

Computer-Aided Breast Cancer Diagnosis

Subjects: Computer Science, Artificial Intelligence | Medical Informatics

Contributor: Xin Yu Liew, Jérémie Clos

A computer-aided diagnosis (CAD) expert system is a powerful tool to efficiently assist a pathologist in achieving an early diagnosis of breast cancer. This process identifies the presence of cancer in breast tissue samples and the distinct type of cancer stages. In a standard CAD system, the main process involves image pre-processing, segmentation, feature extraction, feature selection, classification, and performance evaluation. Breast cancer can be distinguished as benign (non-cancerous) and malignant (cancerous/metastatic) tumours. Benign tissue refers to changes in normal tissue of breast parenchyma, which does not relate to the development of malignancy. Contrarily, malignant tissue can be categorised into two types: in-situ carcinoma and invasive carcinoma.

Keywords: machine learning ; deep learning ; computer aided diagnosis ; breast cancer ; histopathology images ; classification ; medical imaging

1. Introduction

'Cancer' is a term used when a cell divides abnormally or uncontrollably, which can happen in various parts of the body. Amongst the distinct types of cancer, the most common type of cancer happening in females is breast cancer. According to the World Health Organisation (WHO), breast cancer is the most frequent cancer among women, affecting 2.1 million women each year. About 627,000 women died from breast cancer in 2018, which accounted for around 15% of all cancer deaths among women ^[1].

However, it has been proven that an early detection of breast cancer can significantly increase the chances of successful treatment plan and ensure a long-term survival of the patients ^[2]. According to the most common procedure, a 'two-week wait' is the procedure to diagnose breast cancer ^[3]. The standard procedure to diagnose breast cancer by pathologists usually requires extensive microscopic assessment. Therefore, having an automated solution like a computer-aided diagnosis (CAD) system not only contributes to an easier diagnostic process, but also reduces the subjectivity in diagnosis.

With the advanced development of artificial intelligence, many machine learning techniques have been applied for CAD systems. This technique can potentially outperform humans and learn more efficiently with time, therefore integrating machine learning in diagnosis can supply useful knowledge to assist pathologists in evaluating and analysing enormous amounts of medical data ^[4]. It could also speed up the process due to the capability to process large data much faster than manual diagnosis by a pathologist ^[4]. Breast cancer diagnosis can be considered as a classification problem in machine learning, in which the result indicates which class of cancer it belongs to.

Conventionally, several popular machine learning algorithms applied to classification problems include naïve Bayes ^[5], artificial neural network ^[6], support vector machine (SVM) ^[7], and many more. Recently, deep learning methods were introduced to improve on conventional machine learning methods by extracting information automatically as part of the learning process, leading to undoubtedly better solutions ^[8]. Deep learning was shown to outperform state-of-the-art methods in many fields of medical imaging analysis tasks.

Breast cancer varies based on which part of the breast tissue becomes cancerous. Commonly, breast cancer starts in the cells that line the ducts of the breast; however, it may also grow in different areas of the breast such as the lobules, milk ducts or sometimes in between tissues, as illustrated in **Figure 1** ^[9].

