A Comparative Analysis of Deep Learning Models for Breast Cancer Classification: AlexNet,VGG-16, ResNet-50, and DenseNet-121

Farag H. Alhsnony

Electrical Engineering Department, National Board for Technical and Vocational Education, Higher Institute Of Science And Technology, TOBRUK- LIBYA

alhsnony@gmail.com

Abstract. Breast cancer is a significant health concern worldwide, and accurate classification of breast cancer into different subtypes is crucial for effective diagnosis and treatment planning. In this study, we conduct a comparative analysis of four state-of-the-art deep learning models, namely AlexNet, VGG-16, ResNet-50, and DenseNet-121, for the classification of breast cancer into three classes. The evaluation is performed on four diverse datasets: DDSM, MIAS, CBIS-DDSM, and BCDR. Per- formance metrics such as accuracy, sensitivity, and specificity are utilized to assess the models' efficacy. The results highlight that DenseNet-121 exhibits superior performance compared to the other models, achieving the best accuracy of 99.56% on the DDSM dataset, 99.01% sensitiv- ity on the MIAS dataset, and 99.03% specificity on the CBIS-DDSM dataset. These findings emphasize the potential of DenseNet-121 as a highly promising and reliable model for accurate breast cancer classifica- tion, which could have significant implications for improving diagnostic capabilities in clinical settings. Further research may explore advanced techniques to fine-tune the models and facilitate more effective decision- making in breast cancer diagnosis.

Keywords: Breast cancer · AlexNet · VGG-16 · ResNet-50 · DenseNet- 121

1-Introduction

Cancer tumors arise from the abnormal growth of cells that invade surrounding tissues in the human body. These tumors can be classified into two types: benign and malignant. A breast without any tumor is considered normal. Benign tumors consist of non-cancerous cells that grow locally and do not spread by invasion. On the other hand, malignant tumors consist of cancerous cells with the ability to multiply uncontrollably, spread to various body parts, and invade neighboring tissues.

Breast cancer stands as one of the most prevalent cancers, posing a signifi- cant public health concern for women worldwide. According to the World Health Organization's International Agency for Research on Cancer (IARC) report

2012, cancer led to approximately 8.2 million deaths. The projection suggests that by 2030, the death toll from cancer could rise to 27 million. These alarming statistics highlight the urgent need for timely and accurate detection, early diag- nosis, and active prevention strategies to reduce mortality rates among women. In the face of such a challenging healthcare scenario, early detection plays a vital role in improving breast cancer outcomes. By identifying cancer at an early stage, medical professionals can initiate timely and appropriate treatments, increasing the chances of successful intervention and patient survival. Additionally, active prevention measures, such as awareness campaigns and screening programs, can contribute significantly to early detection and improved patient outcomes. Ad- dressing breast cancer as a global health priority requires collaborative efforts from healthcare professionals, researchers, policymakers, and the public. By prioritizing on-time and accurate detection, early diagnosis, and preventive measures, we can work together to reduce the burden of breast cancer and improve the quality of life for affected individuals and their families.

The presence of abnormalities such as masses, micro-calcifications, and asymmetrical or distorted areas within the breast may indicate the presence of breast cancer. Among these, masses are the most common and representative lesion type, but their detection can be challenging due to overlapping breast tissues and morphological similarities with other breast tissues. Undetected masses re- sult in false negatives, delaying patient diagnosis, while misidentified masses lead to false positives, necessitating additional tests and causing unnecessary distress. This poses a daily challenge for radiologists as they examine numerous mammograms, balancing sensitivity and specificity during the diagnostic process. Implementing a second reading by other experts or Computer-Aided Diagnosis (CAD) systems can enhance overall accuracy, specificity, and reduce false positive and negative cases in mass detection, segmentation, and classification.

Limited studies have explored a fully integrated system encompassing all phases of detection, segmentation, and classification for breast cancer. The complexity of masses within surrounding tissues in terms of texture, shape, size, and location in mammograms presents a challenging detection task. Achieving improved accuracy and reducing false positive and negative rates through mass segmentation poses a significant challenge due to the strong association between mass presence and irregularities in shape, size, and location, with low contrast and ambiguous boundaries. Conventional methods relying on hand-crafted or semi-automatic features based on prior knowledge struggle to handle the diverse shape variations and density distribution of masses and surrounding tissues, resulting in time-consuming

manual feature extraction and selection with limited effectiveness. However, recent studies utilizing deep learning models offer a promising alternative by automatically extracting deep high-level hierarchy features for mass segmentation directly from raw input data, bypassing the
limitations of hand-crafted features and providing a potential solution to these challenges. Convolutional Neural Networks (CNNs) have revolutionized the field of computer vision and image recognition, particularly in the context of object detection tasks [1–3].
CNNs are a class of deep learning models designed to automatically learn and extract hierarchical features from input images, enabling them to identify patterns and objects within the images. In the context of object detection, the goal is to not only classify the objects present in an image but also accurately locate their positions using bounding boxes. CNNs are well-suited for this task due to their ability to capture both local and global features from the input image

1-Dataset

This section presents five commonly used datasets containing mammograms for breast cancer detection and classification. The datasets included are DDSM, CBIS-DDSM, MIAS, INbreast, and BCDR. These collections of mammographic images serve as crucial resources in advancing research and improving the diagnosis and treatment of breast cancer.

