

RECENT CNN-BASED TECHNIQUES FOR BREAST CANCER HISTOLOGY IMAGE CLASSIFICATION

ArunaDevi Karuppasamy^{1,*}, Abdelhamid Abdesselam¹, Rachid Hedjam¹,
Hamza Zidoum¹, and Maiya Al-Bahri²

¹ Department of Computer Science, Sultan Qaboos University,

² Department of Pathology, Sultan Qaboos University Hospital,
Muscat, Sultanate of Oman

ABSTRACT: Histology images are extensively used by pathologists to assess abnormalities and detect malignancy in breast tissues. On the other hand, Convolutional Neural Networks (CNN) are by far, the privileged models for image classification and interpretation. Based on these two facts, we surveyed the recent CNN-based methods for breast cancer histology image analysis. The survey focuses on two major issues usually faced by CNN-based methods namely the design of an appropriate CNN architecture and the lack of a sufficient labelled dataset for training the model. Regarding the design of the CNN architecture, methods examining breast histology images adopt three main approaches: Designing manually from scratch the CNN architecture, using pre-trained models and adopting an automatic architecture design. Methods addressing the lack of labelled datasets are grouped into four categories: methods using pre-trained models, methods using data augmentation, methods adopting weakly supervised learning and those adopting feedforward filter learning. Research works from each category and reported performance are presented in this paper. We conclude the paper by indicating some future research directions related to the analysis of histology images.

Keywords: Breast cancer; CNN; Deep learning; Histology image Classification; Machine learning.

التقنيات الحديثة المعتمدة على شبكة CNN لتصنيف صور سرطان الثدي

أرون ديفي كربوسامي* و عبد الحميد عبدالسلام و راشد حجام و حمزة زيدوم و ميا البحري

الملخص: يستخدم علماء الأمراض صور الأنسجة على نطاق واسع لتقييم التشوهات واكتشاف الأورام الخبيثة في أنسجة الثدي. من ناحية أخرى تعد الشبكات العصبية التلافيفية إلى حد بعيد النماذج المميزة لتصنيف الصور وتفسيرها؛ بناءً على هاتين الحقيقتين قمنا بمسح الطرق الحديثة المعتمدة على CNN لتحليل صورة أنسجة سرطان الثدي. يركز المسح على مشكلتين رئيسيتين عادة ما تواجههما الأساليب المعتمدة على CNN وهما تصميم بنية CNN المناسبة وعدم وجود مجموعة بيانات مصنفة كافية لتدريب النموذج. فيما يتعلق بتصميم بنية CNN، تتبنى طرق فحص صور أنسجة الثدي ثلاثة مناهج رئيسية: التصميم اليدوي لهندسة CNN، استخدام نماذج مدربة مسبقاً واعتماد تصميم هندسي تلقائي. يتم تجميع الطرق التي تعالج نقص مجموعات البيانات المصنفة في أربع فئات: الأساليب التي تستخدم النماذج المدربة مسبقاً، والطرق التي تستخدم زيادة البيانات، والطرق التي تعتمد التعلم الخاضع للإشراف الضعيف، وتلك التي تتبنى التعلم التوضيحي. في هذه الورقة، نقوم بعرض الأعمال البحثية من كل فئة والأداء المبلغ عنه. نختم الورقة بالإشارة إلى بعض اتجاهات البحث المستقبلية المتعلقة بتحليل صور الأنسجة.

الكلمات المفتاحية: سرطان الثدي؛ الشبكات العصبية التلافيفية؛ تصنيف صور الأنسجة؛ تعلم الآلة؛ التعلم الآلي.

*Corresponding author's e-mail: s121293@student.squ.edu.om



1. INTRODUCTION

Breast cancer is the most common cancer among women in developed countries (IARC, 2020) (Intercollegiate and Network, 2005) (Sickles, 1997). Currently, pathologists examine histology images to interpret the tissue appearance and detect abnormal conditions. These images usually have a large resolution, different magnification scales, various acquisition processes, and different staining methods, which set them apart from other medical images (See Figure 1). Besides, the result of a histology image analysis may vary from different pathologists, and sometimes from different analyses done by the same pathologist (Gurcan *et al.*, 2009, Germain, 2017), which leads to inter-observer and intra-observer variability and uncertainties in the decision-making process. Hence, there is a need for developing effective and efficient computer-aided techniques to analyze and interpret histology images.

Machine Learning (ML) is a discipline of Artificial Intelligence that has proven to be very efficient in solving classification problems. It allows computers to learn from data without being explicitly programmed. Supervised learning is an ML category in which a model is trained on a set of inputs (features) with known outcomes (labels). Once the training is completed, the defined model will be capable of making predictions when fed with new unseen data. Traditionally, features are explicitly selected and extracted by the user (see Figure 2).

Extracted features are mostly related to intensity, morphology, and texture (Boucheron, 2008) (Dundar *et al.*, 2011) (Liu *et al.*, 2011). Examples of these features include: (i) Gabor-wavelet filters, and density measures features (Marugame *et al.*, 2009), (ii) Gaussian Markov random field and fractal dimension features (Al-Kadi, 2010), (iii) intensity, morphological, co-occurrence, and run-length features (Irshad, 2013),

(iv) Haar-like features and Gaussian filters features (Vink *et al.*, 2013), and (v) local binary patterns, morphometric features, entropy features and gray-Level co-occurrence matrix (Tashk *et al.*, 2015) (Bruno *et al.*, 2016), (Peikari *et al.*, 2017).

With the advent of Deep Learning (DL), image classification and object recognition achieved unprecedented levels of accuracy (Lecun *et al.*, 2015). Similarly, its application to healthcare especially for analyzing medical images showed excellent performance in solving various problems such as segmentation, interpretation and registration (Kim M *et al.* 2019). The success of DL is mainly due to the ability of modern computers to process huge amounts of data, and extract features automatically at different abstraction levels. Convolutional Neural Network (CNN), is a DL technique specially adapted to process images and videos. The first CNN, LeNet (Y. LeCun *et al.*, 1998) was proposed in 1998 to classify handwritten digits, but the real surge in CNN popularity started with AlexNet (Krizhevsky *et al.*, 2012) when it won the ImageNet challenge 2012 and since then several deeper and more performant CNNs were developed (VGG16 (K. Simonyan and A. Zisserman, 2014), GoogleNet (C. Szegedy *et al.*, 2015 and 2016), and ResNet(He *et al.*, 2016).

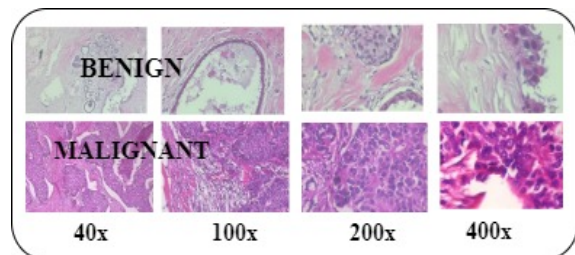


Figure 1. Benign and malignant samples from BreakHis dataset of various magnification (F. Spanhol, 2016).

