Transfer Learning with Pre-trained CNNs for Breast Cancer Stage Identification

Tesfahunegn Minwuyelet Mengistu(✉), Birtukan Shegaw Arega, and Birhanu Hailu Belay

Bahir Dar University, Bahir Dar, Ethiopia
tesfahunegn9@gmail.com

Abstract. Breast cancer stage identification is an important prerequisite for early treatment to increase the chance of survival, and predict the recurrence of cancer. Research works done so far were mainly focused on the classification of breast cancer types while many of them are neglecting to stage of breast cancer. Obtaining an adequate labeled breast cancer image dataset for training machine learning algorithms is challenging. In this paper, we propose a pre-trained Convolutional Neural Network (Pretrained-CNN) model for Breast Cancer Stage Identification. The proposed method is designed by leveraging transfer learning techniques. Further, the performance of the pre-trained model is compared with CNN-based models that are trained from scratch. The performance of the proposed model is tested using a publicly available breast cancer-image dataset taken and achieved a promising result with an overall classification accuracy of 90%

Keywords: Breast cancer · Pre-trained model · CNN · Segmentation · Transfer learning

1 Introduction

In many clinical practices, the detection and identification of diseases including breast cancer diagnostic and medical image interpretation have been made by the expertise of individual clinicians and/or physicians (Chen et al. 2022; Birtukan et al. 2020). The medical decision mainly relies on the physician's knowledge, and experience which results in large variability in interpreting medical images (Chen et al. 2022). To address such problems various types of research proposed machine learning-based models for interpreting and analysis of medical images (Birtukan et al. 2020). Artificial Intelligence (AI)-based models have been widely applied in medical image processing including diagnosis and staging and detection of breast cancer from digital mammography (Li et al. 2020; Dabeer et al. 2019). These tools help to detect the suspicious region in mammogram images and classify the suspicious regions into different classes. Compared to the manual detection of breast cancer, such AI-assisted systems improve the accuracy of diagnosis and detection of the stage of breast cancer (Dembrower et al. 2020; Tahmooresi et al.

B. H. Woldegiorgis et al. (Eds.): ICAST 2022, LNICST 455, pp. 127–136, 2023.
https://doi.org/10.1007/978-3-031-28725-1_8

2018). Furthermore, as a simultaneous assistant to a radiologist, the use of these AI-enabled cancer detectors can identify other cancers (Kim 2022; Schaffter et al. n.d.; McKinney et al. 2020).

Breast cancer is one of the most prominent cancer types and is the leading cause of death in women (Birtukan et al. 2020; Boughorbel et al. 2016). Generally, breast cancer is classified into five stages (Birtukan et al. 2020). In clinical settings, the stages of breast cancer are usually expressed on a scale of 0 through IV where stage 0 denotes non-invasive cancer and stage IV denotes invasive cancers (Birtukan et al. 2020; Boughorbel et al. 2016) (Breastcancer.org. 2018). Sample images of the five breast cancer images taken from (Birtukan et al. 2020) are illustrated in Fig. 1.

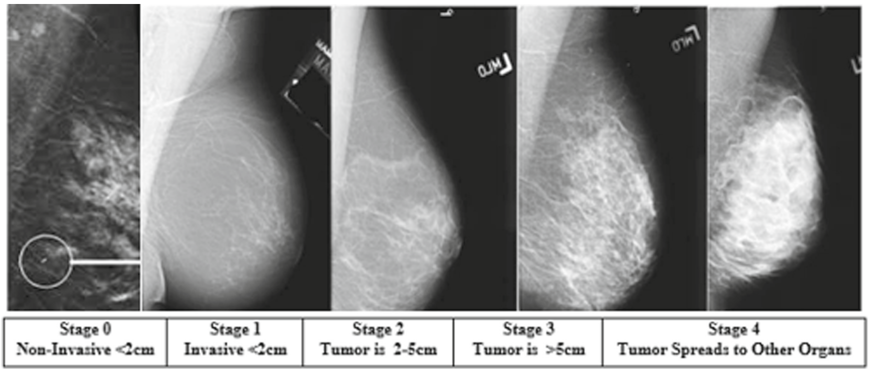


Fig. 1. Sample images of the five breast cancer stages (Birtukan et al. 2020)

