Anshul Jaswal et al. 2021, 9:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

A Review Breast Cancer Classification using Transfer Learning

Anshul Jaswal, Kumari Archana
Department of Computer Science,
Himachal Pradesh Technical University,
Himachal Pradesh, India,
anshuljaswal4@gmail.com, niralaarchana.1991@gmail.com

Abstract- Breast cancer is a significant factor in female mortality. An early cancer diagnosis leads to a reduction in the breast cancer death rate. With the help of a computer-aided diagnosis system, the efficiency increased, and the cost was reduced for the cancer diagnosis. Traditional breast cancer classification techniques are based on handcrafted features techniques, and their performance relies upon the chosen features. They also are very sensitive to different sizes and complex shapes. However, histopathological breast cancer images are very complex in shape. Currently, deep learning models have become an alternative solution for diagnosis, and have overcame the drawbacks of classical classification techniques. Although it still has some challenges, mainly challenge is the lack of training data. To address this challenge and optimize the performance, a new developed technique is transfer learning which is where the deep learning model strains on a task, and then fine-tunes the models for another task.

Keywords: Breast Cancer, Deep learning, Transfer learning, and Convolutional Neural Network etc.

INTRODUCTION

Cancer is one of the most death causing diseases worldwide currently. Any person can be affected by this. Cancer arises when the abnormal body's cells start to separate and come in contact with normal cells and make them malignant. Among women, breast cancer (BC) related deaths are frequent in comparison to the other types of cancer related deaths [1], and this causes a large number of deaths each year worldwide [2]. After skin cancer, breast cancer is the most common cancer diagnosed in women. Reports say that the incidence rate of breast cancer ranges from 19.3 per 100,000 women in East Africa, to 89.7 per 100,000 women in Western Europe [3]. This number of new cases has Continued to grow in recent years, and this number is expected to increase to 27 million in 2030[4]. In India, the annual percentage change in the incidence ranged from 0.46 to 2.56 for breast cancer. Breast cancer seems to be more common in the young age group and 52% of all women suffering from breast cancer in Mumbai are between 40 and 49 years of age. A significant number of patients are below 30 years [5]. BC is a disease in which malignant (cancer) cells are formed in tissues of the breast. BC can occur in men and women both, but it is more common in women. Breast cancer starts in the cells of the breast as a group of cancer cells which can then invade surrounding tissues or spread to other areas of the body. Sometimes, this process of cell growth goes wrong and new cells form when the body does not need them and old or damaged cells do not die as

they should so a buildup of cells often forms a mass of tissue called a lump or tumor.

Breast cancer occurs when these malignant tumors develop in the breast. These cells can spread by breaking away from the original tumor and entering blood vessels which branch into tissues throughout the body. When these cancer cells travel to other parts of the body and starts damaging other tissues and organs, the process is called metastasis. These BC are further classified into Carcinomas and Non Carcinomas.

BC occurs when some breast cells begin to grow abnormally. These cells divide more rapidly than the healthy cells do and continue to accumulate, forming a lump or mass. Cells may spread through the breast to the lymph nodes or to other parts of the body. BC most often begins with cells in milk producing ducts (invasive ductal carcinoma).

BC may also begin in the glandular tissue called lobules (invasive lobular carcinoma) or in other cells or tissue within the breast. Risk factors for developing breast cancer include being female, obesity, lack of physical exercises, alcoholism, hormone replacement therapy during menopause, ionizing radiation, an early age at first menstruation, having children late in life or not at all, older age, a prior history of breast cancer, and a family history of breast cancer[6][7].About 5–10% of cases are the result of a genetic predisposition inherited from a person's parents[6]. It is likely that it is caused by a complex interaction of the genetic makeup and the environment.

There are four types of BC. First type of breast cancer is Ductal Carcinoma in Situ which is found in coating of breast milk ducts and it is pre-stage breast cancer. Second type of breast cancer is the most popular disease and contains up to 70-80% diagnosis. Third type of breast cancer is In flammatory breast cancer which is forcefully and quickly developing breast cancer, in this disease cells penetrate the skin and lymph vessels of the breast. The fourth type of breast cancer is Metastatic breast cancer which spreads to other parts of the body. The diagnosis and treatment of BC in the early stages is essential to prevent the disease progression and reduce its morbidity rates [8]. Detection and diagnosis of BC can be achieved through various non invasive methods and biopsy[9]. Non invasive methods are basically the imaging procedures these are: diagnostic mammograms (x-ray), Magnetic Resonance Imaging (MRI) of the breast, thermography, mammography [10], breast ultrasound [11], biopsy [9] and other methods.

