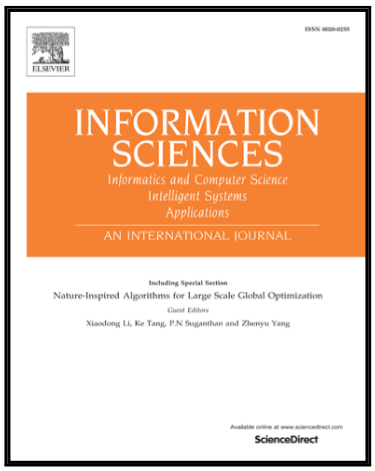
Accepted Manuscript



Classification of Breast Cancer Histology Images using Incremental Boosting Convolution Networks

Duc My Vo, Ngoc-Quang Nguyen, Sang-Woong Lee

PII: S0020-0255(18)31046-6
DOI: https://doi.org/10.1016/j.ins.2018.12.089
Reference: INS 14189

To appear in: *Information Sciences*

Received date: 20 April 2018
Revised date: 22 November 2018
Accepted date: 31 December 2018

Please cite this article as: Duc My Vo, Ngoc-Quang Nguyen, Sang-Woong Lee, Classification of Breast Cancer Histology Images using Incremental Boosting Convolution Networks, *Information Sciences* (2018), doi: https://doi.org/10.1016/j.ins.2018.12.089

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Classification of Breast Cancer Histology Images using Incremental Boosting Convolution Networks

Duc My Vo

Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Korea

Ngoc-Quang Nguyen

Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Korea

Sang-Woong Lee [*]

Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Korea

Abstract

Breast cancer is the most common cancer type diagnosed in women worldwide. While breast cancer can occur in both men and women, it is by far more prevalent in women. Researchers have developed computer-aided systems for efficient diagnosis of breast cancer from histopathological microscopic images. These systems have contributed to increased diagnosis efficiency of biopsy tissue using hematoxylin and eosin stained images. However, most computer-aided diagnosis systems have traditionally used handcrafted feature extraction methods that are both ineffective and time-consuming. In this study, we propose an approach that utilizes deep learning models with convolutional layers to extract the most useful visual features for breast cancer classification. It is shown that these deep learning models can extract better features than handcrafted feature extraction approaches. We also propose a novel boosting strategy to achieve the main goal, whereby the system is efficiently enriched by progressively combining deep learning models (weak classifiers) into a stronger classifier. Our

[1]*: Corresponding author

ACCEPTED MANUSCRIPT

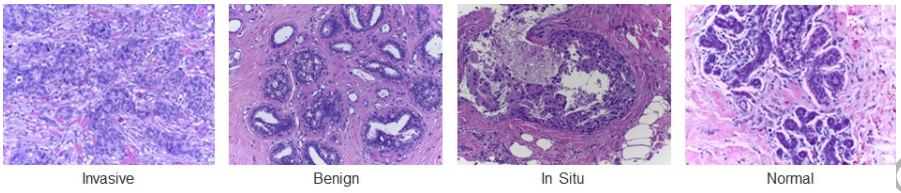


Figure 1: Examples of microscopic biopsy images.

system is used to classify hematoxylin and eosin stained breast biopsy images into two major groups (carcinomas and non-carcinomas) and four classes (normal tissues, benign lesions, in situ carcinomas and invasive carcinomas). We demonstrate applications to breast cancer histopathology images that have been considered challenging to diagnose based on conventional methodologies. Our results demonstrate that our breast cancer classifier with a boosting deep learning model significantly outperforms state-of-the-art methods.

Keywords: Inception network, gradient boosting trees, breast cancer, histopathological microscopic images.

1. Introduction

Breast cancer is still one of the top leading causes of death in women worldwide [27]. To diagnose a wide variety of breast cancer types properly, it is necessary to apply a medical test (commonly performed by a surgeon), followed by a microscopic analysis of breast tissue. In the first stage of this process, the doctor has to cut section biopsy materials and then stain them using hematoxylin and eosin staining. The hematoxylin solution binds deoxyribonucleic acid (DNA) and highlights nuclei, while eosin binds proteins and highlights other structures [43]. In the second stage of this analysis, pathologists evaluate tissue biopsies by visualizing highlighted regions in digitized images using microscopes. The evaluation of tissue biopsies allows the identification of early clues of tissue biopsies. However, professional pathologists must expend considerable time and effort to accomplish this task. The process of breast cancer diagnosis is not only time-consuming and expensive but also strongly depends on the prior

ACCEPTED MANUSCRIPT

knowledge of the pathologist and the consistency of pathologic reports. The average diagnostic accuracy of pathologists is approximately 75% [11].