Advances in deep learning and its remarkable success in clinical applications has attracted the attention of researcher in medical domains (Boughorbel et al. 2016; Goswami 2018; Saha 2015). Despite the success, the lack of a labeled image dataset has been the major bottleneck in developing a robust deep learning-based model for medical image analysis (Chen et al. 2022; Birtukan et al. 2020). This is also true in breast cancer detection and identification. Therefore, various researchers (Breastcancer.org. 2018; Nadig 2017; McCowan et al. 2007) use CNN and handcrafted feature-assisted classical machine learning algorithms for breast cancer detection and staging. Since these handcrafted features consist of limited information about the image, the recognition performance of the models was very limited (Birtukan et al. 2020).

Nowadays, to overcome the shortage of training datasets, researchers employ transfer learning techniques whereby models are first trained on a problem and one or more layers from the trained model are then used in a second new model of a related problem (Li et al. 2020; Kim 2022). As a continuation of our previous work (Birtukan et al. 2020), which employed CNNs and traditional machine learning algorithms as feature extractors and classifiers respectively, this paper presents a model with pre-trained CNN backbones by leveraging transfer learning techniques for Breast cancer staging.

The rest of the paper is organized as follows: Sect. 2 reviews the relevant methods and related works. The proposed methods and training settings are described in Sect. 3.

Section 4 presents all experiments and results obtained from the experiments. Finally, Sect. 5, presents conclusions and future research directions.

2 Related Works

The existing breast cancer detection and staging models can utilize either handcrafted or automatic features. Methods belonging to the first category were mainly applied before the introduction of deep learning and follows step-wise routines. In contrast, the second approach integrated the feature extraction and classification steps and trained from end to end. Therefore, in this section, we review the research trends in breast cancer detection and the existing state-of-the-art techniques that are applied for medical image analysis where labeled images are very limited.

A k-Means based Gaussian Mixture Model (GMM) is proposed to detect and classify breast cancer as benign or malignant (Dheeba 2019). The authors developed a model by following three steps. First, they found a region of interest by using the K-means-based GMM segmentation technique, and then they applied texture feature extraction and optimization of features of the Region of Interest (ROI) by using a Genetic Algorithm (GA). Finally, classified the detected abnormality as benign or malignant. To develop and evaluate the model they used the publicly available Mammographic Image Analysis Society (MIAS) dataset and got an accuracy of 95.8%. They classified images as benign or malignant but they did not work on identifying stages of detected malignant breast cancer.

Researchers in (Shen 2017) also developed a deep learning algorithm that can detect breast cancer on screening mammograms by using an end-to-end training approach. They used a combined dataset of publicly available Digital databases for Screening mammograms (CBIS-DDSM) and a private dataset of Full-Field Digital Mammography (FFDM) from the INbreast database. They achieved 86.1% of sensitivity and 80.1% of specificity in CBIS-DDSM and 86.7% of sensitivity and 96.1% of specificity in FFDM of the INbreast dataset. The mammograms are classified as cancer or normal. It did not identify whether the cancerous are benign or malignant and at which stage the cancerous case is reached.

Another CNN-based approach for cancer diagnosis on the histopathological image was also proposed in (Dabeer et al. 2019). They used datasets from the BreakHis database and reported 99.86% of accuracy. They introduced deep learning architecture for breast cancer detection as benign and malignant but they did not work on which stage malignant breast cancer reach that is used to start treatment, estimate recurrence, and increase survival from cancer.

The breast cancer histopathology image classification scheme was also proposed by researchers in (Zhu et al. 2019) by assembling multiple compact CNNs. They used two breast cancer datasets these are the BreakHis database which contains 7909 images taken from the breast tissue biopsy side and the BreAst Cancer Histology (BACH) which contains 400 breast histology images and achieved 84.4% accuracy. They classified breast cancer histopathological images based on multiple compact CNN as cancer or not cancer.

