Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Transfer Learning vs Training from Scratch for Breast Cancer Histology Image Classification

Janhvi Chauhan[1], Dhaval M. Modi[2]

[1, 2]Lecturer

[1, 2]Electronics & Communication Department, Govt. Polytechnic, Ahmedabad

Abstract

We evaluated the effectiveness of transfer learning compared to fully trained networks for histopathological image classification, using three pre-trained models: VGG16, VGG19, and ResNet50. Their performance was analyzed in the context of magnification-independent breast cancer classification. Additionally, we assessed how varying the training–testing data split impacts model performance. Among the tested configurations, the fine-tuned VGG16 model combined with a logistic regression classifier achieved the highest accuracy of 93.50%, an AUC of 96.00%, and an average precision score (APS) of 96.05% using 85%–15% train–test split. Future work may explore layer-wise fine-tuning and alternative weight initialization strategies to further enhance performance.

Keywords: Breast Cancer; Histopathological Images; Convolutional Neural Network; Full Training; Transfer Learning

1. Introduction

The pink ribbon has become a universal symbol in the medical community to raise awareness about breast cancer. Breast cancer remains one of the leading causes of mortality among women [1]. Several high-risk factors are associated with the development of this disease, including genetic mutations in the BRCA1 and BRCA2 genes, obesity, use of birth control pills, irregular menstrual cycles, and prolonged exposure to radiation therapy and estrogen [2–4]. These factors can trigger mutations at the cellular level, resulting in uncontrolled cell proliferation.

One of the earliest symptoms of breast cancer is breast soreness, which can become life-threatening if not detected at an early stage [38]. Additional initial symptoms may include skin irritation, redness, swelling, and pain, which can progress to more severe signs such as nipple erosion or unexpected watery discharge [3,4]. Therefore, early detection is crucial for enabling timely and effective treatment, ultimately improving the chances of survival [5].

Common techniques used for breast cancer screening and monitoring include mammography, magnetic resonance imaging (MRI), ultrasound, positron emission tomography (PET), thermography, and surgical biopsy [4,6–8]. However, interpreting the data generated from these methods is often complex due to overlapping clinical features among various cancer types, making accurate analysis challenging. The process of analyzing such data is not only time-consuming and labor-intensive but also critical for reaching a differential diagnosis.

To address these challenges, automation in diagnostic workflows has become essential to reduce the workload on radiologists and pathologists. In this context, machine learning has emerged as a powerful tool, offering dependable and intelligent solutions capable of automating several diagnostic tasks [9].

Among various machine learning techniques, deep neural networks (DNNs) have garnered significant attention due to their ability to automatically extract features and perform hierarchical representation learning [10–12]. The rise in computational capabilities has further fueled their widespread adoption [13]. A prominent variant of DNNs, the convolutional neural network (CNN), is extensively used in computer vision applications, primarily because of its weight-sharing and local connectivity properties [14–17]. These characteristics enable CNNs to function as localized filters that can detect consistent patterns across the entire image while requiring fewer trainable parameters.

CNNs excel at representation learning, gradually combining low-level features into complex high-level abstractions to ultimately classify images. However, training CNNs from scratch can be challenging, as it demands a large volume of labeled data to achieve high performance [18,19]. Additionally, powerful graphical processing units (GPUs) are necessary to expedite the training process, given the computational intensity involved in handling large datasets [18]. Moreover, training CNNs is often complex due to issues like convergence instability and overfitting, which necessitate careful tuning of hyperparameters to ensure balanced learning across all layers [20].

Transfer learning presents a viable alternative to full training. In this approach, a model trained on one task is fine-tuned for a different yet related task. Transfer learning can be implemented either as a baseline model or as a feature extractor [21–24]. When used as a baseline, the pre-trained network's parameters are adapted to the new task [21]. Alternatively, when acting as a feature extractor, the network generates features from input data which are then used to train a separate classifier [22].

This study explores a critical question in the context of breast cancer histopathological image classification: Which performs better— a fully trained CNN or a fine-tuned pre-trained model— for magnification-independent classification? To answer this, two primary experiments were conducted: (a) magnification-invariant classification of breast cancer histopathological images using the BreakHis dataset (https://web.inf.ufpr.br/vri/databases/breast-cancer/histopathological-database-breakhis/), and (b) an evaluation of model performance across three different training–testing data splits.

For each experiment, the performance of fine-tuned pre-trained models was compared against that of CNNs trained from scratch. This work was inspired by the study conducted by Tajbakhsh et al., who highlighted the lack of a comparative analysis between fully trained and fine-tuned CNNs on both histopathological and magnetic resonance imaging modalities [25]. Our goal is to extend their research by providing such a comparison specifically for histopathological imaging.