Although the use of imaging procedures for cancer screening is wide spread, however biopsy is the only diagnostic procedure that can definitely determine if the specific area is cancerous. BC diagnosis usually consists of an initial detection via palpation and regular checkups using mammography or ultrasound imaging. This diagnosis is then followed by the breast tissue biopsy if the check-up exam indicates the possibility of growth of malignant tissue [12].

In biopsy techniques, procedures such as Fine Needle Aspiration (FNA), Core Needle Biopsy (CNB), Vacuum Assisted Breast Biopsy (VABB) and Surgical (open) Biopsy (SOB) stand out [13] are available. Biopsy procedures collect samples of the cells or tissue. The tissues collected from patient during biopsy is commonly stained with the

Hematoxylin and Eosin (H&E) as staining enhances nuclei(purple) and cytoplasm(pinkish) as well as other structures of the interest. Computer-aided diagnosis (CAD) is widely used for detection of numerous diseases with accurate decision. It assists the healthcare professionals for analyzing and concluding the stages of the various diseases.

CAD is developed in such a way to provide us with promising result and quality decision making on patient condition so that it helps the health care professionals to diagnose the various stages of disease effectively. It also supports radiologists to avert misconceptions and wrong diagnosis due to inaccurate data, lack of focus, or inexperience, who uses visually screening mammogram of patients.
Deep Learning (DL) is a growing technology in the field of machine learning (ML) and it has got the attention of many researchers [14]. In DL mainly the Convolutional Neural Network (CNN) has achieved great success in large-scale image and video recognition, as it is helping the health care professionals to analyze and on the basis of this analysis conclude the stages of various diseases and reduce the intense workload for

specialized pathologist. The breast cancer classification task is challenging due to the complexity of the breast cancer images.

This requires a deep network with excellent feature representation to extract features, and DL models have the ability to do this task with high performance. So this research proposes a deep Convolutional Neural Network (CNN) model to classify Hematoxylin–Eosin stained breast biopsy images into **four classes (invasive carcinoma, in-situ carcinoma, benign tumor and normal tissue)**. Moreover, one of the main challenges of employing DL in the breast cancer classification task is the lack of training data due to a large amount of time to collect the images and expertise needed to label the images. To tackle this challenge, we have used a **transfer learning technique**.

II.TECHNIQUES FOR BREAST CANCER CLASSIFICATION

The time-consuming task has become efficient with the use of CAD. CAD is very helpful in the medical field to classify the images for different diseases. Similarly for classifying the histopathology image of BC,

Kowal et al. [16] used morphological, topological, and texture capabilities, and after that capabilities have been classified on distinct clustering algorithms for nuclei segmentation on biopsy microscopic images. The classifier performed an accuracy of 84.44% as worst and best as 93% on 500 photographs from 50 patients. Further, it was categorized as patient wise classification by majority voting on 10 photographs each, which in addition gave stepped forward accuracy of 96-100%. With the improvement of technology, the Convolution Neural Network (CNN) done outstanding achievement in large-scale image and video.

Spanhol et al. [17] used AlexNet [15] **f**or the classification of BC histopathology images for benign and malignant classes. The results were more as compared to the previous traditional method with a 6% higher accuracy.

Teresa Araujo et al. [18] used a method to categorize BioImaging challenge BC dataset (2015) by self-designing the CNN model for the diagnosis of BC H&E stained histopathology image. Firstly, the author by using various data augmentation techniques such as rotation, vertical flip, and overlap patches increased the dataset size. After that in the CNN these patches were fed in for patch-wise classification. For predicting image-wise classification accuracy different techniques (majority voting, maximum probability and the sum of probability) are used. The author to categorize these features using an SVM classifier has also extracted features from the second flattened layer of the self-design CNN model which gave an accuracy of 77% for four classes and 83% for two classes.

Kausik et al. [19] proposed a multiple instance learning(MIL) framework for CNN. To extract the most informative features from the patches which constituted an entire slide, without the need of the whole slide scope, the author presented a modern pooling layer. An accuracy of almost 88% on BC images was achieved by the CNN network.

S.A. Adeshina et al. [20] adopted a DCNN architecture combined with an Ensemble learning method using Tensorflow framework ReLU activation function to achieve the accurate automated classification of BC images. An inter-class classification accuracy of 91.5% was achieved by the author with the BreakHis dataset.

Mahbod et al. [21] on the BioImaging challenge dataset performed the RGB histogram normalization method. Then by using the data augmentation method increased the size of the dataset. Then the author trained the dataset on ResNet50 and ResNet101 [22] CNN model using transfer learning to classify the class of BC image.

Habib Dhahri et al. [23] proposed genetic programming and machine learning algorithms which aimed to develop a system to accurately classify benign and malignant breast tumours. The author deduced that Logistic Regression, Linear Discriminant Analysis, and Gaussian Naive Bayes algorithms fit better than any other methods by using the default input parameters for all the machine learning classifiers.