CNN was used by researchers for breast cancer screening as a multi-view deep CNN (Geras et al. 2017). They used a mammography-based breast cancer screening

Breast Imaging Reporting and Database System (BI-RADS) dataset having 886,000 images to study the impact of training set size and image size on CNN cancer prediction accuracy. They focused on identifying the impact of the training set size and image size on the prediction of cancer accuracy. They did not consider how can identify and classify stages of malignant breast cancer since identifying stages of breast cancer. A similar work (Dalmış et al. 2018) also proposed a CNN-based model for breast cancer screening from a Magnetic Resonance Image (MRI). 385 MRI scans, containing 161 malignant lesions.

Nine-layer CNN with the parametric rectified linear unit and rank-based stochastic pooling is also employed for abnormal breast identification. Researchers focused on how to select the optimal number of convolution layers and the effect of data augmentation on breast cancer detection. They were using a mini-MIAS database that contains 209 normal breast images and 113 abnormal breast images. They achieved results over 100 test sets with 94.0% of accuracy, 94.5% of precision, 93.4% of sensitivity, and 94.6 specificity by combining ReLU and rank-based stochastic pooling. They did not consider the classification of stages of abnormal breast images.

Researchers in (Wang et al. 2014) used convolutional neural network features by combining them with handcrafted features for Mitosis detection in breast cancer pathology images. They used the public ICPR12 mitosis dataset that has 226 mitoses and 15 testing HPFs and got an F-measure of 0.7345. They considered mitosis count for grading of breast cancer they did not work on the identification of stages of breast cancer necessary to start treatment and for better treatment suggestions.

CNN was used for automated breast ultrasound lesions detection (Yap et al. 2018). They used two different datasets of US images that were obtained from US systems. Dataset A contains 306 images with 246 benign and 60 malignant cases. Dataset B contains 163 images with 110 benign and 53 malignant cases. The proposed model detects lesions either benign or malignant.

Researchers in (Nadig 2017) proposed stage-specific predictive models for breast cancer survivability by using three different machine learning methods (naïve Bayes, Logistic regression, and decision tree). And compared their accuracy to predict survivability. They used a publicly available SEER dataset. Unlike image-based staging, they considered text-based stage information as a factor. The other method for the prediction of breast cancer using big data analytics by using the K-nearest neighbor algorithm was also proposed in (Shailaja et al. 2018). The Wisconsin breast cancer dataset taken from the UCI machine learning repository that contains 699 instances with 11 attributes are classified as either benign or malignant by using KNN.

Though there are limited labeled breast cancer training datasets, nowadays, the detection of breast cancer for early diagnosis from image data is becoming common and many researchers reported promising results employing Deep convolutional neural networks (Breastcancer.org 2018). Hence, in this paper, we propose a method that can overcome the issue of labeled training image scarcity with a great emphasis on the identification of stages of breast cancer. The following sections give a detailed overview of the proposed breast cancer stinging model.

3 Materials and Methods

In this section, we describe the breast cancer dataset employed in model development and elaborate on the details of the proposed breast cancer staging model architecture. To develop the breast cancer stages identification models we follow experimental research which consists of dataset preparation, model training, and evaluation.

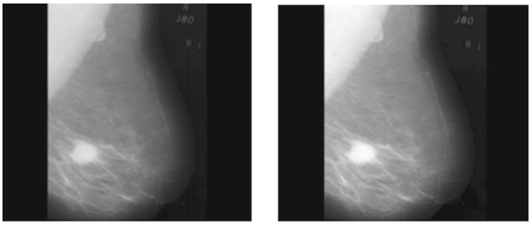
3.1 Dataset

In the era of deep learning, the shortage of training datasets is one of the bottlenecks that limit the development of robust models for disease diagnosis in general and specifically for breast cancer staging. Few datasets have been used in various works on breast cancer detection (Dheeba 2019). Many of the datasets used in the literature consist of very limited images that are prepared for a very specific use case. Recently, an updated and organized breast cancer staging dataset has introduced by (Birtukan et al. 2020) where the images were collected from various sources including the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) (Lee et al. 2017) and it is publicly available in Kaggle at *[*https://www.kaggle.com/datasets/tesfahunegn/breast-cancer-stage-identifications*]*. This dataset consists of 1469 Images in MINIST files that are split as training and testing data. The details of the dataset used in this experiment are summarized in Table 1.