The primary objective of this paper is to evaluate the efficacy of transfer learning versus training from scratch in the classification of breast cancer using histopathological images. Additionally, the study aims to identify the most effective pre-trained model for this task. To the best of our knowledge, this is the first work to conduct such a comparative analysis within the domain of histopathological imaging.

2. Material and method used

In this study, three widely recognized pre-trained deep convolutional neural network (CNN) models—VGG16, VGG19, and ResNet50—are employed for both full training (from scratch) and transfer learning approaches. Due to the complex nature of classifying breast cancer histology images, deep architectures are essential to effectively capture and learn relevant features. The selected models

are particularly suited for this task because of their depth and proven performance across various challenging computer vision problems.

VGG16 and VGG19, for instance, gained significant recognition by securing top positions in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2014), ranking first in localization and second in classification. These models have also demonstrated strong generalization capabilities on other datasets such as Caltech-101, Caltech-256, and PASCAL VOC 2007 and 2012 [26,27]. Similarly, ResNet50, a residual network architecture, won the ILSVRC-2015 and even outperformed human accuracy in classification tasks. A key advantage of ResNet50 is its ease of training, as it learns residual mappings rather than direct features, facilitating better gradient flow in deep networks [28].

In this work, all three pre-trained CNNs are utilized as feature extractors. Specifically, the activations from the layers preceding the final fully connected layer are extracted and used to form feature vectors. These vectors are then used to train a new classifier—logistic regression (LR)—to perform the final classification. This approach differs from that of Tajbakhsh et al. [25], where fine-tuning was applied layer-wise using only a single pre-trained model, AlexNet. Furthermore, the current study also examines how varying training–testing data proportions influence the performance of both fully trained and fine-tuned models, a factor also considered in [25].

2.1. Dataset

In the medical imaging domain, datasets often comprise a limited number of annotated samples due to the complexity and cost of data acquisition. Accessing a large-scale, well-annotated dataset is critical for developing robust and generalizable models, as such datasets provide a standardized benchmark for model validation and comparison.

For this study, the publicly available BreakHis dataset is utilized. This dataset was developed through a collaboration with the Prognostics and Diagnostics (P&D) Laboratory in Parana, Brazil [14,35]. It includes a total of 7,909 breast cancer histopathology image samples collected from 82 patients and captured under four different magnification levels. The dataset is categorized into two main classes: benign and malignant, comprising 2,480 and 5,429 samples, respectively. A sample visualization of the dataset is shown in Figure 1.

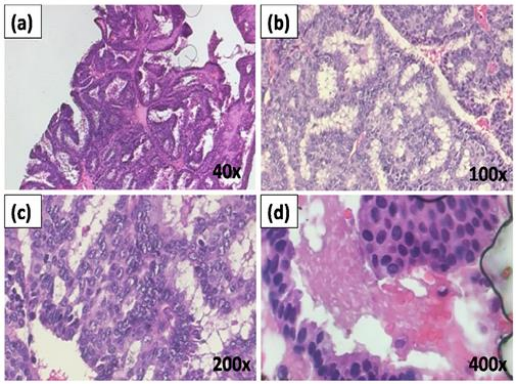

Figure 1. Breast Cancer Histopathological Images from BreakHis Dataset of a Patient Suffered from Papillary carcinoma (Malignant) with four magnification levels (a) 40x, (b) 100x (c) 200x and (d) 400x. [10].

2.2. Data augmentation

A major challenge in developing robust computer-aided diagnosis (CAD) systems lies in dealing with unbalanced and limited datasets. To address this limitation, data augmentation is commonly used in deep learning models to artificially expand the training data. Techniques such as flipping, cropping, scaling, rotation, interpolation, translation, and noise injection have been widely applied in prior studies [15,29,30]. However, augmentation methods effective for natural images may not be directly applicable to medical imaging, as many medical images follow a top-down interpretation paradigm, unlike the bottom-up approach used in natural images. Moreover, pixel intensity values often carry critical diagnostic information in medical images, making the selection of augmentation strategies even more sensitive.

Thus, it is essential to tailor augmentation techniques based on the specific characteristics of the dataset. In the case of histopathological images, which inherently exhibit rotation and reflection symmetry [31], using inappropriate augmentation may inadvertently discard crucial discriminative features. Therefore, in this study, only rotation is employed as the augmentation technique for both training-from-scratch and transfer learning approaches. Images are rotated around their center at three angles: 90°, 180°, and 270°. In addition to enriching the dataset, this augmentation strategy also helps mitigate overfitting—a common challenge in training machine learning models on limited data [32].