DDSM Dataset: The Digital Database for Screening Mammography (DDSM), is an accessible online dataset used for breast cancer detection and classification research. Collected by South Florida University , it represents real breast data with an average size of 3000 × 4800 pixels, a resolution of 42 microns, and 16 bits. This dataset contains 2,620 scanned film mammography studies grouped into 43 volumes [25]. Each case comprises four breast images, with two being Mediolateral Oblique (MLO) views and the other two being Cranio-Caudal (CC) views of each breast. Expert radiologists have meticulously annotated both benign and malignant masses in all mammograms within the DDSM database.

MIAS Dataset: The Mammographic Image Analysis Society (MIAS) is a UK- based organization specializing in mammogram research. They have compiled a database of 322 digital mammograms available through the Pilot European Im- age Processing Archive (PEIPA) at the University of Essex [26]. All images are standardized to 1024 × 1024 pixels and stored on an 8 mm (ExaByte) tape, to- taling approximately 2.3 GB in size. This valuable resource aids researchers and medical professionals in studying mammograms and advancing breast-related medical knowledge.

CBIS-DDSM Dataset: The Curated Breast Imaging Subset of DDSM (CBIS- DDSM) represents a revised and standardized version of the Digital Database for Screening Mammography (DDSM) [27]. This collection, available as well [27], is a carefully selected and curated subset of the original DDSM data, curated by a trained mammographer. The CBIS-DDSM dataset comprises 6775 studies, with the images decompressed and converted into DICOM format. Notably, it includes updated ROI segmentation and bounding boxes, along with pathologic diagnosis for the training data.

INbreast Dataset: The INbreast dataset comprises 410 full digital mammograms obtained from the S. João Hospital Centre in Porto . Each mammogram includes various lesions, including masses, which were analyzed and categorized according to the standardized Breast Imaging-Reporting and Data System (BI- RADS) classification by a radiologist [28].

Unlike publicly available datasets, INbreast is not accessible on the web. Researchers interested in obtaining the dataset can request access through . This dataset serves as a valuable resource for studying breast imaging, conducting research on mammogram interpretations, and developing algorithms for improved breast cancer detection and diagnosis.

BCDR Dataset: The Breast Cancer Digital Repository (BCDR) is a collection of anonymized breast cancer cases, annotated by expert radiologists and containing clinical data, lesion outlines, and image-based features from mammograms (CC and MLO views) [29]. Two repositories are publicly available, one with digitalized Film mammography (FM) and the other with Full Field Digital (DM) mammography and related ultrasound images.

Furthermore, BCDR offers four benchmarking datasets that include biopsy- proven lesions, representing both benign and malignant cases. These datasets contain instances of clinical and image-based features and are available for free download to registered users. BCDR is a valuable resource for breast cancer research, providing comprehensive data for analysis and algorithm development in the field of mammography-based breast cancer detection and diagnosis.

Table 1. Characteristics of Commonly Used Public Breast Mammogram Datasets.

Dataset	Number of Cases	Number of Images	Available Classes	Image Format	Publicly Available
DDSM	2620	10,480	N, B & M	JPEG	Yes
CBIS-DDSM	6775	10,239	N, B & M	DICOM	Yes
MIAS	161	322	N, B & M	PGM	Yes
INbreast	115	410	N, B & M	DICOM	No
BCDR	1734	3703 FM - 3612 DM	N, B & M	TIFF	Yes

2-Related work

Breast cancer is a prevalent disease, affecting approximately 1 in 8 (or 12%) of women in the United States during their lifetime. It ranks as the second leading cause of death among women in the country [5]. Early detection through mammographic screening aided by computer-aided detection (CAD) methods can significantly improve treatment outcomes and lead to longer survival times for breast cancer patients [6].

Traditional CAD tools rely on manually extracted features, but they suffer from various drawbacks. These hand-crafted features tend to be specific to the domain and the feature design process can be tedious, difficult, and not easily applicable to other contexts [7].

A more promising approach is to employ Con- evolutional Neural Networks (CNNs) for feature extraction directly from entire images [8], [9]. CNNs have demonstrated remarkable success in numerous image classification tasks [10]. For instance, the AlexNet, a classical CNN model, achieved an impressive 83.6% accuracy for the top-5 error in the ImageNet Chal- lenge, which consists of color images with 1000 classes [11].