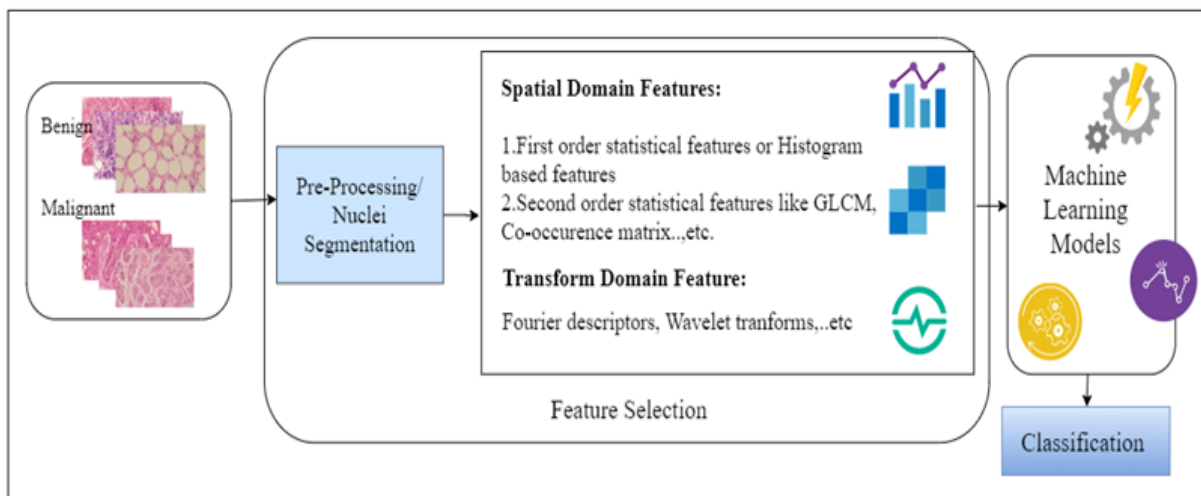


Figure 2. The traditional machine learning process for breast cancer histology image classification.

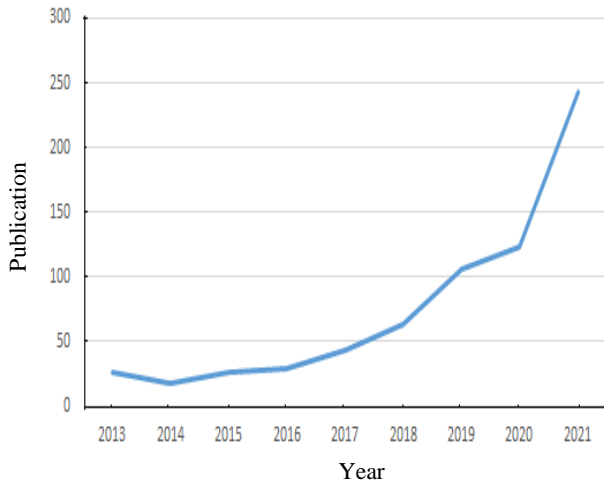


Figure 3. The number of papers addressing the problem of breast cancer histology image classification published between 2013 and 2021.

CNN-based methods, including those addressing histopathology image classification, usually face two major challenges: 1)- there is no systematic approach for designing an appropriate CNN architecture and 2)-there is a need for large-scale training dataset that is, unfortunately not available for many real-world problems, including those related to the medical field.

This review aims at exploring how these two major challenges have been addressed in the literature, especially by those research works using breast cancer histopathological images. To the best of our knowledge, this is the first article conducting such a study.

The selection of the reviewed papers is based on the search keywords “Deep Learning, Histopathology and Breast cancer” in Springer, ScienceDirect, and IEEE Explore during the period 2013-2021. The total number of publications in the breast cancer histopathology classification gradually increased during that period (See Figure3). The search returned with 987 research papers, then we removed 136 duplications, and reported only the papers related to the classification of breast cancer using CNN.

The remainder of the paper is organized as follows: Section 2 describes the main components of a CNN model and introduces the datasets used by the methods included in this survey.

Section 3 describes the proposed models for breast cancer histology image classification and categorizes them based on the way their architecture is designed.

Section 4 describes methods proposed to address the problem of the lack of labelled datasets. The performance, in terms of accuracy and training time, of all surveyed works, were reported.

The paper concludes by summarizing the main characteristics of the proposed approaches and presenting some future research directions related to the analysis of histology images.

2. CONVOLUTIONAL NEURAL NETWORK FOR HISTOLOGY IMAGE CLASSIFICATION

CNNs are capable of learning automatically features from raw input data. They consist of several feature extraction layers where the first layers extract basic features such as edges and blobs and deeper layers extract more complex and abstract features. A CNN includes three main building blocks: (i) Convolutional and pooling layers, (ii) Fully-Connected layers and (iii) a classification layer (See Figure 4). The convolution operation is the most expensive computation of a CNN, it extracts information by convolving the input data with a set of filters. A feature map is obtained by applying an activation function to the output of the convolution operation. This ensures non-linearity and enforces a sparse representation of the feature map. Early models used Sigmoid and hyperbolic tangent functions as activation functions (LeCun *et al.*, 2012). Recent works adopted other activation functions such as ReLU (Krizhevsky *et al.*, 2012), Leaky ReLU(LReLU) (Maas *et al.*, 2013), Parametric-ReLU(PReLU) (He *et al.*, 2015), Randomised-ReLU (RReLU) (Xu *et al.*, 2015), S-shaped ReLU(S-ReLU) (Jin *et al.*, 2016), Maxout (Goodfellow *et al.*, 2013), and Exponential Linear Unit (ELU) (Trottier *et al.*, 2017). A pooling operation usually follows convolutional and activation operations to reduce the size of the feature maps. There are different pooling types in CNN (Maas *et al.*, 2013) such as max-pooling, mean-pooling, stochastic pooling, spatial pyramid pooling, and deformation pooling. Max pooling (Boureau, Bach, *et al.*, 2010)(Boureau, Ponce, *et al.*, 2010) is the most used method; it consists of replacing a pooling window by the maximum value in the window. Then come a set of fully-connected layers, where each neuron in a layer is connected to all neurons of the following layer. The last feature map is

passed to the first fully connected layer in a form of a 1-D vector (Kröse *et al.*, 1993) to produce an output more appropriate to the classification task that is performed either by applying a function like a Softmax for multi-label classification or a Sigmoid for binary classification. One of the traditional classification models such as Logistic Regression (LR), Support Vector Machines (SVM) or Random Forests (RF) can also perform the classification.

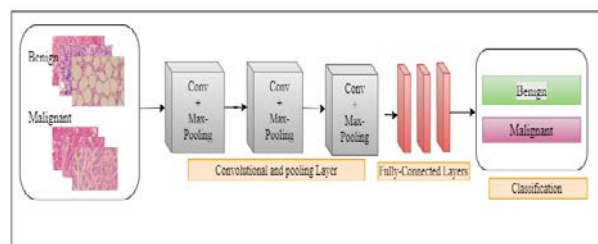


Figure 4. Convolutional neural network for breast cancer histology image classification.

Table 1. Dataset and links for breast cancer histology classification.

Dataset	Links
ICPR-2012	http://ludol7.free.fr/mitos_2012/
TUPAC2016	http://tupac.tue-image.nl/
BreakHis 2016	https://web.inf.ufpr.br/vri/databases/breastcancer-histopathological-
Camelyon16	https://camelyon16.grand-challenge.org/
Patchcamelyon	https://patchcamelyon.grand-challenge.org/Download/
BACH-ICIAR18	https://iciar2018-challenge.grand-challenge
Kaggle data repository	https://www.kaggle.com/c/histopathology-cancer-detection

Most of the models discussed in this paper are trained and tested on breast cancer datasets publicly available on the Internet. Their links are shown in Table 1.

