences and coping strategies. *Journal of Psychosomatic Research*. Vol. **55**, no. 5, pg. 437–443, 2003, https://doi.org/10.1016/S0022-3999(03)00512-9.

6. N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, A. Tsirigos. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine*. Vol. **24**, no. 10, pg. 1559–1567, 2018, https://doi.org/10.1038/s41591-018-0177-5.

7. J. Ayeni. Convolutional neural network (CNN): The architecture and applications. *Applied Journal of Physical Science*. Vol. **4**, no. 4, pg. 42–50, 2022, https://doi.org/10.31248/AJPS2022.085.

8. N. K. Manaswi. Understanding and working with Keras. *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras*. Apress, Berkeley, CA, pg. 31–43, 2018, https://doi.org/10.1007/978-1-4842-3516-4.

9. S. Tammina. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications*. Vol. **9**, no. 10, pg. 143–150, 2019, https://doi.org/10.29322/IJSRP.9.10.2019.p9420.

10. K. R. Bhatele, S. S. Bhadauria. Classification of neurodegenerative diseases based on VGG 19 deep transfer learning architecture: a deep learning approach. *Bioscience Biotechnology Research Communications*. Vol. **13**, no. 4, pg. 1972–1980, 2020, https://doi.org/10.21786/bbrc/13.4/51.

11. N. Behar, M. Shrivastava. ResNet50-based effective model for breast cancer classification using histopathology images. *Computer Modeling in Engineering & Sciences*. Vol. **130**, no. 2, pg. 823–839, 2022, https://doi.org/10.32604/cmes.2022.017030.

12. BreakHis. *Kaggle*. 2019, https://www.kaggle.com/datasets/ambarish/breakhis.

13. D. W. Visscher, A. Nassar, A. C. Degnim, M. H. Frost, R. A. Vierkant, R. D. Frank, Y. Tarabishy, D. C. Radisky, L. C. Hartmann. Sclerosing adenosis and risk of breast cancer. *Breast Cancer Research and Treatment*. Vol. **144**, no. 1, pg. 205–212, 2014, https://doi.org/10.1007/s10549-014-2862-5.

14. M. P. Foschini, V. Eusebi. Microglandular adenosis of the breast: a deceptive and still mysterious benign lesion. *Human Pathology*. Vol. **82**, pg. 1–9, 2018, https://doi.org/10.1016/j.humpath.2018.07.021.

15. I. Gershon-Cohen, H. Ingleby. Roentgenography of fibroadenoma of the breast. *Radiology*. Vol. **59**, no. 1, pg. 77–87, 1952, https://doi.org/10.1148/59.1.77.

16. S. J. Parker, S. A. Harries. Phyllodes tumours. *Postgraduate Medical Journal*. Vol. **77**, no. 909, pg. 428–435, 2001, https://doi.org/10.1136/pmj.77.909.428.

17. R. A. Rodrigues, C. L. A. P. Azevedo, M. C. R. Djahjah, T. S. S. Pereira. Tubular adenoma of the breast: radiological and ultrasound findings. *Radiologia Brasileira*. Vol. **51**, no. 5, pg. 341–342, 2018, https://doi.org/10.1590/0100-3984.2017.0012.

18. S. R. Wellings. A hypothesis of the origin of human breast cancer from the terminal ductal lobular unit. *Pathology Research and Practice*. Vol. **166**, no. 4, pg. 515–535, 1980, https://doi.org/10.1016/S0344-0338(80)80248-2.

19. D. P. Winchester, J. M. Jeske, R. A. Goldschmidt. The diagnosis and management of ductal carcinoma in-situ of the breast. *CA: A Cancer Journal for Clinicians*. Vol. **50**, no. 3, pg. 184–200, 2000, https://doi.org/10.3322/canjclin.50.3.184.

20. N. Wilson, A. Ironside, A. Diana, O. Oikonomidou. Lobular Breast Cancer: A Review. *Frontiers in Oncology*. Vol. **10**, pg. 591399, 2021, https://doi.org/10.3389/fonc.2020.591399.

21. S. K. Pal, S. K. Lau, L. Kruper, U. Nwoye, C. Garberoglio, R. K. Gupta, B. Paz, L. Vora, E. Guzman, A. Artinyan, G. Somlo. Papillary carcinoma of the breast: an overview. *Breast Cancer Research and Treatment*. Vol. **122**, no. 3, pg. 637–645, 2010, https://doi.org/10.1007/s10549-010-0961-5.