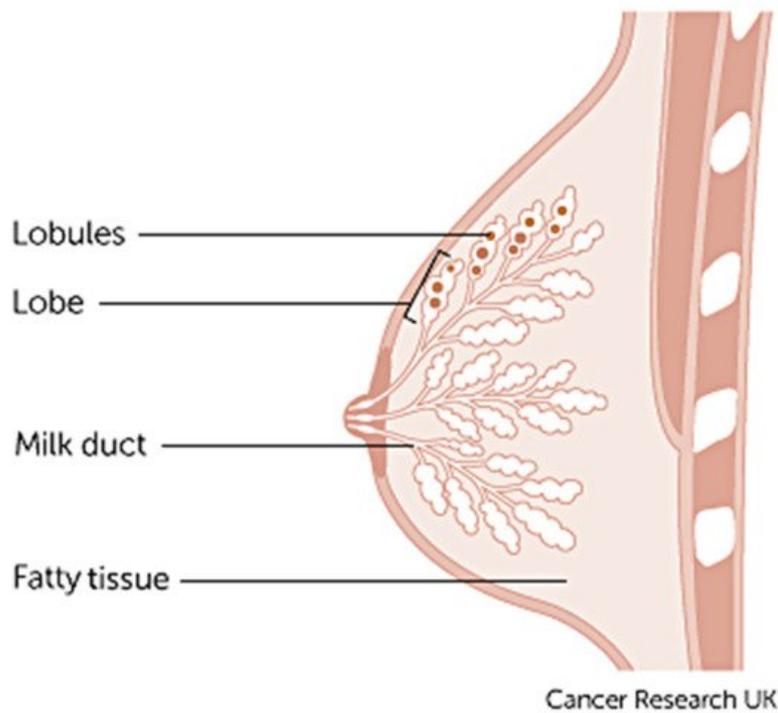


Figure 1. Anatomy of the breast credits to Cancer Research UK ^[9].

The term 'breast cancer' refers to a malignant tumour that has developed from cells in the breast that are considered cancerous and cause danger to health. The stage of this cancer is usually expressed as a number on a scale of 0 through IV, with stage 0 describing non-invasive cancers that are still within their original location and stage IV describing invasive cancers that have spread outside the breast ^[10]. In cases where cancer is detected, but no cancer cells are visible in the lymph glands, the breast cancer is of a lower risk. When spreading occurs, it carries a substantial risk of death, meaning that the cancer cells from the breast tissue have broken away, which can be carried to nearby lymph nodes by the lymph fluid (fluid that gathers waste products and drains into veins to be removed) ^[10].

Breast cancer can be distinguished as benign (non-cancerous) and malignant (cancerous/metastatic) tumours. Benign tissue refers to changes in normal tissue of breast parenchyma, which does not relate to the development of malignancy ^[11]. Contrarily, malignant tissue can be categorised into two types: in-situ carcinoma and invasive carcinoma. Additionally, in some cases benign breast tumours can be further divided into four subclass types, adenosis, fibroadenoma, phyllodes tumour, and tubular adenoma, whereas malignant breast tumours can be further divided into ductal carcinoma, lobular carcinoma, medullary carcinoma, mucinous carcinoma, tubular carcinoma, and papillary carcinoma ^[12].

Histopathology (histology) image samples of breast lesions are obtained through either needles or surgical operation, which are then later processed and allocated to a glass slide to undergo a staining process. Currently, histopathological images play a vital role in cancer diagnosis because of the large amount of information they provide for medical image analysis ^[13]. Whole-slide images (WSI) can have multiple regions of breast lesion tissue, whereas microscopy images are patches derived from WSI, each representing one type of breast lesion only. In this paper we have chosen to study histopathology images of breast cancer in developing a machine learning based CAD system.

The main contribution of this paper is to discuss the process, methods, comparisons, and remarks on developing a CAD expert system for breast cancer. The rest of the research paper is organized as follows: Section 2 explains the publicly available datasets for breast cancer histopathology images. The process of using a computer-aided expert system using histopathology images is presented in Section 3, which includes techniques employed in (1) image pre-processing, (2a) conventional CAD methods that employ segmentation, feature extraction, feature selection (dimension reduction) and classification; (2b) deep-learning-based CAD and (3) Performance evaluation. Finally, Section 4 discusses the conclusion and future directions for researchers are given Section 5.

2. Datasets for Breast Cancer Classification

In the field of medical image analysis, machine learning methodologies applied for histopathological images are developing rapidly. To obtain a large and representative annotated dataset to develop a machine learning method for CAD system is a challenging task ^[14]. Recently, there has been a rise in public challenges for breast cancer diagnosis which

has attracted many researchers to this area of study. This section describes various publicly accessible datasets to assist future research and development.