2.3. Magnification independent classification

The magnification factor plays a vital role in the interpretation and analysis of histological images. It alters the size of visualized structures, enabling more comfortable and detailed observation [33]. While histological images contain a wide range of tissue types, analyzing these tissues becomes increasingly difficult at lower magnification levels due to reduced clarity and detail. Additionally, capturing images at varying magnification levels introduces variability in background textures and image composition, making it challenging for automated computer-aided diagnosis (CAD) systems to consistently extract discriminative features across different magnifications.

To address this, many previous studies have restricted their classification tasks to a single magnification level, thereby minimizing background variation [34,35]. Other works have considered multi-magnification datasets but implemented separate classifiers tailored to each magnification level [36,37]. However, such magnification-dependent approaches require multiple stages of training and prior knowledge of the magnification level, making them less scalable. Furthermore, the introduction of images with previously unseen magnification factors often degrades model performance, highlighting a key limitation of these strategies.

Therefore, it is imperative to develop a magnification-independent CAD system capable of generalizing across varying magnification levels and adapting to new ones without significant loss in accuracy or diagnostic reliability.

3. Results & Discussion

In this study, we present consistent and reliable results for the application of breast cancer classification using histopathological imaging. A balanced dataset was employed for both full training and fine-tuning of convolutional neural networks (CNNs). To achieve dataset balance, the class with a greater number of samples (malignant) was downsampled to match the number of samples in the other class (benign). All experiments were conducted on a system with the following specifications: Intel(R)

Core(TM) i7-7500U CPU @ 2.90 GHz, NVIDIA GeForce 940MX GPU, Windows 10 OS, 8 GB RAM, and implemented using TensorFlow and Keras libraries.

To evaluate classification performance, the dataset was divided into training and testing subsets using three different ratios: 85%–15%, 80%–20%, and 75%–25%. This partitioning strategy is a standard practice in neural network experiments to assess model generalization. Each configuration was tested for both fully-trained and fine-tuned networks, with training times for each experiment ranging from 1 to 2 hours.

Performance metrics, including precision, recall, and F1-score, were calculated separately for each class, followed by averaging to facilitate comparison. Additionally, Receiver Operating Characteristic (ROC) analysis and the Area Under the Curve (AUC), along with the Average Precision Score (APS), were used to comprehensively evaluate classification performance.

For full training, the CNNs were initialized with random weights and trained on the BreakHis dataset from scratch, using only the model architecture of pre-trained networks. In contrast, the transfer learning approach retained the pre-trained weights, assuming the networks had already learned robust feature representations.

Table 1 Performance Analysis for Histopathological Image Classification using Fine-tuned Pre-trained Network (VGG16, VGG19 and ResNet50)

Classifier	Training-Testing Data Splitting	Class Type	Precision	F1 Score	Recall	Accuracy	AUC	APS
VGG16	85%-15%	B M Avg/Total	0.92 0.92 0.92	0.92 0.92 0.92	0.92 0.94 0.93	93.50%	96.00%	96.05%
	80%-20%	B M Avg/Total	0.93 0.94 0.93	0.93 0.94 0.93	0.93 0.92 0.92	93.40%	93.95%	95.20%
	75%-25%	B M Avg/Total	0.93 0.93 0.93	0.93 0.92 0.92	0.92 0.91 0.92	92.20%	93.49%	94.34%
VGG19	85%-15%	B M Avg/Total	0.88 0.93 0.90	0.91 0.90 0.90	0.93 0.90 0.91	90.00%	90.45%	91.27%
	80%-20%	B M Avg/Total	0.89 0.90 0.90	0.90 0.91 0.90	0.91 0.91 0.91	89.50%	90.45%	91.13%
	75%-25%	B M Avg/Total	0.90 0.90 0.90	0.91 0.90 0.90	0.91 0.91 0.91	90.40%	92.20%	90.38%
ResNet50	85%-15%	B M Avg/Total	0.77 0.81 0.79	0.79 0.80 0.79	0.81 0.80 0.80	79.40%	79.39%	82.03%
	80%-20%	B	0.80	0.79	0.80	78.90%	79.23%	80.56%

		M Avg/Total	0.81 0.79	0.80 0.79	0.80 0.80			
	75%-25%	B M Avg/Total	0.78 0.80 0.79	0.79 0.78 0.79	0.81 0.80 0.80	78.73%	79.12%	79.09%