Fortunately, the development of computer vision and machine learning potentially offers more reliable classification methods for the histological assessment of hematoxylin and eosin stained sections. These methods can automatically classify breast tissues into different categories with high classification rates. Thus, many researchers have developed fast and precise image analysis algorithms for breast cancer detection tasks. However, their results are still far from meeting accepted clinical requirements. For this reason, researchers have been expending most of their efforts into the development of new algorithms for histopathological image analyses [18], [41], [26]. These algorithms aim to achieve the precise classification of breast tissues as normal tissues, nonmalignant (benign) tissues, in situ carcinomas, or invasive carcinomas. In the category of benign lesions, images show changes in the normal structures of breast parenchyma that do not progress to malignancy. In situ carcinoma indicates cells that are restrained inside the mammary ductal–lobular system. Unlike in situ carcinomas, invasive carcinomas present a profile where cells spread beyond the structure of the mammary ductal–lobular system.

One of the challenges in analyzing breast cancer histopathology is to deal with a wide variety of hematoxylin and eosin stained sections, which is attributed to differences among people, different protocols used in labs, skills of pathologists in scanning images, and different staining procedures [30]. Figure 1 shows some breast cancer histopathology images that are considered challenging for classification. Each of them belongs to one of four tissue classification categories, including normal tissues, nonmalignant tumors, in-situ carcinomas, and invasive carcinomas. To overcome this challenging problem, we developed an automated breast cancer detection method to classify breast tissues precisely into the four listed categories above. In particular, our proposed method is synthesized from several novel concepts.

In this research, we present an ensemble of deep convolutional neural networks (DCNNs) trained to extract the most useful visual features from multi-

ACCEPTED MANUSCRIPT

scale training images. The use of DCNNs increases the accuracy of classifying multi-label breast cancers by aggregating multi-scale contextual information. In addition, this network can extract both global and local information from original images. Conversely, by using lower-resolution input images, the receptive field of the network in the original image can be expanded to cover global features more adequately. Furthermore, higher resolution input images are used in our network to extract multi-scale features in local regions. These advantages can be applied effectively in breast cancer detection tasks because breast cancer tumors and cells have a wide variety of shapes, sizes, margins, and densities. First, our network can detect differently sized breast cancer tumors by extracting multi-scale local features that are very important for doctors to determine the stage of breast cancer. In most cases, if the doctor only detects small tumors, the patient has better chances for long-term survival. Otherwise, a larger cancer can be more aggressive and doctors may recommend a mastectomy or chemotherapy before surgery. Second, our network can recognize characteristic abnormalities in breast cancer tissues, such as shapes and margins, based on the detailed information obtained in local regions. The shapes of breast cancer tumors are usually round, oval, lobular, or irregular. Poorly defined or spiculated margins are often worrying signs of breast cancer cells. Most breast lesions and tumors have ill-defined borders and certainly need more investigations. Finally, this ensemble of deep convolutional neural networks is able to extract global information that is used to estimate the tumor densities and the number of breast tissues. This is an important advantage because a high-tumor density in terms of the amounts of fatty elements usually constitutes a highly suspicious sign for breast cancer.

The other challenging problems of breast cancer histopathology image analyses are related to the limited number of available training samples and the imbalanced data problem. First, DCNNs are only effective when the number of available training samples is large enough during a training stage. Conversely, these networks often suffer from overfitting if the training samples are limited. Unlike the case of natural image classification tasks, there are far fewer medical