Three main performance metrics are used by the methods surveyed in this paper. Classification accuracy, F-score metric and Area Under Curve (AUC), where

$$1. \text{ Accuracy} = \frac{\# \text{ of correct classifications}}{\text{Total number of tested images}} \quad (1)$$

$$2. \text{ F-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (2)$$

where precision is the fraction of correctly classified images among those labelled as positive, and recall is the fraction correctly classified images out of all positive images in the test dataset.

$$3. \text{ Area Under the Curve (AUC) indicates how well the positive classes are classified compared to negative classes.} \quad (3)$$

F-score and AUC are more appropriate for evaluating the classification performance on datasets with imbalanced classes.

3. CNN ARCHITECTURE DESIGN

Based on the way architectures are designed, we can categorize CNN-based methods for classifying cancer histology images into three categories: (i) Manually-designed CNN architectures, (ii) Pre-trained CNN architectures, and (iii) Automatically-designed CNN architectures.

- **Manually-Designed CNN Architectures**

Most CNN-based histology image classification methods are of this category as shown in **Table 2**. They use relatively shallow architectures (small number of layers) and backpropagation algorithms for filter

learning. They are trained mostly on datasets of limited size using a limited number of epochs and achieve moderate to good levels of performance.

(Circsan *et al.*, 2013) proposed a 6-layer CNN (four Convolutional and pooling layers and two fully-connected layers) to detect mitosis in a dataset of 50 images. It was trained for one day with less than 30 epochs in an optimized GPU implementation. This method won the ICPR2012 competition by achieving an F1-score of 0.78. (Cruz-Roa *et al.*, 2014) designed a 3-layer CNN to automatically detect invasive ductal carcinoma from whole-slide images (WSI) of breast cancer. The dataset consists of 162 WSIs from which about 165000 patches of size 100x100 were extracted and used for model training and testing. The proposed CNN was trained with 25 epochs and achieved an F1-score of 0.71. (Wang *et al.*, 2014) proposed a cascade ensemble that combines hand-crafted features (morphology, colour, and texture features) with features extracted from a 3-layer CNN. These two feature sets were passed to a random forest classifier to detect mitotic nuclei on the ICPR dataset consisting of 50 High Power Field images (scanned into 2084x2084 pixels RGB images). The model was trained for 9 epochs and it required nearly 18 hours without GPU implementation. They reported an F1-score of 0.73. (Litjens *et al.*, 2016) proposed a patch-based CNN model to classify prostate cancer and detect micro-metastases and macro-metastases breast cancer in sentinel lymph nodes. Their dataset consists of 271 WSIs from which 2.6 million patches of size 128x128 were obtained through data augmentation. For metastases detection, the CNN was trained with 12 epochs, each epoch requiring 200 minutes on GeForce GTX970, and obtained an AUC of 0.88. (Araújo *et al.*, 2017) proposed an 8-layer CNN model that classifies histology images into four categories: i) normal tissue, ii) benign lesion, iii) in situ carcinoma, and iv) invasive carcinoma. Their dataset consists of 269 high resolution images (2049x1536 pixels) obtained from the Bio-imaging 2015 breast histology classification challenge dataset from which 70000 (512x512) patches were extracted. The CNN model was trained with 50 epochs and achieved a validation accuracy of 80.6% for binary classification (non-carcinoma and carcinoma). The accuracy improved to 83.3% when using the CNN as a feature selection tool and an SVM as a classifier. (Bayramoglu *et al.*, 2017), proposed two 6-layer CNN architectures for breast cancer histology image classification characterized by their independence of the magnification level. A single task CNN (3 convolutional layers followed by 3 fully connected layers) was used to predict malignancy and a multi-task CNN (3 convolutional layers followed by two sequential fully connected layers) is used to predict both malignancy and image magnification level simultaneously. The CNNs were trained with images from different magnification levels (40x, 100x, 200x, and 400x) of BreakHis dataset.

Table 2. Overview of manually-designed cnn architectures used in breast cancer histology image classification.

Reference	CNN Architecture	Dataset	Performance
(Cirecsan <i>et al.</i> , 2013)	6-layer CNN(4-Convolutional layers and 2 Fully-connected layers)	ICPR2012(50 images)	F1-score :0.78
(Cruz-Roa <i>et al.</i> , 2014)	3-layer CNN (2-Convolutional layers and 1-Fully-connected layer)	Hospital of University of Pennsylvania and The cancer Institute of New Jersey dataset (162 images)	F-score : 0.71 Acc: 84.23%
(Wang <i>et al.</i> , 2014)	3-layer CNN (2-Convolutional layers and a Fully-connected layer)	ICPR dataset (50 images)	F-score: 0.73
(Litjens <i>et al.</i> , 2016)	6-layer CNN(4-Convolutional layers and 2 Fully-connected layers)	Radboud University Medical Center dataset(173 images)	AUC: 0.88
(Araújo <i>et al.</i> , 2017)	8-layer CNN(5 Convolutional layers and 3 Fully-Connected layers)	Bioimaging 2015 breast histology classification challenge dataset (269 images)	Acc: multi-class:77.8% and binary-class: 83.3%
(Bayramoglu <i>et al.</i> , 2017)	Two 6-layer CNNs	BreakHis dataset (7909 images)	Average recognition rate:83.25%
(Wahab <i>et al.</i> , 2017)	Two-Phase CNN model	ICPR2012 dataset(50 images) and TUPAC16 dataset(73 images)	F-score : 0.79
(Cruz-Roa <i>et al.</i> , 2018)	2-layer CNN	Hospital of the Univ. of Pennsylvania(239 images), Case Western Reserve Univ(110 images), Cancer Institute of New Jersey(52 images), and The Cancer Genome Atlas(195 images)	Acc: 0.90 Dice Coefficient: 76%
(Roy <i>et al.</i> , 2019)	Patch-Based classifier with CNN (6-convolutional layers and 2-fully connected layers)	ICIAR-2018 Dataset (500 images)	Acc: Binay-class: 92.5% Multi-Class:90%

The authors reported an average recognition accuracy of 83.25% for single CNN and averages of 82.13% and 80.10% for benign or malignant classification and magnification detection respectively. (Wahab *et al.*, 2017) proposed a two-phase CNN to handle imbalanced histology data. The first phase of the CNN was trained with 80x80 pixels non-mitosis patches to classify them into three imbalanced classes: easy, normal and hard using 5 epochs. The resulting classes are then balanced and combined with augmented mitosis patches to be used for retraining the CNN with 25 additional epochs during 15 hours on i7-3770 CPU of 3.4GHz to classify the patches into mitotic and non-mitotic nuclei. The patches were extracted from two datasets (50 HPFs(high-power fields) from the ICPR2012 context dataset and 73 HPFs from the TUPAC16 dataset). The authors reported an F1-score of 0.79. (Cruz-Roa *et al.*, 2018) proposed a framework called HASHI for invasive breast cancer detection in WSI. The training data is obtained by applying a regular sampling on 600 labelled WSIs. The patches resulting from a pseudorandom sampling of the new unseen WSIs are passed to a 2-layer CNN classifier; resulting predictions are used to build an interpolated probability map. Dense sampling is further applied to regions with high uncertainty, which produces an improved probability map estimation. The authors reported an

AUC of 0.90. (Roy *et al.*, 2019) proposed an 8-layer CNN model for binary (Non-Malignant and Malignant) and multi-class (normal, benign, in-situ and invasive) classification on the ICIAR-2018 dataset, which consists of 500 images (400 for training and 100 for testing) of size 2048x1536. The model was trained with about 70 epochs on 128GB RAM of CPU based system with Xenon processor to yield an accuracy of 92.5% and 90% for binary classification and multi-classification, respectively.