Table 1. Number of breast cancer images in each stage

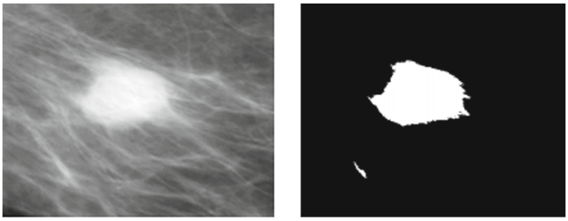
No	Classes	Image format	Number of images
1	Stage 0	Png	395
2	Stage 1	Png	495
3	Stage 2	png	310
4	Stage 3	Png	173
5	Stage 4	Png	96
Total			1469

Sample breast cancer images are illustrated in Figs. 2 and 3.



(A) (B)

Fig. 2. Sample enhanced breast cancer image (Birtukan et al. 2020). (A) Original image, (B) image after filtering and enhancement



(A) (B)

Fig. 3. Example of ROI segmentation (Birtukan et al. 2020)

3.2 The Proposed Model Architecture

Considering the nature and size of the dataset, we propose two experimental setups. The first approach employs a CNN-based architecture which is trained from scratch while the second breast cancer stage identification approach uses pre-trained CNN models as a backbone. The two proposed approached and the overall model architectures are depicted in Figs. 4 and 5 respectively.

The network architecture in the first approach consists of the feature learner that takes the segmented image of the breast cancer and encodes it to a high-level feature vector representation and the classification layers (which consists of a fully connected layer with softmax) which are responsible to produce the probability distribution of breast cancer stages over a given input feature vectors. The architecture of the CNN-based model that is trained from the scratch is illustrated in Fig. 4.

This CNN-based model consists of seven convolution layers and the number of filters that are used to control the depth of the output volume was 64, 128, and 256 filters also a 3×3 filter size at a single layer have used. We applied the ReLU activation function, a pooling size of two (2×2) after the two consecutive convolution layers, dropout layers with dropping probability (p) = 0.25, 0.4, and 0.5 at each fully connected layer, batch normalization, Adam optimizer as an optimization function, a learning rate of 0.003, categorical cross-entropy as a loss function. The model was for 25 epochs with a batch size of 32. Unlike our previous work (Birtukan et al. 2020), in this paper, we haven't applied data augmentation.

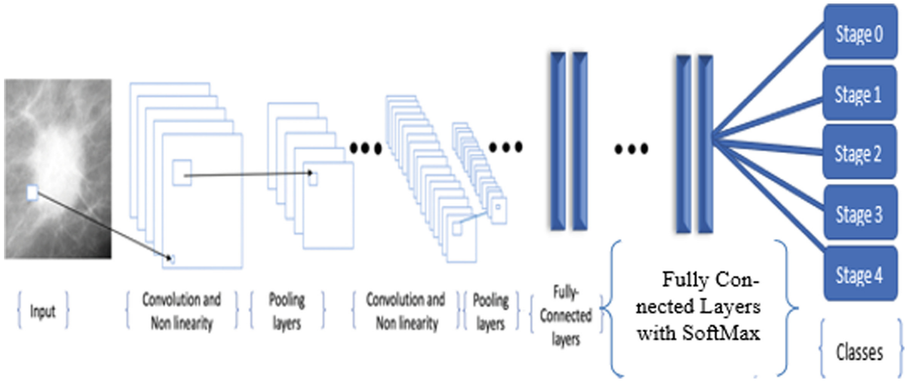


Fig. 4. Proposed CNN model architecture

The second approach employed a pre-trained CNN in which the model reuses the already learning method for other tasks. This approach is commonly called transfer learning. In this experimental setup, the pre-trained VGG16 model trained with ImageNet recognition tasks is applied for breast cancer stage identification where the number of class labels is set to five. To demonstrate the effect of the depth of fine-tuned layers on recognition performance, we fine-tuned some of the specific layers (up to 3 layers), while the remaining layers are set to freeze. Since the pre-training VGG net input size was 224 × 224, we changed the input tensor size of the breast cancer image to a similar size of 224 × 224. During loading the pre-trained VGG16 model, we don't load the fully connected layers; thus, we add custom fully connected (FC) layers which will be trained together. The fine-tuned and frozen layers of the pre-trained VGG net model are illustrated in Fig. 5.