ACCEPTED MANUSCRIPT

images available for use for the effective training of a deep learning network for medical applications, such as breast cancer detection. This is because of data privacy issues and the increased cost of data collection. Second, like in many other medical image applications, breast cancer detection methods are also associated with the problem of unbalanced training images, because collecting data from patients is not an easy task. To address these problems, we propose a novel boosting method for improving classification performance and for preventing overfitting. This boosting method is based on the combination of a DCNN model and a boosting trees classifier. In this combination, the DCNN model can be used to map high-dimensional inputs to a low-dimensional space of discriminative feature vectors. In fact, the classification performance of each DCNN degrades after a certain number of training iterations owing to the overfitting problem. In this situation, we can stop training the DCNN model early on, and use it to extract low-dimensional discriminated features. In particular, the last convolutional layer of the DCNN model is converted into a one-dimensional feature vector with a length of 1056. Finally, this one-dimensional feature vector is used as the input to the gradient boosting trees classifier. Unlike DCNNs, the boosting trees classifier is able to tackle the challenging problems of the limited number of available training samples and the imbalanced data problem. The boosting trees classifier can further improve the DCNN performance provided that the training data belongs to a low dimensional subspace. This is because the boosting trees classifier is able to convert weak classifiers to strong classifiers by adding more weight to training examples that were misclassified by weaker classifiers in earlier rounds. Furthermore, a data augmentation method is used to tackle imbalanced data, and to increase the available data samples for training the DCNNs and for boosting trees classifiers. For these reasons, the combination of DCNN and a boosting trees classifier leads to a better classification performance even though the number of breast cancer samples is not large, owing to privacy policy constraints and other conditions.

In general, the training process for this method consists of two stages. In the first stage, we aimed at building a set of deep convolutional neural networks

ACCEPTED MANUSCRIPT

trained to extract the most useful visual features from multi-scale training images. Recent research demonstrated that an ensemble of DCNNs can perform better than a single DCNN [34] [33]. Hence, we trained independently multiple DCNNs with different input scales and combined their final feature vectors by using a new boosting strategy. In fact, these advance feature vectors were used as inputs to gradient boosting trees classifiers [14]. In the second training stage, each gradient boosting trees classifier was trained based on the new dataset of deep feature vectors extracted from the corresponding DCNN in the first stage. After the training processes, each boosting trees classifier can achieve a better accuracy rate than its corresponding DCNN classifier. Finally, we applied the majority voting strategy for combining the boosting trees classifiers. This combination resulted in a final boosting classifier that achieved the best classification performance thanks to the improved extraction of multi-scale deep features of breast cancer tumors.

In this study, we tested our proposed algorithms and its competitors to evaluate the accuracy of breast detection in a challenging dataset. Based on extensive experiments, our algorithms are shown to significantly outperform state-of-the-art algorithms. In particular, our proposed method provides several contributions:

- First, we present an ensemble of DCNNs trained to extract the most useful visual features from multi-scale training-images. By building the ensemble of DCNNs, both global and local features of breast cancer tumors can be extracted properly, and the accuracy of multi-label breast cancer detection is thus significantly improved.
- Second, we propose to use the gradient boosting trees algorithm to boost the classification performance of the DCNN classifiers. We proved that the combination of DCNN and a boosting trees classifier leads to a better classification performance despite the limited number of breast cancer samples and imbalanced training data.
- Finally, we use the majority voting strategy to combine the boosting trees

ACCEPTED MANUSCRIPT

classifiers. This combination leads to the best classification performance among competitive boosting trees classifiers.

The remainder of this study is organized as follows. In Section II, we briefly present some related state-of-the-art algorithms for breast cancer detection that motivated our research. In Sections III and IV, we review the methods of Inception networks and gradient boosting trees and show their recent applications to classification tasks. In Section V, we describe our proposed method of incremental boosting convolution networks. In Section VI, the experimental results obtained from challenging datasets of H&E stained histology images are presented. We conclude this paper, mentioning our intentions for our future work, in Section VII.

2. Related Work

In this section, we briefly review and discuss the state-of-the-art algorithms of breast cancer detection and related fields, namely feature extraction using computer-aided diagnosis systems, sparse representation, deep learning for breast cancer detection, and boosting algorithms for classification tasks.