• Pre-trained CNN Architectures

This category of methods implements transfer learning, where features learned to solve a problem from one domain constitute the baseline for the features used to solve a problem from a different but related domain (Mehra, 2018). These techniques take an existing CNN architecture usually trained on a large dataset, and re-train it on a smaller dataset related to the problem under investigation. During the training, only the weights of the last few layers are updated using the backpropagation approach for filter learning. These CNNs are usually deep (more than seven layers), initially trained for a long time and with a large number of epochs which make them very performant in solving the original problem. Retraining takes usually, much less time (less epochs) on much smaller datasets and resulting models

Table 3. Overview of pre-trained CNN architectures used in breast cancer histology image classification.

Reference	CNN Architecture	Dataset	Performance
(Fabio Alexandre Spanhol <i>et al.</i> , 2016)	AlexNet	BreakHis dataset (7909 images)	Acc:85.6 (40x), 83.0(100x), 83.1(200x), 80.8 (400x)
(Fabio A Spanhol <i>et al.</i> , 2017)	BVLC CaffeNet model (pre-trained AlexNet as a CNN) along with DeCaf features	BreakHis dataset (7909 images)	Acc: 84.3 (40x), 84.7 (100x), 84.1(200x), 81.6 (400x)
(Song <i>et al.</i> , 2017)	Component Selective Encoding CNN (combination of Fisher vector and VggNet19)	BreakHis dataset (7909 images)	Acc: 87.5 (40x), 88.6(100x), 85.5(200x), 85.0(400x)
(Bejnordi <i>et al.</i> , 2017)	CAS-CNN(a combination of wide-ResNet and VggNet)	Radboud University Medical Centre Dataset(224 images)	AUC: 0.96(Binary) Acc: 89%(Binary) Acc: 81.3%(multi-class)
(Han <i>et al.</i> , 2017)	Class Structure-based Deep Convolutional Neural Network (GoogleNet)	BreakHis dataset (7909 images)	Acc: 95.8 (40x), 96.9 (100x), 96.7 (200x), 94.9 (400x)
(Vang <i>et al.</i> , 2018)	InceptionV3 and Dual-path network	ICIAR-2018 Dataset(400 images)	Acc: 87.5%
(Motlagh <i>et al.</i> , 2018)	ResNet152	TMA database(6,402) and BreakHis dataset(7909 images)	TMA - Acc:99.8% BreakHis Acc:98.7%(Binary) Acc:94.6%(multi-class)
(Vesal <i>et al.</i> , 2018)	ResNet50	BACH-2018 dataset (320 images)	Acc: 97.50%
(Kausar <i>et al.</i> , 2019)	VggNet16	ICIAR 2018 (400 images) and BreakHis dataset(7909 images)	Acc: 98.2%(ICIAR2018) 96.85%(40x-BreakHis)
Budak <i>et al.</i> (2019)	FCN based on AlexNet and Bi-LSTM (Bidirectional Long Short-Term Memory)	BreaKHis dataset(7909 images)	Acc: 95.69(40x), 93.61(100x), 96.32(200x), 94.29(400x).
Alom <i>et al.</i> (2019)	Inception Recurrent Residual CNN (IRRCNN)	Bioimaging Challenge 2015	Acc: 99.05% (for binary) and 98.59%(for multi)
Vo <i>et al.</i> (2019)	Inception and Gradient boosting trees	BreaKHis dataset(7909 images)	Acc: 93.5(40x), 95.3(100x), 96.1(200x), 91.1(400x)
(Anwar <i>et al.</i> , 2020)	ResNet50, HoG, WPD	BreakHis dataset (7909 images)	Acc: 97.10(40x), 97.56(100x), 96.41(200x), 94.32(400x)
(Chakraborty <i>et al.</i> , 2020)	ResNet50	Kaggle data repository (220,025 patches of images)	Acc:96.48% F1-score:96.32%
Laxmisagar and Hanumantharaju (2021)	MobileNet2.10ex	Bioimaging Challenge 2015	Acc: 88.92%
Munien and Viriri (2021)	EfficientNet	ICIAR2018 dataset	Acc: 98.33%

are usually of high performance. The most re-used CNN architectures are AlexNet (Krizhevsky *et al.*, 2012), VggNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy *et al.*, 2015), and ResNet (He *et al.*, 2016) (See Table 3).

AlexNet is an 8-layer CNN trained on an ILSVRC dataset (Deng *et al.*, 2009). The dataset consists of images of 1000 different classes. 1.2 million images are used for training, 50,000 images for validation and 150,000 images for testing. It took 6 days for a 90-epoch training and achieved a winning error rate of 15.3%. The second version of AlexNet trained on CIFAR-10 dataset that contains 60000 images (50000 for training and 10000 for testing) from 10-classes has also been used as a pre-trained model (Krizhevsky and Hinton, 2010) (Krizhevsky, 2009). VggNet is available in two architectures, 16 layers and 19 layers; they were trained on the ILSVRC14 challenge dataset for 74 epochs and won second place by obtaining an error rate of 14.7% and 7.3% respectively. GoogLeNet is a 22-layer architecture also trained on the ILSVRC14 challenge dataset and won the competition by obtaining an error rate of 6.7%. ResNet is a deep network that is available in different versions, the most common ones have 50, 101, and 152 layers respectively. The ResNet50 was trained on the ImageNet dataset and achieved an error rate of 5.25%. The ResNet101 model was trained using 80k iterations for detection and segmentation of the COCO 2015 challenge dataset and won the COCO 2015 competitions by achieving a 28% relative improvement on object detection. The ResNet152 model was trained on the ILSVRC15 challenge dataset for 60×10^4 iterations and won the ILSVRC15 competitions by achieving an error rate of 3.57%. Classification of histology images with pre-trained networks yields better performance compared with manually designed CNNs.

Fabio Alexandre Spanhol *et al.* (2016) applied the AlexNet pre-trained on the CIFAR-10 dataset on a multi-magnification dataset called BreakHis that was introduced by (Fabio A Spanhol *et al.*, 2015). A subset of 1000 patches of size 64x64 was used to train the model for about 80,000 iterations to obtain image-level accuracy of $85.6 \pm 4.8\%$ (40x), $83.0 \pm 3.9\%$ (100x), $83.1 \pm 1.9\%$ (200x), and $80.8 \pm 3\%$ (400x). (Fabio A Spanhol *et al.*, 2017) applied a modified version of AlexNet (the order of pooling and normalization layer was exchanged) pre-trained on the ImageNet dataset. On the breaks dataset their model achieved an image-level accuracy of $84.3 \pm 2.9\%$ (40x), $84.7 \pm 4.4\%$ (100x), $84.1 \pm 1.5\%$ (200x), and $81.6 \pm 3.7\%$ (400x). (Song *et al.*, 2017) proposed a model, which extracts features from the final convolutional layer of VggNet19, encoded by a fisher-vector to feed an SVM classifier. The proposed model was evaluated on the BreakHis dataset and obtained a binary classification accuracy of 87.5 ± 1.6 (40x), 88.6 ± 3.6 (100x), 85.5 ± 2.0 (200x), and 85.0 ± 4.6 (400x). (Bejnordi *et al.*, 2017) proposed a cascaded CNN (CAS-CNN) model, which