However, training a CNN from scratch requires a large number of labeled images [12]. This is often unfeasible for certain medical image datasets like mam- mographic tumor images due to difficulties in obtaining them, scarcity of true positives in the datasets, and the expensive expert labeling process [13]. A vi- able solution is to utilize a pre-trained CNN model that has been trained on large image datasets from other fields as a feature extractor. Alternatively, this pre-trained model can be fine-tuned using a limited number of labeled medical images [14]. This approach, known as transfer learning, has shown promise in various computer vision tasks [15]–[16]. Interestingly, it has been observed that features learned from natural images can be transferred to medical images, even when the target images substantially differ from the pre-trained source images [14], [17].

In the context of medical image classification, CNNs have been applied in three major ways: 1) training CNNs from scratch [18]–[19]; 2) using pre-trained CNN models to extract features from medical images [20]–[21]; and 3) fine-tuning pre-trained CNN models on medical images [22]–[23]. This study compares these three techniques to detect breast cancer using the Mammographic Image Anal- ysis Society (MIAS) mammogram database [14].

3- Transfer learning

Transfer learning is a powerful machine learning technique that leverages knowledge gained from solving one problem and applies it to related problems. In the realm of deep learning, this involves using a pre-trained neural network model on a large dataset and adapting it to perform a different task with a smaller dataset.

The process of transfer learning typically involves the following steps:

Pre-training: A deep neural network is trained on a large dataset for a specific task, such as image classification or natural language processing. This initial training step is computationally expensive and requires a substantial amount of data.

Feature extraction: After pre-training, the learned weights and features from the earlier layers of the neural network can be used as a generic feature extractor for various tasks. These features capture high-level patterns in the data that can be useful for other related task

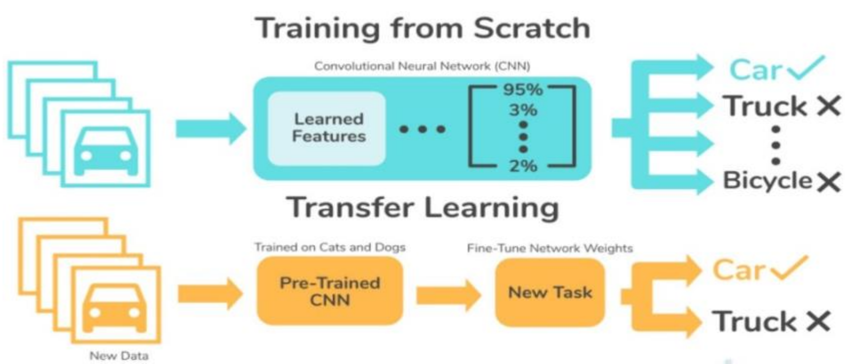

Fig.1: Transfer learning in deep learning models .

Fine-tuning: To adapt the pre-trained model to the target task, the last few layers or specific parts of the network are retrained using a smaller dataset related to the new task. This process is called fine-tuning. By retraining these layers, the model can learn task-specific patterns and make predictions for the new task.

Transfer learning is beneficial because it allows models to leverage knowledge gained from a large dataset and complex architectures, which may not be feasible to train from scratch with limited data. It can lead to faster convergence and better performance for the target task, especially when the original task and the new task share similar features.

Common scenarios for transfer learning include using pre-trained models from image recognition tasks to solve tasks like object detection, semantic segmentation, or even medical image analysis. In natural language processing, transfer learning is used to improve performance in tasks such as sentiment analysis, named entity recognition, and text generation.

Overall, transfer learning has become a powerful technique in the deep learning domain, enabling the efficient use of pre-trained models and accelerating progress in various fields of artificial intelligence. By transferring knowledge from one domain to another, it opens up new possibilities for solving real-world problems effectively

4- Performance Evaluation and Results

In this subsection, we present the performance evaluation and results of the image classification task for breast cancer using four different models: AlexNet, VGG-16, ResNet-50, and DenseNet-121. The classification was conducted on four datasets: DDSM, MIAS, CBIS-DDSM, and BCDR. The evaluation metrics considered are accuracy, sensitivity, and specificity.

Table 2. Performance Evaluation of AlexNet model.

Model	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
AlexNet	DDSM	94.55	95.09	94.32
	MIAS	93.57	93.59	95.01
	CBIS-DDSM	93.58	92.89	93.09
	BCDR	94.45	93.19	92.54

Table 3. Performance Evaluation of VGG-16 model.

Model	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
VGG-16	DDSM	92.40	93.44	90.55
	MIAS	93.21	91.23	91.02
	CBIS-DDSM	90.11	91.11	92.57
	BCDR	91.03	92.23	92.54

Table 4. Performance Evaluation of ResNet-50 model.