Table 2 Performance Analysis for Histopathological Image Classification using Full-trained Network (VGG16, VGG19 and ResNet50)

Classifier	Training-Testing Data Splitting	Class Type	Precision	Recall	F1 Score	Accuracy	AUC	APS
VGG16	85%-15%	B M Avg/Total	0.63 0.66 0.64	0.64 0.64 0.64	0.64 0.65 0.64	64.40%	75.65%	76.80%
	80%-20%	B M Avg/Total	0.63 0.63 0.63	0.63 0.64 0.63	0.63 0.63 0.63	61.20%	75.95%	75.22%
	75%-25%	B M Avg/Total	0.66 0.62 0.64	0.60 0.68 0.64	0.63 0.65 0.64	62.73%	75.49%	73.29%
VGG19	85%-15%	B M Avg/Total	0.50 0.60 0.55	0.87 0.19 0.52	0.64 0.29 0.46	53.40%	75.85%	72.27%
	80%-20%	B M Avg/Total	0.57 0.65 0.61	0.76 0.43 0.60	0.65 0.52 0.59	55.40%	74.76%	73.13%
	75%-25%	B M Avg/Total	0.81 0.50 0.66	0.02 0.99 0.50	0.04 0.66 0.35	53.21%	75.14%	75.38%
ResNet50	85%-15%	B M Avg/Total	0.77 0.80 0.79	0.78 0.78 0.79	0.79 0.79 0.79	74.40%	73.39%	73.03%
	80%-20%	B M Avg/Total	0.79 0.80 0.79	0.80 0.78 0.79	0.79 0.79 0.79	73.90%	74.23%	74.56%
	75%-25%	B M Avg/Total	0.70 0.81 0.80	0.82 0.78 0.80	0.80 0.79 0.80	78.53%	75.12%	75.09%

Tables 1 and 2 summarize the classification results of both transfer learning and full training for the VGG16, VGG19, and ResNet50 models. The models were evaluated on their ability to classify histopathological breast images into benign (B) and malignant (M) categories. From Table 1, it is evident that the fine-tuned VGG16 network significantly outperformed ResNet50, while VGG16 and VGG19

yielded comparable performance. ResNet50, when used in transfer learning, exhibited poor generalization on the BreakHis dataset, likely due to overfitting—a result of its high model capacity. This issue could potentially be mitigated by freezing additional layers to reduce effective model complexity, a solution proposed in [25] but not explored in this work due to space constraints. We plan to address this in a future extended version of the paper.

In contrast, for full training, ResNet50 demonstrated superior performance over VGG16 and VGG19. Table 2 indicates that both VGG16 and VGG19 were biased toward one class, as reflected in the imbalanced recall values. ResNet50, however, showed balanced sensitivity across both classes, making it more effective in the full-training scenario.

To analyze the effect of training data size, the models were tested with three different training–testing splits. The impact of dataset size on CNN performance was assessed using ROC curves and AUC values, as shown in Figures 2–4. For the 85%–15% split (Figure 2), pre-trained VGG16 (AUC: 96.00%) and VGG19 (AUC: 90.45%) outperformed their fully-trained counterparts—VGG16 (AUC: 75.65%) and VGG19 (AUC: 75.85%). While the pre-trained ResNet50 (AUC: 79.39%) was slightly outperformed by its fully-trained version (AUC: 73.39%), the margin was minimal. Similar trends were observed for the 80%–20% and 75%–25% splits (Figures 3 and 4).

Across these experiments, the performance of the fine-tuned VGG16, VGG19, and ResNet50 networks remained relatively stable despite reductions in training data size. All three models performed well with 85% training data, but full training yielded mixed results. For instance, VGG19 performed best at the 80%–20% split, while ResNet50 and VGG16 showed consistent performance across all splits. The deviation in VGG19's performance may be attributed to its class sensitivity, which varied depending on the data split—favoring benign samples in the 85%–15% split and malignant samples in the 75%–25% split.