is a combination of a modified version of ResNet called Wide-ResNet (Zagoruyko and Komodakis, 2016) and VggNet. The proposed CAS-CNN model was trained and tested on 224 WSI (100 normal/benign, 69 DCIS, and 55 IDC WSIs) from Radboud University Medical Center, Netherland. An Accuracy of 89% was obtained for binary classification (Benign and Cancer) and 81% for multi-classification (Benign, Ductal Carcinoma, and Invasive Ductal Carcinoma). (Han *et al.*, 2017) proposed a Class Structure-based Deep CNN (CSDCNN), which is based on a pre-trained GoogLeNet for binary and multi-classification. The data augmentation method was applied on BreakHis dataset to increase the training images. Moreover, an over-sampling based on Gaussian distribution was applied to deal with the imbalance of the classes. The training took 10 hours on the augmented dataset on Intel i7- NVIDIA Quadro K22200 GPU and achieved image-level accuracies of 95.8 ± 3.1 (40x), 96.9 ± 1.9 (100x), 96.7 ± 2.0 (200x), and 94.9 ± 2.8 (400x). (Vang *et al.*, 2018) used the pre-trained GoogLeNet model on the ICIAR-2018 challenge dataset, which contains 400 microscopy images of size 2040x1536 pixels. The training was conducted on 4 GPUs (2 NVIDIA TitanX GPUs and 2 NVIDIA GTX 1080Ti) for 30 epochs to achieve an accuracy of 87.5% for a multi-classification (normal, benign, in-situ carcinoma and invasive carcinoma). (Motlagh *et al.*, 2018) used the pre-trained ResNet152 to extract hierarchical features on TMA (Tissue Micro Array) and BreakHis datasets. The model was trained on ASUS GeForce GTX1080 for 3000 epochs. It achieved an accuracy of 98.7% and 96.4%, for binary and multi-classification, respectively. (Vesal *et al.*, 2018) applied ResNet50 on the BACH-2018 challenge dataset, which consists of 400 WSI of size 2040x1536 pixels. From the 320 training images, 67,200 patches of size 512x512 were extracted for training. The model was trained for about 100 epochs and achieved an accuracy of 97.5%. (Kausar *et al.*, 2019) proposed a modified version of VggNet16 by replacing the last three fully connected layers with a global pooling layer. The input of the model consists of Haar wavelet-decomposed images. ICIAR2018 and BreakHis datasets have been used to train the model for about 150 epochs on NVIDIA Tesla M40 GPU to achieve an accuracy of 98.2% and 96.85% on the two datasets, respectively. (Budak *et al.*, 2019) used an integrated model combining AlexNet, and Bi-LSTM (Bidirectional Long Short-Term Memory) to classify BreakHis dataset. The authors reported that pre-trained AlexNet was re-trained for 20 epochs in the NVIDIA Quadro P6000 GPU, and achieved the accuracy of 95.69% (40x), 93.61% (100x), 96.31% (200x), and 94.29% (400x). (Alom *et al.*, 2019) proposed a model that combines Inception-V4, ResNet and Recurrent CNN. The model was re-trained for 150 epochs on the BreakHis dataset using a 56G of RAM GPU with NVIDIA GeForce GTX-980 Ti. It achieved an accuracy of 97.95% (40x), 97.57% (100x), 97.32 (200x), and 97.36 % (400x). (Vo *et al.*, 2019) proposed

a model based on Inception-v3 and ResNet-152. It was trained on BreaKHis dataset for 50 epochs on GeForce GTX 1080 Ti that yields an accuracy of 93.5% (40x), 95.3%(100x), 96.1%(200x), and 91.1%(400x). (Anwar *et al.*, 2020) proposed a model that combines features produced by ResNet50, wavelets packet decomposition, and histograms of the gradient. PCA is then applied to reduce the feature dimensionality. A Data Augmentation was applied to the BreaKHis dataset to obtain 237270 training patches. The authors reported accuracies of 97.10 (40x), 97.56 (100x), 96.41 (200x), and 94.32 (400x). (Chakraborty *et al.*, 2020) proposed a Dual Channel Residual CNN (DCRCNN) which decomposes input images into two separate stain channels (Eosin and Hemotoxylin) (Ruifrok, 2019). Stain channels are fed into two separate ResNet50; resulting features are then fused for classification. The proposed model was applied on Kaggle data repository (*Kaggle Dataset*, 2018), which contains nearly 220,025 images of size 96x96. The authors trained the model for 50 epochs and achieved an accuracy, and F1-score of 96.48% and 96.32%, respectively. Laxmisagar and Hanumantharaju (2021) proposed a model that incorporated MobileNet2.10e with fully convolutional deep neural network on bio-imaging challenge 2015 dataset. This model was trained for 280 epochs in Google colab with GPU 2496 CUDA, and achieved an accuracy of 88.92%. (Munien and Viriri, 2021) proposed a model based on EfficientNet-B2 trained for 50 epochs in Google colab with NVIDIA Tesla K80 GPU, and achieved an accuracy of 98.3% in ICIAR2018 dataset.

• Automatically-Designed CNN Architecture

Network Architecture Search (NAS) is a state-of-the-art approach for designing efficient CNN-based solutions; It was recently introduced to automate and optimize the hyper-parameters and the architecture design of neural networks (Zoph and Le, 2017)(Wistuba, 2019). NAS methods can be categorized according to three dimensions (Elsken *et al.* 2019): search space, search strategies and performance evaluation strategies. The search space defines which architectures a NAS method might discover, the search strategy defines the way the search space is explored to discover an optimized architecture in terms of some parameters such as number of layers, number of filters, size of the filters and pooling window, and strides while a performance evaluation strategy aims at estimating the performance of a neural network architecture in terms of predefined criteria such as the accuracy. The reader is referred to (Elsken *et al.* 2019) for more details on these strategies.

(Koohbanani *et al.*, 2018) proposed a model that applies the NAS approach to design a neural network for breast cancer histology image classification. In this work, a Bayesian optimization approach (Zoph and Le, 2017) was applied to estimate the optimal values of the main hyper-parameters of the model (filter size, dropout layer, learning rate, and momentum). The

optimized model was used to detect metastasis on CAMELYON16 dataset that consists of 400 WSI (270 WSI for training and 130 WSI for testing). The authors reported an AUC of 0.99. (Oyelade and Ezugwu, 2021) proposed a NAS based model using Ebola Optimization Search Algorithm for search space to optimize the hyper-parameters of deep neural networks that applied on the breast cancer histopathology images of BreaKHis and BACH dataset. The optimized model trained for 500 epochs on both datasets and attained a 100% accuracy

4. LACK OF LABELED HISTOLOGY DATASETS

To achieve high prediction performance, a CNN model requires large labelled training datasets which is, unfortunately, not available for many problem domains including the medical domain. There are two main reasons behind the lack of labelled histology image datasets: 1) Labeling histology images is time-consuming and requires medical experts, which make the task very expensive. 2) Medical data, in general, are usually confidential and therefore not easily available to the research community. Several approaches have been adopted to address this problem; the most common methods used for histology image classification are patch generation, data augmentation and pre-trained CNNs. The two other methods addressing this problem are weakly supervised learning and feed-forward filter learning.