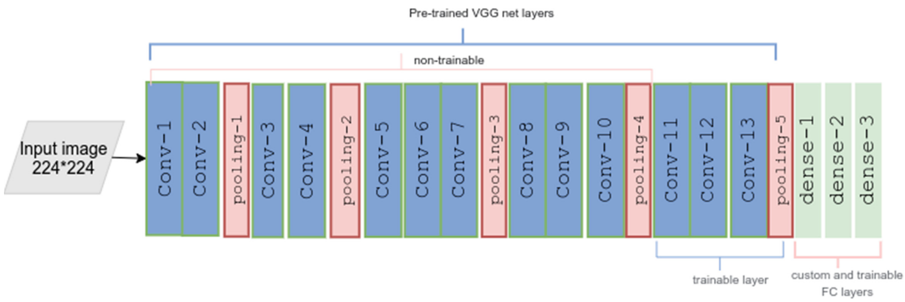


Fig. 5. Pre-trained VGG net setups: all frozen layers except the last three and custom fully connected layers

In both experimental setups, the whole dataset is split as 80/20 for training and testing respectively and after that again we classified the remaining training data as 80/20 for the training and validation phase respectively that is we used 60% for training the model, 20% for validating the model and to remove bias to training dataset and 20% for testing and evaluating the model.

4 Experimental Results

Experiments were conducted using the dataset (Birtukan et al. 2020), which is a freely available breast cancer dataset. The model architecture and experimental setups described in Sect. 3 are implemented using Keras Application Program Interface (API) with a Tensor Flow as a backend. To select a suitable model parameter, different values of these parameters were considered and tuned during model training. In addition to dropouts, we also employ early stopping to avoid over-fitting The best results recorded during experimentation are reported in Table 2. The performance of both models is measured using an accuracy metric.

We compared the accuracy of the CNN-based model trained from the scratch with a pre-trained CNN model, and our pre-trained CNN model outperforms the classification accuracy of the model that was trained from scratch by a large margin.

Table 2. Summary of comparison of models based on testing accuracy

Dataset used	Model	Testing accuracy
Full mammogram images	CNN-SoftMax	39%
	Pretrained-CNN model	42%
Segmented ROI images	CNN- SoftMax	84%
	Pretrained-CNN model	90%

5 Conclusion

Early identification of stages of breast cancer allows one to get better treatment, expect recurrence and survival, and control the spread of cancer to the other part of the body. Extensive research has been done on breast cancer detection, and many of them are focusing on breast cancer type classification while neglecting the stage of breast cancer. Therefore, in this paper, two CNN-based models are proposed for breast cancer stage identification. The first model is trained from scratch while the second model leverages the concept of transfer learning through which knowledge is reused from other pre-trained models. The proposed models are then evaluated using the publicly available breast cancer dataset and it achieves promising results with an over-recognition accuracy of 90% Based on the results observed during experimentation, using pre-trained CNN models gives significant discrimination performance compared to the CN-based models that are trained from scratch. To improve the performance of the proposed model, as part of future work, other pre-trained CNN models could be explored and investigated. In addition, instead of developing independent models, we plan to develop a multi-tasking deep learning model that can learn the breast cancer type and stage simultaneously.

Author Contributions. Conceptualization, T.M. and B.S; Methodology, T.M. and B.S.; Writing—original draft preparation, T.M. and B.S.; Writing—review and editing, B.H. All authors have read and agreed to the published version of the manuscript.