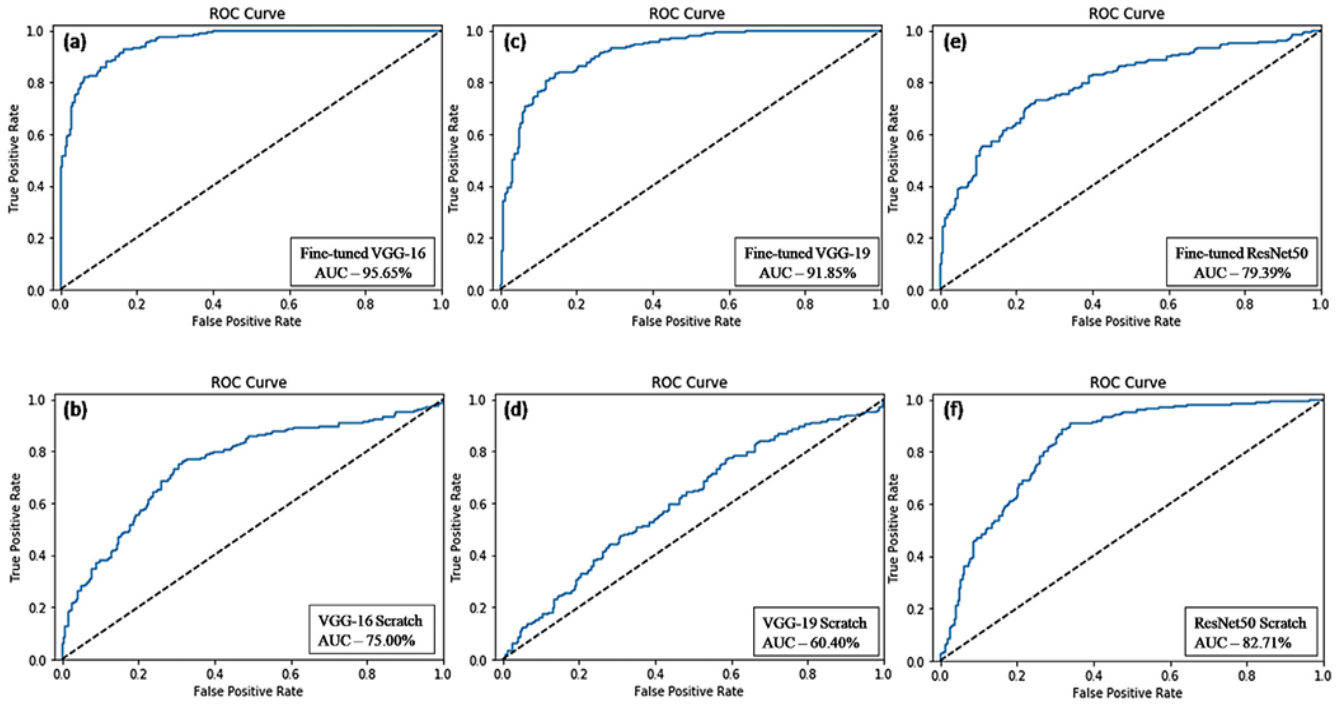

Figure 2ROC analysis for breast cancer classification with 85%–15% trainingandtestingsetsplitting (a)Fine-tuned pre-trained VGG16 (b)Fine-tuned pre-trained VGG19 (c)Fine-tuned pre-trained ResNet 50 (d)Fully-trainedVGG16 (e)Fully-trained VGG19 and (f)Fully-trained ResNet50