• Commonly Used Approaches Dealing with Small Histology Datasets

Histology images are in general of large size (Wide Slide Images 2040 x 1536 pixels) which allows increasing the number of the dataset samples by generating thousands of smaller labelled images (patches) from the original WSI dataset (Litjens *et al.*, 2016) (Araújo *et al.*, 2017) (Wahab *et al.*, 2017) (Vesal *et al.*, 2018). Data augmentation is another commonly adopted method for increasing the size of image datasets. It consists of generating additional labelled images by applying various image transformations, such as rotation, translation, zooming, and cropping, to the original dataset. (Fabio Alexandre Spanhol *et al.*, 2016) (Vang *et al.*, 2018) (Anwar *et al.*, 2020). Finally, pre-trained models are often adopted because they can lead to high accuracy even when re-trained on relatively small datasets (Song *et al.*, 2017) (Bejnordi *et al.*, 2017).

• CNN using Weakly-Supervised Learning

Weakly supervised learning includes three categories (Zhou, 2018): 1) incomplete supervision, in which only a subset of training data is given with labels; 2) inexact supervision, in which the training data are given with only coarse-grained labels, and 3) inaccurate supervision, in which the given labels are not always ground truth (Albarqouni *et al.*, 2016).

Table 3. Overview of methods dealing with the lack of large labelled datasets in breast cancer histology classification.

Reference	Supervision/ Filter Learning	Dataset	Performance
(Albarqouni <i>et al.</i> , 2016)	Inaccurate supervision	MICCAI-AMIDA13 challenge dataset (311 images)	F1-Score: 0.74 AUC: 0.869
(Das <i>et al.</i> , 2018)	Inexact supervision (MIL)	BreakHis dataset (7909 Images)	Acc:89.52(40x), 89.06 (100x), 88.84(200x), 87.67(400x)
(Sudharshan <i>et al.</i> , 2019)	Inexact supervision (MIL)	BreakHis dataset (7909 Images)	Acc:86.9(40x), 85.7 (100x), 85.7 (200x), 83.4 (400x)
(Jaiswal <i>et al.</i> , 2019)	Incomplete supervision	PatchCamelyon (327680 patches of images)	AUC: 0.98
(Shi <i>et al.</i> , 2016)	PCA	MGH dataset (66 images)	AUC: 0.88 Acc: 81.51±2.91
(Huang <i>et al.</i> , 2017)	PCA	Netherland Cancer Institute (778 images), Vancouver General Hospital (664 images) and Public dataset. [Stanford Tissue Microarray Consortium Web Portal and The UCSB Bio-Segmentation Benchmark dataset] (45 images)	Acc: 85%

Incomplete supervision is implemented via two major techniques, namely active learning which involves a human expert to label selected unlabeled instances considered as the most valuable, and semi-supervised learning that makes use of labelled data and some basic assumptions about data distribution to predict labels of unlabeled data. For more details, the reader is referred to (Zhou, 2018). Inexact supervision uses the multiple-instance learning (MIL) technique introduced by (Dietterich *et al.*, 1997). In MIL, CNNs inputs consist of multiple instances of images (called a bag of images), and the bag labels are obtained using two methods that are instance-based and bag-based methods. For more details, the reader is referred to (Herrera *et al.*, 2016). In inaccurate supervision, training data may contain erroneous labelling. A typical scenario is when some of the labels are obtained through the crowdsourcing technique introduced by (Brabham, 2008). In this technique, non-expert users registered to a common platform provide labels for the unlabeled data. All published works applying methods grouped under this category are using the backpropagation approach for filter learning.

Albarqouni *et al.* (2016) proposed a 5-layer CNN model (AggNet) for mitotic detection. A crowdsourcing platform, called Crowdflower was developed to annotate the unlabeled images. Users registered on this platform are able to access the samples of labelled images and participate in the annotation process. This model was trained for 100 epochs in GeForce GT 750M on the MICCAI-AMIDA13 challenge dataset that consists of 318 HPF images for training and 22 HPF images for testing, along with 5500 labelled patches from the Crowdflower platform (550 patches annotated by 10

participants). The authors reported an AUC of 0.87. (Das *et al.*, 2018) proposed a MIL-CNN with new multiple instance-pooling layers that extract features from the final fully connected layer for classification. A bag with multiple instances is obtained by extracting regions of interest (ROI) from WSI. Each patient is assigned a bag containing 25 to 60 randomly chosen patches of size 224x224 extracted from the BreakHis dataset. Their model achieved patient-level accuracies of 89.52% (40x), 89.6% (100x), 88.4% (200x) and 87.67 (400x). (Sudharshan *et al.*, 2019) applied on BreakHis dataset MILCNN model initially proposed by (Sun *et al.* 2016). In their work, they considered two different settings; in the first set, each patient is considered as a bag which is labelled by its diagnosis while in the second set 1000 patches are randomly extracted from each image to constitute bags assigned the image label. The model was trained for 80 to 120 epochs and achieved patient-level accuracies of $86.9 \pm 5.4\%$ (40x), $85.7 \pm 4.8\%$ (100x), $85.9 \pm 3.9\%$ (200x), and $83.4 \pm 5.3\%$ (400x) and image-level accuracies of $86.1 \pm 4.2\%$ (40x), $83.8 \pm 3.1\%$ (100x), $80.2 \pm 2.6\%$ (200x), and $80.6 \pm 4.6\%$ (400x). (Jaiswal *et al.*, 2019) applied a modified DenseNet-201 proposed by (G. Huang and Weinberger, 2018) where they substituted the last two fully connected layers with a global-max and global-average pooling layers. The model was trained with a small labelled dataset then the trained model is applied on a batch of the unlabeled dataset to obtain a pseudo-labelled dataset that was used to explore discriminating features of the unlabeled dataset. During the training, the backpropagation algorithm optimizes the objective function that combines the loss from the predictions of the pseudo-labelled dataset with that from the labelled dataset.

Their model was applied to PCam (PatchCamelyon) dataset that includes labelled and unlabeled images consisting of 327,680 patches of size 96x96 pixels. They trained the model for 5 to 7 epochs on NVIDIA Tesla P100 and reported an AUC of 0.98.

- **Methods Using Forward Filter Learning**

In this approach, predefined filters are used in the convolutional layers instead of learning them through the traditional backpropagation algorithm. These methods are trained for a short period at a low computational cost and they can perform well with small-sized datasets. Two pioneer research works have adopted this approach for learning filters, namely ScatNet (Bruna, J. and Mallat, S. 2013) and PCANet (Chan *et al.*, 2015). The formerly used wavelet operators to define the filters while the latter used PCA method to define them.

Shi *et al.* (2016) proposed a method inspired by PCANet called Color pattern Random Binary Hashing-Based PCANet (C-RBH-PCANet). In this method, filters learned using a PCA-based technique were applied to a 2-block CNN (each including a convolutional layer, non-linearity layer and a pooling layer). They reported an AUC and accuracy of 0.88 and 81.51%, respectively on a dataset that consists of 1000 patches from 66 WSI of size 800x1800 from the MGH (Massachusetts General Hospital) dataset. (Huang *et al.*, 2017) proposed a 5-layer CNN model. The filters were learned using principal component analysis on the public datasets (University of Pittsburgh School of Medicine, UPSM and the UCSB Bio-Segmentation, UCSB-BS Benchmark dataset), and a sample of different types of histology images obtained through the Google search engine. They achieved an accuracy of 85% on randomly selected 12,000 patches of size 50x50 pixels from three datasets (Netherland Cancer Institute dataset, Vancouver General Hospital dataset, and the UPSM and UCSB-BS public datasets).