For binary classification, there are two categories of benign and malignant to determine cancer or non-cancerous. The category benign type of breast tumour consists of adenosis (A), The malignant type of breast tumour consists of ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). This dataset is the most used dataset by many researchers for CAD breast cancer in histopathology images [\[14\]\[15\]\[16\]\[17\]\[18\]\[19\]\[20\]\[21\]\[22\]\[23\]\[24\]\[25\]\[26\]\[27\]\[28\]](#).

Bioimaging Challenge 2015 dataset [\[29\]](#): This dataset contains 269 images of haematoxylin and eosin (H&E)-stained breast cancer histology images with image size of 2048×1536 pixels. To further classify, the non-cancerous categories can be categorized as normal and benign, while the cancerous ones can be categorized as in situ carcinoma and invasive carcinoma. The training set has a total of 249 images form by 55 normal class, 69 benign class, 63 in situ carcinoma class, and 62 invasive carcinoma class, while the test set has a total of 20 images with 5 images for each class. Additionally, there is an extended test set with more diversity provided with a total of 16 images available.

BACH (BreAst Cancer Histology) dataset [\[30\]](#): The ICIAR 2018 challenge resulted in the BreAst Cancer Histology (BACH) image dataset, which is an extended version of the Bioimaging 2015 breast histology classification challenge dataset with similar image sizes and magnification levels [\[29\]](#). The dataset has a total number of 400 images, respectively classified to a total number of 100 normal class, 100 benign class, 100 in situ carcinoma class, and 100 invasive carcinoma class. The test set has a total of 100 images without any labels.

The Cancer Metastases in Lymph Nodes Challenge breast cancer metastasis detection dataset combines two datasets collected from CAMELYON16 and CAMELYON17 challenges, with each image approximately 1×105 by 2×105 pixels at the highest resolution. Each image is annotated with a binary label for classification, showing normal and presence of metastases tissue. There are two sets of training datasets, the first has a total number of 170 images, formed of 100 normal class and 70 metastases class, while the second has a total number of 100 images formed of 60 normal class and 40 metastases class. This version is extended from the CAMELYON16 which include patients testing for breast cancer from the CAMELYON16 challenge with an additional three medical centres from the Netherlands, specifically: slides from 130 lymph node resections from Radboud University Medical Center in Nijmegen (RUMC), 144 from Canisius-Wilhelmina Hospital in Nijmegen (CWZ), 129 from University Medical Center Utrecht (UMCU), 168 from Rijnstate Hospital in Arnhem (RST), and 140 from the Laboratory of Pathology East-Netherlands in Hengelo (LPON) [\[31\]](#).

Thus, this version of dataset is derived from the CAMELYON dataset with a total number of 327.680 histopathologic scans of lymph node sections images, each in the size of 96×96 px pixels. Like the CAMELYON dataset, each image is annotated with binary label for classification, showing normal and presence of metastases tissue. The main difference and advantage of this dataset is that it is bigger than CIFAR10, smaller than ImageNet, additionally it is trainable on a single GPU to able to achieve competitive scores in the CAMELYON16 tasks of cancer detection and WSI diagnosis. PCam contributed by supplying the segmented tissue parts that separated tissue and background from the whole slide images.

The conference ICPR 2012 supplied the MITOS dataset benchmark that consists of 50 histopathology images of haematoxylin and eosin (H&E)-stained slides of breast cancer images from 5 different breast biopsies at $40\times$ magnification level. However, this dataset is too small to produce an exceptionally reliable performance and the robustness of the diagnosis system is limited. Therefore, an extended version of the dataset (MITOS-ATYPIA-14) was presented at ICPR 2014.

MITOS-ATYPIA-14 dataset [\[32\]](#): (H&E)-stained slides of breast cancer images with the size of 1539×1376 pixels at $20\times$ and $40\times$ magnification level. There is a training set with a total number of 1200 images acquired from 16 different biopsies and testing set with a total number of 496 images acquired from 5 different breast biopsies. The dataset consists of a significantly diverse variation of stained images in many conditions to elevate the challenge to achieve a more exceptional performance.