Sumaiya Dabeer et al. [24] proposed a system by training 2480 benign and 5429 malignant breast photographs which belonged to the RGB colour

model. The author highlighted the visual patterns by using the labelled input image from raw pixels. Then to differentiate between the benign and malign tissue he utilized these patterns, with the help of a classifier network working akin to digital staining, which highlighted the image segments critical for diagnostic decisions. The system gives an effective classification model for classifying breast tissue between benign or malignant.

Mahesh Gour et al. [25] proposed a method in which the data augmentation techniques were used initially for increasing BreakHis dataset. Then he obtained images that were fed to the ResiHis network for training. The accuracy reported was 92.52% and F-score reported was 93.54%.

Table -1: Summary of Literature Review

Year	Author	Dataset	Methodology	Result
2013	Kowal et al. [16]	500 images from 50 patients	**Features**: morphological, topological, and texture **Classifier**: Different machine learning classifier	Worst Accuracy = 84.44% Best Accuracy = 93%
2016	Spanhol et al. [17]	BreakHis	**Pre-processing**: Data Augmentation **CNN**: AlexNet	Patient level= 88.4% Image level= 83.5%
2017	Araujo et al. [18]	BioImaging Challenging Dataset (2015)	**Pre-processing**: Augmentation patches extraction **CNN**: self design	Binary class= 83% Multiclass= 77%
2018	Kausik et al. [19]	BreakHis	Multiple Instance Learning (MIL) framework for CNN for extractive most informative patches	Accuracy = 88%
2019	Adebanjo et al. [20]	BreakHis	**Classifier**: DCNN architecture combined with ensemble learning	Accuracy = 91.5%
2019	Mahbod et al. [21]	BioImaging Challenge Dataset (2015)	**Pre-processing**: RGB normalization. **CNN**: ResNet50 and ResNet101	Test Acuracy = 95% Extended Accuracy = 87.7% Overall Accuracy = 93.7%
2016	Habib Dhahri et al. [23]	Benett Dataset	**Pre-processing**: Genetic Algorithm Classifier: Machine learning algorithm	AdaBoost classifier Accuracy = 98.24%
2019	Sumaiya Dabeer et al. [24]	BreakHis	**Pre-processing**: Highlight visual patterns Highlight image segment using classifier network.	Accuracy = 93.44%

2020	Mahesh Gour et al. [25]	BreakHis	**Pre-processing**: Data Augmentation **CNN**: ResiHis	Accuracy = 95.52% F-score= 93.54%

III.CONCLUSION

Breast cancer detection is a challenging problem because it is most popular and harmful disease. Breast cancer is growing year by year and there is a less chance to recover from this disease.

Machine learning and deep learning techniques are used to detect breast cancer. A CNN-based approach for the classification of Hematoxylin and Eosin stained histological breast cancer images is proposed. All the relevant features are learned by the network, which reduces the need of field knowledge. Images are then classified into four types as normal tissue, benign lesion, in situ carcinoma or invasive carcinoma. Alternatively, a binary classification as carcinoma or non-carcinoma is also performed. For this, architecture of the network is designed for information extraction from different relevant scales, including the nuclei and overall tissues organization. The network is then trained on an augmented patch dataset and tested on a separate set of images.

Dataset augmentation1and scale-based1network design both are important for the1success of the approach. The extracted features can also be used for training a SVM classifier. Both CNNs achieve comparable results. The classification scheme allows1obtaining high sensitivity for the carcinoma cases, which is for interest of pathologists.

The performance of the system is similar or superior to the state-of-the-art methods, even though1a smaller and more challenging dataset is1used. Finally, since the network is designed to consider multiple biological scales, the proposed system can be extended for whole-slide breast histology image classification relevant for clinical settings.

REFERENCES

[1]. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," CA. Cancer J. Clin., vol. 67, no. 1, pp. 7–30, 2017, doi: 10.3322/caac.21387.

[2]. D. C. Koboldt et al., "Comprehensive molecular portraits of human breast tumours," Nature, vol. 490, no. 7418, pp. 61–70, 2012, doi: 10.1038/nature11412.

[3]. M. M. Al Rahhal, "Breast cancer classification in histopathological images using convolutional neural network," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 3, pp. 64– 8, 2018, doi: 10.14569/IJACSA.2018.090310.

[4]. D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," IEEE Access, vol. 6, pp. 24680–24693, 2018, doi: 10.1109/ACCESS.2018.2831280.

[5]. A. Khokhar, "Breast cancer in india: Where do we stand and where do we go?," Asian Pacific J. Cancer Prev., vol. 13, no. 10, pp. 4861–4866, 2012.