5. DISCUSSION

In this paper, the focus was on how CNN-based methods addressed the two major challenges faced by any CNN-based model namely the design of an appropriate CNN architecture and the acquisition of an adequate labelled dataset.

Regarding the CNN architecture design, the survey showed that the early methods adopted a handcrafted approach that builds from scratch the adequate CNN architecture based on trial and error. These methods are characterized by their relatively small depths (mostly less than 7 layers), reasonable training time and epoch number and moderate to good accuracy (mostly from 70% to 90%). Tuning the values of the CNN hyperparameters such as the number of layers, number of filters and their sizes, etc. is tedious and time-consuming. Besides, the identified architectures are difficult to adapt to new datasets. With the event of performant deep CNNs such as AlexNet (8 layers), VggNet (16 and 19 layers), GoogleNet (22 layers) and

ResNet (152 layers), transfer learning became the preferred approach for designing image classifiers. The proposed methods attained better accuracies on breast cancer histology images (from mid-80% to mid-90%) with much less design effort, as most of the time the design consists of retraining and /or modifying the few last layers of a predesigned CNN architecture. The main drawback of these methods is the large number of learned weights that have to be stored. Very recently Network Architecture Search (NAS) techniques have been proposed. They aim to automatically design optimized CNN architectures with efficient hyperparameter setting that suits the dataset under consideration. Designed CNN can achieve very high classification accuracies, the only work we found addressing breast cancer histology image classification reported an accuracy of 99%. The main drawback of these methods resides in the significant search time required to identify the appropriate architecture.

Regarding the acquisition of an adequate labelled dataset, the survey showed that proposed breast cancer histology image classification methods addressed the problem in different ways. Some methods make use of the large size of the Whole-Slide Images or WSI (about 2040 x 1536) to extract a significant number of small labelled patches of a few hundred pixels in each dimension. The resulting images are then used to train and test the designed CNN models which lead to achieving accuracies in the range of mid-80% to mid-90%. Other methods use pre-trained models that usually do not need a large number of images for training because of the huge size of the dataset used to design the pre-trained CNN. When there is a need to further increase the size of the dataset, two approaches are applied: data augmentation and weakly supervised learning. In data augmentation, new images/patches are generated from collected labelled data by applying image transformations such as rotations, zooming, illumination variations, cropping etc. Obtained classification accuracies are usually very high (from 85% to 99%). Weakly supervised learning acquires labelled images through three techniques crowdsourcing, bagging or the semi-supervised technique proposed by Chappelle *et al.* Methods adopting these techniques achieved good classification accuracies (from 83% to 89%). Recently, feedforward approaches for learning filters were proposed to overcome the drawbacks of the traditional backpropagation method including the need for large labelled datasets. Proposed methods achieve good accuracy levels (from 81% to 85%) with a few layers and reasonable size of labelled datasets.

6. CONCLUSION

In this paper, we surveyed recent CNN-based methods for breast cancer histology image classification. The surge in high-performance GPU computing coupled with the availability of a large amount of training data has made CNN achieve state-of-the-art performance in

the classification tasks. Although the feature representation from pre-trained models has been learned from natural images, these models achieve high classification accuracy on histological images and therefore are widely applied. Recently, training with few samples is getting more attention from the research community using histology images because of the limited size of this type of dataset. This survey showed that the use of automatic architecture design and feedforward filter learning approaches have not been fully explored for the analysis of breast histology images despite the impressive results obtained by the former in classifying images and the promising perspectives offered by the latter to deal with the lack of labelled histology image datasets. Therefore, we believe these two directions need to be further explored and could be of great potential and value in the future.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding this publication.

FUNDING

No Funding was received for this study.

REFERENCE

- Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., and Navab, N. (2016). AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images. *IEEE Transactions on Medical Imaging*, 35(5), 1313–1321. <https://doi.org/10.1109/TMI.2016.2528120>
- Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., and Asari, V. K. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of Digital Imaging*, 32(4), 605–617.
- Anwar, F., Attallah, O., Ghanem, N., and Ismail, M. A. (2020). Automatic breast cancer classification from histopathological images. *2019 International Conference on Advances in the Emerging Computing Technologies, AECT 2019*, 16–21. <https://doi.org/10.1109/AECT47998.2020.9194194>
- Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., and Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PLOS One*, 12(6).
- Bayramoglu, N., Kannala, J., and Heikkila, J. (2017). Deep learning for magnification independent breast cancer histopathology image classification. *Proceedings - International Conference on Pattern Recognition*, 2440–2445. <https://doi.org/10.1109/ICPR.2016.7900002>
- Bejnordi, B. E., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., van Ginneken, B., Karssemeijer, N., Litjens, G., and van der Laak, J. (2017). *Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images*. 1–13. <https://doi.org/10.1117/1.JMI.4.4.044504>
- Boureau, Y.-L., Bach, F., LeCun, Y., and Ponce, J. (2010). Learning mid-level features for recognition. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2559–2566.
- Boureau, Y.-L., Ponce, J., and LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 111–118.
- Brabham, D. C. (2008). Crowdsourcing as a model for problem-solving: An introduction and cases. *Convergence*, 14(1), 75–90. <https://doi.org/10.1177/1354856507084420>
- Budak, Ü., Cömert, Z., Rashid, Z. N., Sengür, A., and Çibuk, M. (2019). Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. *Applied Soft Computing*, 85, 105765.
- Chakraborty, S., Aich, S., Kumar, A., Sarkar, S., Sim, J. S., and Kim, H. C. (2020). Detection of cancerous tissue in histopathological images using Dual-Channel Residual Convolutional Neural Networks (DCRCNN). *International Conference on Advanced Communication Technology, ICACT, 2020*, 197–202. <https://doi.org/10.23919/ICACT48636.2020.9061289>
- Ciresan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 411–418.
- Cruz-Roa, A., Basavanthally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., and Madabhushi, A. (2014). *Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks*. 9041(216), 904103. <https://doi.org/10.1117/12.2043872>
- Cruz-Roa, A., Gilmore, H., Basavanthally, A., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A., and González, F. (2018). High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. *PLOS One*, 13(5).
- Das, K., Conjeti, S., Roy, A. G., Chatterjee, J., and Sheet, D. (2018). Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. *Proceedings - International Symposium on Biomedical Imaging, 2018-April(Isbi)*, 578–581. <https://doi.org/10.1109/ISBI.2018.8363642>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer*

Vision and Pattern Recognition, 248–255.

Dietterich, T. G., Lathrop, R. H., and Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1–2), 31–71. [https://doi.org/10.1016/s0004-3702\(96\)00034-3](https://doi.org/10.1016/s0004-3702(96)00034-3)

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. (2013). Maxout Networks. *ArXiv Preprint ArXiv:1302.4389*.

Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., and Li, S. (2017). Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports*, 7(1), 1–10. <https://doi.org/10.1038/s41598-017-04075-z>

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.

Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., Vluymans, S., Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., and Vluymans, S. (2016). Multiple Instance Learning. In *Multiple Instance Learning*. https://doi.org/10.1007/978-3-319-47759-6_2

Huang, Y., Zheng, H., Liu, C., Ding, X., and Rohde, G. K. (2017). Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. *IEEE Journal of Biomedical and Health Informatics*, 21(6), 1625–1632.

<https://doi.org/10.1109/JBHI.2017.2691738>

IARC. (2020). *GLOBOCAN 2020: New Global Cancer Data*. <https://www.uicc.org/news/globocan-2020-new-global-cancer-data>

Intercollegiate, S., and Network, G. (2005). *SIGN Guideline No. 84 Management of breast cancer in women. December*.

Jaiswal, A. K., Panshin, I., Shulkin, D., Aneja, N., and Abramov, S. (2019). *Semi-Supervised Learning for Cancer Detection of Lymph Node Metastases*. <http://arxiv.org/abs/1906.09587>

Jin, X., Xu, C., Feng, J., Wei, Y., Xiong, J., and Yan, S. (2016). Deep learning with s-shaped rectified linear activation units. *Thirtieth AAAI Conference on Artificial Intelligence*.

Kaggle dataset. (2018). <https://www.kaggle.com/c/histopathologic-cancer-detection>

Kausar, T., Wang, M. J., Idrees, M., and Lu, Y. (2019). HWCNN: Multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network. *Biocybernetics and Biomedical Engineering*, 39(4), 967–982. <https://doi.org/10.1016/j.bbe.2019.09.003>

Koohbanani, N. A., Qaisar, T., Shaban, M.,

Gamper, J., and Rajpoot, N. (2018). Significance of hyperparameter optimization for metastasis detection in breast histology images. In *Computational Pathology and Ophthalmic Medical Image Analysis* (pp. 139–147). Springer.

Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. ... *Science Department, University of Toronto, Tech. ...*, 1–60. <https://doi.org/10.1.1.222.9220>

Krizhevsky, A., and Hinton, G. (2010). Convolutional deep belief networks on cifar-10. *Unpublished Manuscript*, 1–9. <http://scholar.google.com/scholar?hl=en&btnG=Search&dq=intitle:Convolutional+Deep+Belief+Networks+on+CIFAR-10#0>

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Kröse, B., Krose, B., van der Smagt, P., and Smagt, P. (1993). *An introduction to neural networks*.

Laxmisagar, H. S., and Hanumantharaju, M. C. (2021). Design of an Efficient Deep Neural Network for Multi-level Classification of Breast Cancer Histology Images. In *Intelligent Computing and Applications* (pp. 447–459). Springer.

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9–48). Springer.

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Kaa, C. H. Van De, Bult, P., and Ginneken, B. Van. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Nature Publishing Group, January*, 1–11. <https://doi.org/10.1016/j.media.2020.101813>

Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *Proc. Icml*, 30(1), 3.

Mehra, R. (2018). Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express*, xxx. <https://doi.org/10.1016/j.ict.2018.10.007>

Motlagh, N. H., Jannesary, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., and Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. *BioRxiv*, 242818.

Munien, C., and Viriri, S. (2021). Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets. *Computational Intelligence and Neuroscience*, 2021.

Oyelade, O. N., and Ezugwu, A. E. (2021). A bioinspired neural architecture search based convolutional neural network for breast cancer detection using histopathology images. *Scientific Reports*, 11(1), 1–28. <https://doi.org/10.1038/s41598-021-98978-7>

Roy, K., Banik, D., Bhattacharjee, D., and

- Nasipuri, M. (2019). Patch-based system for Classification of Breast Histology images using deep learning. *Computerized Medical Imaging and Graphics*, 71, 90–103. <https://doi.org/10.1016/j.compmedimag.2018.11.003>
- Ruifrok, A. C. (2019). quantification of histochemical staining by color deconvolution. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Shi, J., Wu, J., Li, Y., Zhang, Q., and Ying, S. (2016). Histopathological image classification with color pattern random binary hashing-based PCANet and matrix-form classifier. *IEEE Journal of Biomedical and Health Informatics*, 21(5), 1327–1337.
- Sickles, E. A. (1997). *Breast Cancer Screening Outcomes in Women Ages 40 – 49: Clinical Experience With Service Screening Using Modern Mammography*. 99–104.
- Simonyan, K., and Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 1–14. <https://doi.org/10.1016/j.infsof.2008.09.005>
- Song, Y., Zou, J. J., Chang, H., and Cai, W. (2017). Adapting fisher vectors for histopathology image classification. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 600–603.
- Spanhol, F. (2016). *A Dataset for Breast Cancer Histopathological A Dataset for Breast Cancer Histopathological Image Classification*. 63(November 2015), 1455–1462. <https://doi.org/10.1109/TBME.2015.2496264>
- Spanhol, Fabio A, Cavalin, P. R., Oliveira, L. S., Petitjean, C., and Heutte, L. (2017). *Deep Features for Breast Cancer Histopathological Image Classification*. 1868–1873.
- Spanhol, Fabio A, Oliveira, L. S., Petitjean, C., and Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7), 1455–1462.
- Spanhol, Fabio Alexandre, Oliveira, L. S., Petitjean, C., Heutte, L., Cavalin, P. R., Oliveira, L. S., Petitjean, C., Heutte, L., Sudharshan, P. J., Petitjean, C., Spanhol, F. A., Eduardo, L., Heutte, L., and Honeine, P. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. *Proceedings of the International Joint Conference on Neural Networks, 2016-October*, 2560–2567. <https://doi.org/10.1109/IJCNN.2016.7727519>
- Sudharshan, P. J., Petitjean, C., Spanhol, F., Eduardo, L., Heutte, L., and Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification. *Expert Systems With Applications*, 117, 103–111. <https://doi.org/10.1016/j.eswa.2018.09.049>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.
- Trottier, L., Gigu, P., Chaib-draa, B., and others. (2017). Parametric exponential linear unit for deep convolutional neural networks. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 207–214.
- Vang, Y. S., Chen, Z., and Xie, X. (2018). Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10882 LNCS, 914–922. https://doi.org/10.1007/978-3-319-93000-8_104
- Vesal, S., Ravikumar, N., Davari, A., ... S. E.-... C. I., and 2018, undefined. (n.d.). Classification of breast cancer histology images using transfer learning. *Springer, I*, https://link.springer.com/chapter/10.1007/978-3-319-93000-8_92
- Vo, D. M., Nguyen, N. Q., and Lee, S. W. (2019). Classification of breast cancer histology images using incremental boosting convolution networks. *Information Sciences*, 482, 123–138. <https://doi.org/10.1016/j.ins.2018.12.089>
- Wahab, N., Khan, A., and Lee, Y. S. (2017). Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. *Computers in Biology and Medicine*, 85(November 2016), 86–97. <https://doi.org/10.1016/j.compbiomed.2017.04.012>
- Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., and Madabhushi, A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3), 034003. <https://doi.org/10.1117/1.JMI.1.3.034003>
- Wistuba, M. (n.d.). *A Survey on Neural Architecture Search*.
- Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *ArXiv Preprint ArXiv:1505.00853*.
- Zagoruyko, S., and Komodakis, N. (2016). Wide Residual Networks. *British Machine Vision Conference 2016, BMVC 2016, 2016-Septe*, 87.1–87.12. <https://doi.org/10.5244/C.30.87>
- Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, 5(1), 44–53.
- Zoph, B., and Le, Q. V. (2017). Neural architecture search with reinforcement learning. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 1–16.