References

Birtukan, S., et al.: Breast cancer Stage identification using Machine learning. Univerity of Gondar, Gondar, Ethiopia (2020)

Breastcancer.org.: Breast cancer stages: 0 through iv. (2018). Retrieved from http://www.breastcancer.org/symptoms/diagnosis/stagi

Zhu, C., et al.: Breast cancer histopathology image classification through assembling multiple compact CNNs. BMC Med. Inform. Decis. Mak. **19**(1), 1–17 (2019). https://doi.org/10.1186/s12911-019-0913-x

Dheeba, S.S.: A research on detection and classification of breast cancer using k-means gmm & CNN algorithms. Int. J. Eng. Adv. Technol. **8**(6), 501–509 (2019). https://doi.org/10.35940/ijeat.F1102.0886S19

Goswami, T.: Impact of deep learning in image processing and computer vision. In: Anguera, J., Satapathy, S.C., Bhateja, V., Sunitha, K.V.N. (eds.) Microelectronics, Electromagnetics and Telecommunications. LNEE, vol. 471, pp. 475–485. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7329-8_48

Wang, H., et al.: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J. Med. Imaging **1**(3), 034003 (2014). https://doi.org/10.1117/1.jmi.1.3.034003

McCowan, I.A., et al.: Collection of cancer stage data by classifying free-text medical reports. J. Am. Med. Informatics Assoc. **14**, 736–745 (2007). https://doi.org/10.1197/jamia.M2130

Dembrower, K., et al.: Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit. Health **2**(9), e468–e474 (2020). https://doi.org/10.1016/S2589-7500(20)30185-0

Geras, K.J., et al.: High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv, 1–9

Shailaja, K., et al.: Prediction of breast cancer using big data analytics. Int. J. Eng. Technol **7**, 223–226 (2018). https://doi.org/10.14419/ijet.v7i4.6.20480

Kim, H., e.: Transfer learning for medical image classification: a literature review. BMC Med. Imag. (2022). https://doi.org/10.1186/s12880-022-00793-7

Yap, M.H., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Heal. Inform. **22**(4), 1218–1226 (2018). https://doi.org/10.1109/JBHI.2017.2731873

Tahmooresi, M., et al.: Early detection of breast cancer using machine learning techniques. J. Telecommun. Electron. Comput. Eng. **10**, 21–27 (2018)

Dalmış, M.U., et al.: Fully automated detection of breast cancer in screening MRI using convolutional neural networks. J. Med. Imaging **5**(1), 1 (2018). https://doi.org/10.1117/1.jmi.5.1.014502

Nadig, R.J.: Stage-specific predictive models for breast cancer survivability. Int. J. Med. Inform. **97**, 304–311 (2017). https://doi.org/10.1016/j.ijmedinf.2016.11.001

Lee, R.S., et al.: A curated mammography data set for use in computer-aided detectionand diagnosis research. (2017). https://doi.org/10.1038/sdata.2017.177

Boughorbel, S., et al.: Model comparison for breast cancer prognosis based on clinical data. PLoS ONE **11**, 1–15 (2016). https://doi.org/10.1371/journal.pone.0146413

McKinney, S.M., et al.: International evaluation of an AI system for breast cancer screening. Nature **577**, 89–94 (2020). https://doi.org/10.1038/s41586-019-1799-6

Saha, S.: A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Towards Data Science, pp. 1–19 (2015). Retrieved from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neuralnetworks-the-eli5-way-3bd2b1164a53

Shen, L.: End-to-end training for whole image breast cancer diagnosis using an all convolutional design. arXiv, pp. 1–12 (2017). https://doi.org/10.1038/s41598-019-48995-4

Dabeer, S., et al.: Cancer diagnosis in the histopathological image: CNN based approach. Inform. Med. Unlocked **16**. https://doi.org/10.1016/j.imu.2019.100231

Schaffter, T., et al.: Evaluation of combined artificial intelligence and radiologist assessment to interpret screening Mammograms. JAMA Netw. Open. **3**(3), e200265 (n.d.). https://doi.org/10.1001/jamanetworkopen.2020.0265

Li, X., et al.: Transfer learning in computer vision tasks: remember where you come from. Image Vision Comput. **93** (2020)

Chen, X., et al.: Recent advances and clinical applications of deep learning in medical image analysis. Med. Image Anal. **79**. (2022)doi:https://doi.org/10.1016/j.media.2022.102444