[6]. PDQ Adult Treatment Editorial Board, "Breast Cancer Treatment (Adult) (PDQ®): Patient Version," PDQ Cancer Inf. Summ., pp. 1–32, 2002.

[7]. S. McGuire, "World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015," Adv. Nutr., vol. 7, no. 2, pp. 418–419, 2016, doi: 10.3945/an.116.012211.

[8]. R. A. Smith, V. Cokkinides, and H. J. Eyre, "American Cancer Society Guidelines for the Early Detection of Cancer, 2006," CA. Cancer J. Clin., vol. 56, no. 1, pp. 11–25, 2006, doi:10.3322/canjclin.56.1.11.

[9]. J. V. Horvat, D. M. Keating, H. Rodrigues-Duarte, E. A. Morris, and V. L. Mango, "Calcifications at digital breast tomosynthesis: Imaging features and biopsy techniques," Radiographics, vol. 39, no. 2, pp. 307–318, 2019, doi: 10.1148/rg.2019180124.

[10]. C. Lim, A. Mra, C. Suliong, C. V. Rao, T. Aung, and J. Sieman, "Recent Advances in Breast Cancer Diagnosis Entering an Era of Precision Medicine," Borneo J. Med. Sci., vol. 13, no. 1, pp. 3–9, 2019.

[11]. Y.-L. Huang, PhD, D.-R. Chen, MD, and Y.-C. Lin, "3D Contouring for Breast Tumor in

Sonography," 2019. Breast Cancer Diagnosis 2015," 2015.

[12]. Breastcancer.org. Biopsy. 2012. url: http://www.breastcancer.org/symptoms/testing/types/biopsy

[13]. Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," PLoS One, vol. 14, no. 3, pp. 1–21, 2019, doi: 10.1371/journal.pone.0214587.

[14]. M. I. Nounou, F. ElAmrawy, N. Ahmed, K. Abdelraouf, S. Goda, and H. SyedSha-Qhattal, "Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies," Breast Cancer Basic Clin. Res., vol. 9s2, p. BCBCR.S29420, Jan. 2015, doi: 10.4137/BCBCR.S29420.

[15]. . Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computeraided diagnosis of breast cancer based on fine needle biopsy microscopic images," Comput. Biol. Med., vol. 43, no. 10, pp. 1563–1572, 2013, doi: 10.1016/j.compbiomed.2013.08.003.

[16]. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 2560–2567, doi: 10.1109/IJCNN.2016.7727519.

[17]. T. Araújo, G. Aresta, et al., "Classification of breast cancer histology images using Convolutional Neural Networks," PLoS One, vol. 12, no. 6, p. e0177544, Jun. 2017, doi: 10.1371/ journal. pone. 0177544.

[18]. K. Das, S. Conjeti, A. G. Roy, J. Chatterjee, and D. Sheet, "Multiple Instance Learning Of Deep Convolutional Neural Networks For Breast Histopathology Whole Slide Classification," 2018 IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018), no. Isbi, pp. 578–581, 2018, doi: 10.1109/ISBI.2018.8363642.

[19]. S. A. Adeshina, A. P. Adedigba, A. A. Adeniyi, and A. M. Aibinu, "Breast Cancer Histopathology Image Classification with Deep Convolutional Neural Networks," 2018 14th International Conference on Electronics Computer and Computation (ICECCO), 2018, pp. 206–212, doi: 10.1109/ICECCO.2018.86346 90.

[20]. A. Mahbod, I. Ellinger, R. Ecker, Ö. Smedby, and C. Wang, "Breast Cancer Histological Image Classification Using Fine-Tuned Deep Network Fusion," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Note Bioinformatics), vol. 10882 LNCS, pp. 754–762, 2018, doi: 10.1007/978-3-319-93000-8_85.

[21]. C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0

[22]. H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," J. Healthc. Eng., vol. 2019, pp. 1–11, Nov. 2019, doi: 10.1155/2019/4253641.

[23]. S. Dabeer, M. M. Khan, and S. Islam, "Cancer diagnosis in histopathological image: CNN based approach," Informatics Med. Unlocked, vol. 16, p. 100231, 2019, doi: 10.1016/j.imu.2019.100231.

[24]. M. Gour, S. Jain, and T. Sunil Kumar, "Residual learning based CNN for breast cancer histopathological image classification," Int. J. Imaging Syst. Technol., vol. 30, no. 3, pp. 621–635, Sep. 2020, doi: 10.1002/ima.22403

Author's details

M. Tech scholar, **Anshul Jaswal**
Department of Computer Science, Himachal Pradesh Technical University, Himachal Pradesh, India, anshuljaswal4@gmail.com

Assistant Professor **Kumari Archana**
Department of Computer Science, Himachal Pradesh Technical University, Himachal Pradesh, India, niralaarchana.1991@gmail.com