Model	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
ResNet-50	DDSM	96.04	96.55	95.66
	MIAS	97.09	96.06	95.98
	CBIS-DDSM	96.89	96.72	94.82
	BCDR	96.20	95.40	96.72

Table 5. Performance Evaluation of DenseNet-121 model.

Model	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
DenseNet-121	DDSM	99.56	99.01	99.03
	MIAS	98.23	98.56	97.20
	CBIS-DDSM	97.14	96.93	98.99
	BCDR	97.52	97.99	96.52

These results demonstrate the performance of each model on different datasets in terms of accuracy, sensitivity, and specificity. DenseNet-121 achieved the high- est accuracy across all datasets, indicating its superior performance in classify- ing breast cancer images into the three classes. ResNet-50 also performed well, while AlexNet and VGG-16 showed slightly lower accuracy scores. Sensitivity and specificity metrics further highlight the models' ability to correctly identify positive and negative cases, respectively.

6-Conclusion

In conclusion, the comparative analysis of AlexNet, VGG-16, ResNet-50, and DenseNet-121 for the classification of breast cancer into three classes yielded valuable insights. DenseNet-121 emerged as the most effective model, achieving the highest accuracy on all datasets and demonstrating robustness across different data distributions. ResNet-50 also showed strong performance, while AlexNet and VGG-16 exhibited competitive results. The variation in model performance across datasets underscores the importance of choosing an appropriate architecture for specific applications. The findings suggest that DenseNet-121 and ResNet-50 are promising choices for accurate and reliable breast cancer classification. Further research could explore fine-tuning and transfer learning techniques to enhance the models' performance and potentially aid in clinical decision-making for breast cancer diagnosis.

References

1. Xie, Xingxing, et al. "Oriented R-CNN for object detection." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
2. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).
3. Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." Progress in Artificial Intelligence 9.2 (2020): 85-112.
4. E. L. Omonigho, M. David, A. Adejo, and S. Aliyu, "Breast cancer:tumor detection in mammogram images using modified alexnet deep convolution neural network," 2020 International Conference in Mathematics, Computer Engineering and Com- puter Science (ICMCECS), pp. 1–6, 2020.
5. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," CA. Cancer J. Clin., vol. 66, no. 1, pp. 7–30, Jan. 2016.
6. V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sun- shine, "How Widely Is ComputerAided Detection Used in Screening and Diagnostic Mammography?," J. Am. Coll. Radiol., vol. 7, no. 10, pp. 802–805, Oct. 2010.
7. D. Yi et al., "Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors," ArXiv170506362 Cs, May 2017.
8. S.-C. B. Lo, H.-P. Chan, J.-S. Lin, H. Li, M. T. Freedman, and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," Neural Netw., vol. 8, no. 7–8, pp. 1201–1214, 1995

9. A. R. Jamieson, K. Drukker, and M. L. Giger, “Breast image feature learning with adaptive deconvolutional networks,” presented at the Medical Imaging 2012: Computer-Aided Diagnosis, 2012, vol. 8315, p. 831506.
10. V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, “Efficient Processing of Deep Neural Networks: A Tutorial and Survey,” ArXiv170309039 Cs, Mar. 2017.
11. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
12. D. Erhan, P.-A. Manzagol, Y. Bengio, S. Bengio, and P. Vincent, “The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training,” 2009.
13. H. C. Shin et al., “Deep Convolutional Neural Networks for Computer-Aided De- tection: CNN Architectures, Dataset Characteristics and Transfer Learning,” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, May 2016.
14. N. Tajbakhsh 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, May 2016.
15. A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off- the-shelf: an astounding baseline for recognition,” in Proceedings of the IEEE con- ference on computer vision and pattern recognition workshops, 2014, pp. 806–813.
16. O. A. Penatti, K. Nogueira, and J. A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 44–51.
17. A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
18. J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, “Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 589–596.
19. W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convolutional neural networks for lung nodule classification,” in International Conference on Information Processing in Medical Imaging, 2015, pp. 588–599.
20. B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, “Off-the-shelf con- volutional neural network features for pulmonary nodule detection in computed tomography scans,” in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015, pp.

286–289.

21. F. Ciompi et al., “Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box,” Med. Image Anal., vol. 26, no. 1, pp. 195–202, 2015.
22. T. Schlegl, J. Ofner, and G. Langs, “Unsupervised pretraining across image domains improves lung tissue classification,” in International MICCAI Workshop on Medical Computer Vision, 2014, pp. 82–93.
23. R. Li et al., “Deep learning based imaging data completion for improved brain disease diagnosis,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2014, pp. 305–312.
24. J. Suckling et al., “The Mammographic Image Analysis Society Digital Mammo- gram Database,” Exerpta Medica, vol. 1069, pp. 375–378, 1994.

Lévy, Daniel, and Arzav Jain. "Breast mass classification from mammograms using deep convolutional neur