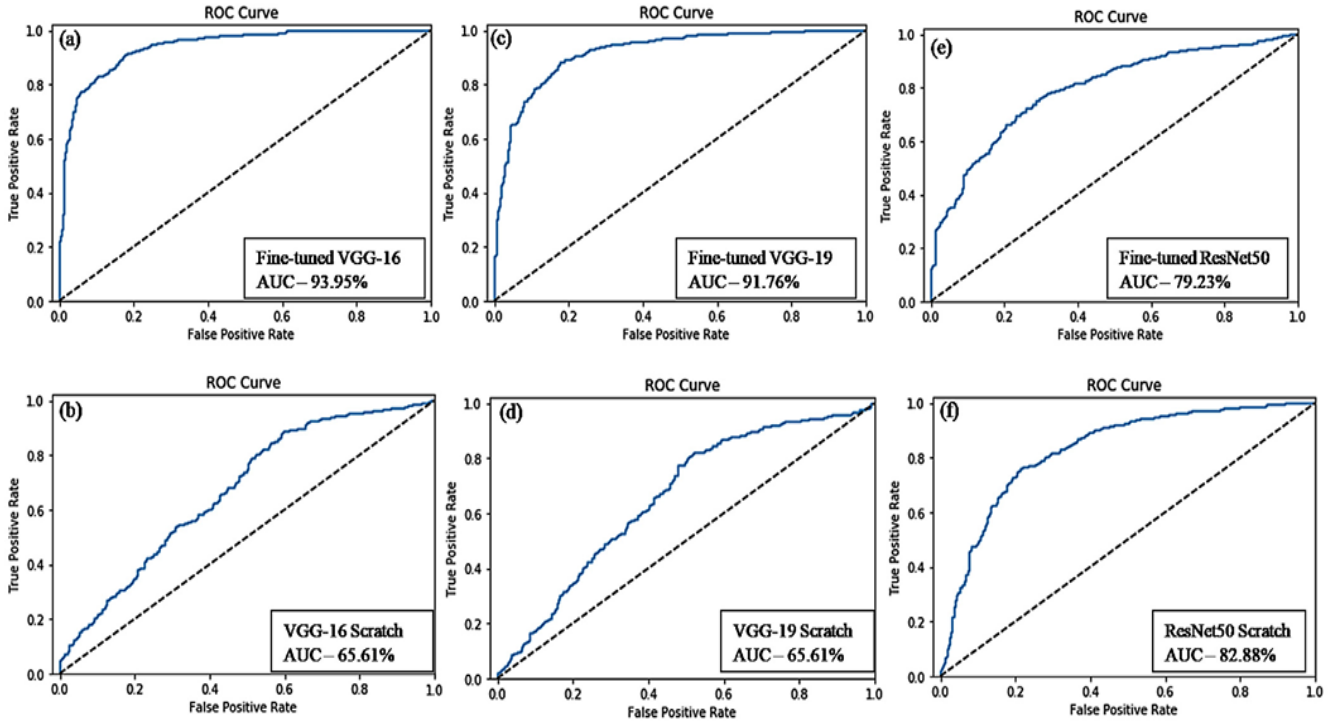

Figure 3 ROC analysis for breast cancer classification with 80%–20% training and testing set splitting (a)Fine-tuned pre-trained VGG16 (b)Fine-tuned pre-trained VGG19 (c)Fine-tuned pre-trained ResNet50 (d)Fully-trained VGG16 (e)Fully-trained VGG19 and (f)Fully-trainedResNet50