TUPAC16 dataset [\[33\]](#): The dataset consists of a total number of 73 breast cancer histopathology images at $40\times$ magnification level from three pathology centres in the Netherlands. The dataset is composed of 23 test images with a size of 2000×2000 pixels and 50 training images with a size of 5657×5657 pixels collected from two separate pathology centres. The images contained in the training dataset are later cropped randomly to the size of 2000×2000 pixels. The dataset can be obtained from <http://tupac.tue-image.nl/node/3> (accessed on 16 March 2021).

UCSB bio segmentation benchmark (UCSB-BB) [34]: This dataset contains 50 haematoxylin and eosin (H&E)-stained histopathology images used in breast cancer cell detection with the size of 896×768 pixels and a ground truth table. Each image is annotated with binary label for classification, it contains half malignant class and half benign class.

References

1. Ferlay, J.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available online: (accessed on 16 March 2021).
2. Sizilio, G.R.M.A.; Leite, C.R.M.; Guerreiro, A.M.G.; Neto, A.D.D. Fuzzy Method for Pre-Diagnosis of Breast Cancer from the Fine Needle Aspirate Analysis. *Biomed. Eng. Online* 2012, 11.
3. Cancer Research UK Breast Cancer Statistics|Cancer Research UK. Available online: (accessed on 16 March 2021).
4. Robertson, S.; Azizpour, H.; Smith, K.; Hartman, J. Digital Image Analysis in Breast Pathology—From Image Processing Techniques to Artificial Intelligence. *Transl. Res.* 2018, 194, 19–35.
5. Krawczyk, B.; Schaefer, G.; Woźniak, M. A Hybrid Cost-Sensitive Ensemble for Imbalanced Breast Thermogram Classification. *Artif. Intell. Med.* 2015, 65.
6. Bhardwaj, A.; Tiwari, A. Breast Cancer Diagnosis Using Genetically Optimized Neural Network Model. *Expert Syst. Appl.* 2015, 42.
7. Chen, H.L.; Yang, B.; Liu, J.; Liu, D.Y. A Support Vector Machine Classifier with Rough Set-Based Feature Selection for Breast Cancer Diagnosis. *Expert Syst. Appl.* 2011, 38.
8. Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* 2015, 521, 436–444.
9. What Is Breast Cancer?|Cancer Research UK. Available online: (accessed on 16 March 2021).
10. Breast Cancer Organization. What Is Breast Cancer?|Breastcancer.Org. 2016. pp. 1–19. Available online: (accessed on 16 March 2021).
11. Alom, M.Z.; Yakopcic, C.; Nasrin, M.S.; Taha, T.M.; Asari, V.K. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *J. Digit. Imaging* 2019, 32.
12. Akram, M.; Iqbal, M.; Daniyal, M.; Khan, A.U. Awareness and Current Knowledge of Breast Cancer. *Biol. Res.* 2017, 50, 33.
13. Khan, S.U.; Islam, N.; Jan, Z.; Ud Din, I.; Rodrigues, J.J.P.C. A Novel Deep Learning Based Framework for the Detection and Classification of Breast Cancer Using Transfer Learning. *Pattern Recognit. Lett.* 2019, 125.
14. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng.* 2009, 2.
15. Bayramoglu, N.; Kannala, J.; Heikkila, J. Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification. In Proceedings of the International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2440–2445.
16. Vo, D.M.; Nguyen, N.Q.; Lee, S.W. Classification of Breast Cancer Histology Images Using Incremental Boosting Convolution Networks. *Inf. Sci.* 2019, 482.
17. Kassani, S.H.; Kassani, P.H.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Classification of Histopathological Biopsy Images Using Ensemble of Deep Learning Networks. *arXiv* 2019, arXiv:1909.11870.
18. Murtaza, G.; Shuib, L.; Mujtaba, G.; Raza, G. Breast Cancer Multi-Classification through Deep Neural Network and Hierarchical Classification Approach. *Multimed. Tools Appl.* 2020, 79.
19. Toğaçar, M.; Özkurt, K.B.; Ergen, B.; Cömert, Z. BreastNet: A Novel Convolutional Neural Network Model through Histopathological Images for the Diagnosis of Breast Cancer. *Phys. A Stat. Mech. Its Appl.* 2020, 545.
20. Alkassar, S.; Jebur, B.A.; Abdullah, M.A.M.; Al-Khalidy, J.H.; Chambers, J.A. Going Deeper: Magnification-Invariant Approach for Breast Cancer Classification Using Histopathological Images. *IET Comput. Vis.* 2021, 15, 151–164.
21. Chan, A.; Tuszynski, J.A. Automatic Prediction of Tumour Malignancy in Breast Cancer with Fractal Dimension. *R. Soc. Open Sci.* 2016, 3.
22. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. In Proceedings of the International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; IEEE: Piscataway, NJ, USA, 2016.