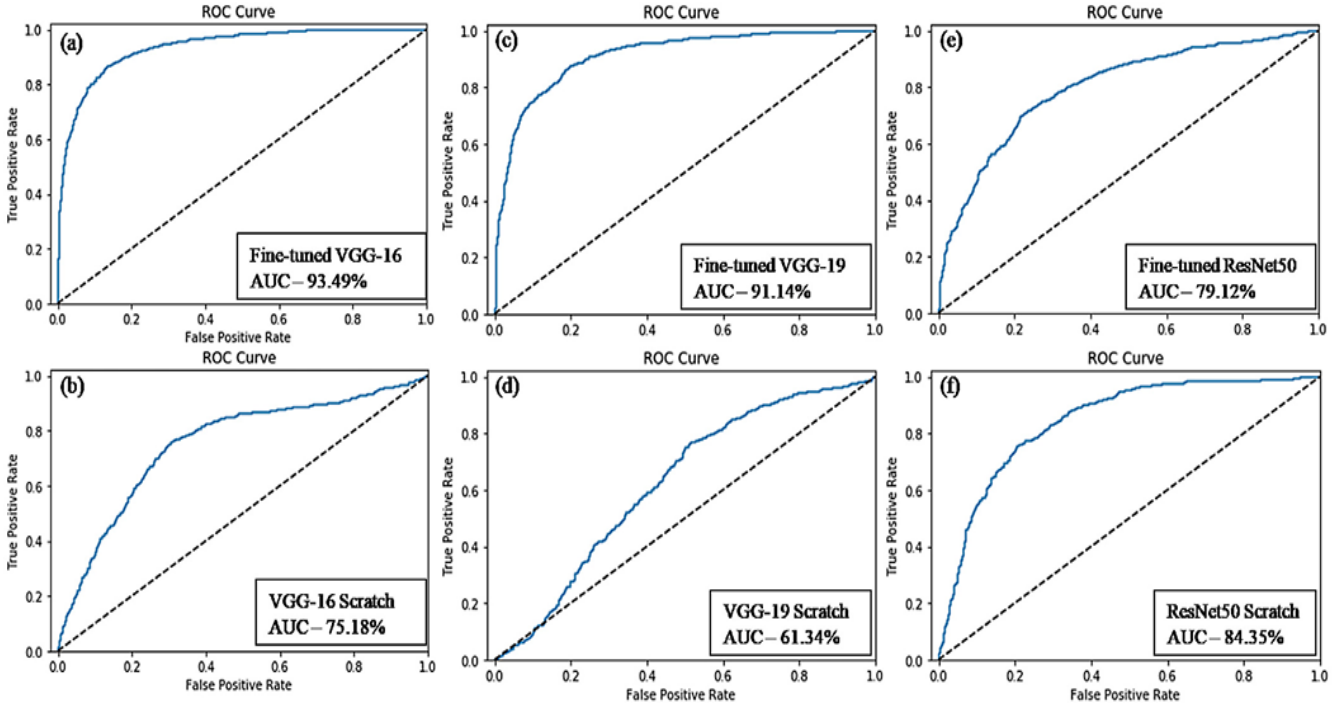

Figure: 4 ROC analysis for breast cancer classification with 75%–25% training and testingset splitting (a) Fine-tuned pre-trained VGG16 (b)Fine-tuned pre-trained VGG19 (c)Fine-tuned pre-trained ResNet50 (d)Fully-trained VGG16 (e)Fully-trained VGG19 and (f)Fully-trained ResNet50

In conclusion, this study demonstrates that transfer learning offers a highly effective approach for breast cancer classification using histopathological images, particularly when training data is limited. It consistently outperforms full training across various data splits, validating its robustness and potential for real-world medical applications.

4. Conclusions and future directions

This study explores the feasibility of transferring learned knowledge from natural images to histopathological images by leveraging three pre-trained convolutional neural networks—VGG16, VGG19, and ResNet50—through both fine-tuning and full-training strategies. In the transfer learning approach, these pre-trained models were used as feature extractors, and the resulting features were used to train a logistic regression classifier.

Key findings from this work include:

- **Superior performance with transfer learning:** Among all the models and training strategies, the fine-tuned VGG16 combined with a logistic regression classifier achieved the highest performance, recording an accuracy of **93.50%**, an **AUC of 96.00%**, and an **Average Precision Score (APS) of 96.05%** for the 85%–15% training–testing split.
- **Robustness to training data size:** The fine-tuned models demonstrated greater resilience to variations in training data size compared to fully-trained models. Their performance remained relatively stable even when the amount of training data was reduced.
- **Class bias impacts model effectiveness:** A model biased toward a specific class significantly compromises classification performance. Therefore, achieving balanced sensitivity across all classes is critical for reliable diagnostics.
- **Network capacity influences generalization:** The capacity of the model plays a crucial role in its effectiveness. Excessively large networks tend to overfit, while overly simplistic architectures may underfit. Optimal network complexity should be tailored to the specific application.

For future work, several enhancements can be considered to improve model performance further. These include layer-wise fine-tuning, the use of larger and more diverse datasets, advanced data augmentation techniques—such as conditional generative adversarial networks (GANs) and deep photo style transfer—and improved weight initialization methods like Xavier, He, MSRA, or Gaussian distributions during full training.

Conflict of interest: The authors declare that they have no conflict of interests in this paper.

References

1. R.L. Siegel, K.D. Miller, S.A. Fedewa, D.J. Ahnen, R.G. Meester, A. Barzi, A. Jemal, "Colorectal cancer statistics, 2017," *CA Cancer J. Clin.*, vol. 67, no. 3, pp. 177–193, 2017.

2. D. Ford, D.F. Easton, M. Stratton, et al., "Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families," *Am. J. Hum. Genet.*, vol. 62, pp. 676–689, 1998.
3. M. Kallenberg, K. Petersen, M. Nielsen, et al., "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.
4. K. Ganesan, U.R. Acharya, C.K. Chua, et al., "Computer-aided breast cancer detection using mammograms: A review," *IEEE Trans. Biomed. Eng.*, vol. 6, pp. 77–98, 2012.
5. Editorial, "Early detection: a long road ahead," *Nat. Rev.*, vol. 18, p. 401, 2018.
6. Y. Zheng, "Breast cancer detection with Gabor features from digital mammograms," *Algorithms*, vol. 3, no. 1, pp. 44–62, 2010.
7. B.D. Fleet, J. Yan, D.B. Knoester, et al., "Breast cancer detection using Haralick features of images reconstructed from ultra-wideband microwave scans," *Transl. Res. Med. Imaging*, pp. 9–16, 2014.
8. M. Tan, B. Zheng, J.K. Leader, D. Gur, "Association between changes in mammographic image features and risk for near-term breast cancer development," *IEEE Trans. Med. Imaging*, vol. 35, no. 7, pp. 1719–1728, 2016.
9. D. Shen, G. Wu, H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.
10. Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
11. R. Mehra, S. Shallu, S. Kumar, "An insight into the convolutional neural network for the analysis of medical images," in *IEEE Int. Conf. on Nanotechnology for Instrumentation and Measurement Workshop (NANOfIM)*, 2017.
12. E. Hoffer, N. Ailon, "Deep metric learning using triplet network," in *Int. Workshop on Similarity-Based Pattern Recognition*, pp. 84–92, 2015.
13. A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, 2017.
14. F.A. Spanhol, P.R. Cavalin, L.S. Oliveira, et al., "Breast cancer histopathological image classification using convolutional neural networks," in *Int. Joint Conf. on Neural Networks*, pp. 2560–2567, 2016.
15. Z. Han, B. Wei, Y. Zheng, et al., "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, pp. 1–10, 2017.
16. M. Anthimopoulos, S. Christodoulidis, L. Ebner, et al., "Lung pattern classification for interstitial lung disease using a deep convolutional neural network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
17. M. Aubreville, C. Knipfer, N. Oetter, et al., "Automatic classification of cancerous tissue in laser endomicroscopy images of the oral cavity using deep learning," *Sci. Rep.*, vol. 7, pp. 1–10, 2017.
18. M. Anthimopoulos, S. Christodoulidis, L. Ebner, et al., "Lung pattern classification for interstitial lung disease using a deep convolutional neural network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