23. Han, Z.; Wei, B.; Zheng, Y.; Yin, Y.; Li, K.; Li, S. Breast Cancer Multi-Classification from Histopathological Images with Structured Deep Learning Model. *Sci. Rep.* 2017, 7.
24. Bardou, D.; Zhang, K.; Ahmad, S.M. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. *IEEE Access* 2018, 6.
25. Gandomkar, Z.; Brennan, P.C.; Mello-Thoms, C. MuDeRN: Multi-Category Classification of Breast Histopathological Image Using Deep Residual Networks. *Artif. Intell. Med.* 2018, 88.
26. Budak, Ü.; Cömert, Z.; Rashid, Z.N.; Şengür, A.; Çibuk, M. Computer-Aided Diagnosis System Combining FCN and Bi-LSTM Model for Efficient Breast Cancer Detection from Histopathological Images. *Appl. Soft Comput. J.* 2019, 85.
27. George, K.; Faziludeen, S.; Sankaran, P.; Paul, J.K. Deep Learned Nucleus Features for Breast Cancer Histopathological Image Analysis Based on Belief Theoretical Classifier Fusion. In *Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Kochi, India, 17–20 October 2019*; IEEE: Piscataway, NJ, USA, 2019.
28. Sudharshan, P.J.; Petitjean, C.; Spanhol, F.; Oliveira, L.E.; Heutte, L.; Honeine, P. Multiple Instance Learning for Histopathological Breast Cancer Image Classification. *Expert Syst. Appl.* 2019, 117.
29. Araujo, T.; Aresta, G.; Castro, E.; Rouco, J.; Aguiar, P.; Eloy, C.; Polonia, A.; Campilho, A. Classification of Breast Cancer Histology Images Using Convolutional Neural Networks. *PLoS ONE* 2017, 12.
30. Aresta, G.; Araújo, T.; Kwok, S.; Chennamsetty, S.S.; Safwan, M.; Alex, V.; Marami, B.; Prastawa, M.; Chan, M.; Donovan, M.; et al. BACH: Grand Challenge on Breast Cancer Histology Images. *Med. Image Anal.* 2019, 56.
31. Bándi, P.; Geessink, O.; Manson, Q.; Van Dijk, M.; Balkenhol, M.; Hermesen, M.; Ehteshami Bejnordi, B.; Lee, B.; Paeng, K.; Zhong, A.; et al. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Trans. Med. Imaging* 2019, 38.
32. MITOS-ATYPIA-14 Grand Challenge. Available online: (accessed on 17 March 2021).
33. Veta, M.; Heng, Y.J.; Stathonikos, N.; Bejnordi, B.E.; Beca, F.; Wollmann, T.; Rohr, K.; Shah, M.A.; Wang, D.; Rousson, M.; et al. Predicting Breast Tumor Proliferation from Whole-Slide Images: The TUPAC16 Challenge. *Med. Image Anal.* 2019, 54.
34. Drelie Gelasca, E.; Obara, B.; Fedorov, D.; Kvilekval, K.; Manjunath, B.S. A Biosegmentation Benchmark for Evaluation of Bioimage Analysis Methods. *BMC Bioinform.* 2009, 10.

Retrieved from <https://encyclopedia.pub/entry/history/show/25713>