19. S. Hoo-Chang, H.R. Roth, M. Gao, et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1297, 2016.
20. O. Irsoy, E. Alpaydın, "Continuously constructive deep neural networks," *arXiv preprint*, arXiv:1804.02491, 2018.
21. A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 512–519, 2014.
22. H. Azizpour, A.S. Razavian, J. Sullivan, et al., "From generic to specific deep representations for visual recognition," *arXiv preprint*, arXiv:1406.5774, 2014.
23. K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Constr. Build. Mater.*, vol. 157, pp. 322–330, 2017.
24. O. Penatti, K. Nogueira, J. Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 44–51, 2015.
25. N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
26. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. on Learning Representations*, pp. 1–14, 2015.
27. P. Wang, G.W. Cottrell, "Basic level categorization facilitates visual object recognition," in *Int. Conf. on Learning Representations*, pp. 1–13, 2016.
28. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," *arXiv preprint*, arXiv:1512.03385, 2015.
29. J. Shijie, W. Ping, J. Peiyi, H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *Chinese Automation Congress (CAC)*, pp. 4165–4170, 2017.
30. J. Wang, L. Perez, "The effectiveness of data augmentation in image classification using deep learning," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1–8, 2017
31. B.S. Veeling, J. Linmans, J. Winkens, T. Cohen, M. Welling"Deep neural network-based recognition of plant diseases by leaf image classification," *Comput. Intell. Neurosci.*
32. T.L. Sellaro, R. Filkins, C. Hoffman, et al., "Relationship between magnification and resolution in digital pathology systems," *J. Pathol. Inform.*, vol. 4, 2013.
33. M. Peikari, M.J. Gangeh, J. Zubovits, et al., "Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture-based approach," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 307–315, 2016.
34. W. Zhou, C. Wu, D. Chen, Y. Yi, and W. Du, “Automatic Microaneurysm Detection Using the Sparse Principal Component Analysis-Based Unsupervised Classification Method,” IEEE Access, vol. 5, pp. 2563–2572, 2017.
35. C. Loukas, S. Kostopoulos, A. Tanoglidi, et al., "Breast cancer characterization based on image classification of tissue sections visualized under low magnification," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–10, 2013.

36. F. Spanhol, L. Oliveria, C. Petitjean, L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2015.
37. F. Keskin, A. Suhre, K. Kose, et al., "Image classification of human carcinoma cells using complex wavelet-based covariance descriptors," *PLoS One*, vol. 8, no. 1, 2013.
https://www.nationalbreastcancer.org/breast-pain
38. A. Sopharak, B. Uyyanonvara, and S. Barman, “Fine Microaneurysm Detection from Non-dilated Diabetic Retinopathy Retinal Images Using a Hybrid Approach,” vol. II, pp. 6–9, 2012.
39. K. M. Adal, D. Sidibé, S. Ali, E. Chaum, T. P. Karnowski, and F. Mériaudeau, “Automated detection of microaneurysms using scale adapted blob analysis and semi-supervised learning,” Comput. Methods Programs Biomed., vol. 114, no. 1, pp. 1–10, 2014.
40. R. A. Welikala et al., “Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification,” Comput. Methods Programs Biomed., vol. 114, no. 3, pp. 247–261, 2014.
41. B. Antal and A. Hajdu, “Improving microaneurysm detection in color fundus images by using context-aware approaches,” Comput. Med. Imaging Graph., vol. 37, no. 5–6, pp. 403–408, 2013.