Computer-aided diagnosis (CAD) has rapidly developed in recent years to assist doctors in diagnosing patients quicker and with a higher accuracy in many hospitals. In general, CAD is used to provide objective results and has assisted medical image diagnosis. One of the major CAD applications is to distinguish normal, benign, in situ carcinoma, and invasive carcinoma. A CAD system can also be incorporated into the diagnostic process of breast cancer detection to decrease inter observer variation, effectively providing biopsy recommendations and reducing unnecessary false-positive biopsies. A conventional CAD system consists of three main steps: feature extraction, feature selection, and image classification. Among these steps, effective feature extraction is the most important step because it can improve the performances of the feature selection and classification steps. In general, breast cancer features can be categorized into morphological and textural features. Veta et al. [43] extracted the features

ACCEPTED MANUSCRIPT

of nuclei morphology to distinguish benign and malignant tissues. Kowal et al. [26] aimed to segment nuclei using clustering algorithms for nuclei segmentation from biopsy microscopic images. This algorithm achieved high classification accuracy owing to a good feature extraction strategy in which morphological and texture features were employed. Filipczuk et al. [12] and George et al. [17] extracted circular Hough features for training a nuclei classifier. Particularly, George et al. [17] used the watershed method to extract shape and texture features of nuclei. Both types of features are useful for training an accurate nuclei classifier. For more complex breast histology H&E datasets, Belsare et al. [3] used color-texture graphs for exploring tissue organization followed by the use of texture features for training a breast histology image classifier. Brook et al. [5] built a three-class support vector machine (SVM) classifier to distinguish breast cancer images into three categories: normal, in situ carcinoma and invasive carcinoma. By using multiple threshold values, binary images were generated for the feature extraction and training tasks. Zhang et al. [48] used a set of parallel SVM classifiers and a set of artificial neural networks (ANN) to create a strong cascade classifier. In their research, the SVM and ANN classifiers were considered as weak classifiers that could be combined together to create a final, strong classifier. A breast cancer image was only rejected if it was rejected by the majority of SVM and ANN classifiers. In this algorithm, training features extracted by the Curvelet transform and local binary pattern (LBP) methods were used to train the set of SVM classifiers. Similarly, Zhang et al. [49] presented an ensemble of a single-class kernel principal component analyses models trained using different characteristic features from each breast cancer class. This ensemble method achieved high-classification accuracy for classifying benign and malignant lesions from breast cancer histopathological images. Interestingly, Dimitropoulos et al. [10] used the Grassmannian vector of local aggregated descriptor (VLAD) method to extract local features of breast cancer tumors and present them as a set of multidimensional spatially-evolving signals that can be efficiently modeled through a higher-order linear dynamical systems analysis. Although they did not use deep learning models, their method

ACCEPTED MANUSCRIPT

is one of the state-of-the-art methods for detecting breast cancers.

Sparse representation [51] [52] [47] [45], developed from the theory of sparse coding, has attracted much attention from researchers in the fields of face recognition, object detection, and cancer detection. Helal et al. [23] used a sparse representation-based classifier to classify benign and malignant breast lesions. In this type of representation, a breast cancer sample can be sufficiently represented by training samples from the same subject. Then, this breast cancer sample is classified based on the least representation residual computed by the sparse representation coefficients and training samples. Thus, this method can considerably improve the effectiveness of breast cancer detection. Similarly, Nayak et al. [31] used a method of sparse feature learning and classification to decompose whole slide images of histology sections into distinct patches. These patches were then classified into tumor types. Kong et al. [25] also proposed a method of jointly sparse discriminant analysis to discriminate the category of the breast cancer types. This method was used to extract the key factors in breast cancer, which are helpful for improving the accuracy in diagnosis and prediction.

Recently, deep learning networks have been developed to extract the most discriminative features and to improve the effectiveness of medical image analysis. There are two advantages regarding the use of deep learning networks for feature extraction. First, we can automatically extract more complex feature sets by using deep feature learning models than those that we may have using other machine learning tools. Second, joint and hierarchal learning features can be extracted from different layers of a deep learning network. As a result, deep learning networks are also efficiently used in the feature selection step. However, deep learning networks are still a drawback for H&E breast tissue biopsy classification. Training a deep learning network for recognizing an entire H&E breast tissue biopsy image is extremely time-consuming because of the huge number of training parameters. To deal with this drawback, Spanhol et al. [37] trained his deep learning network on the dataset of 32×32 and 64×64 pixels patches collected from original H&E breast tissue biopsy images. The resulting H&E

ACCEPTED MANUSCRIPT

breast tissue biopsy image was classified based on the sum of patch probabilities or the highest patch probability. Spanhol et al. [31] also proved that the feature extraction task was more difficult if the deep learning network was trained on breast tissue images at higher magnifications. This is because only nuclei edges were extracted from images at high magnifications and the most useful features were not found by the deep learning network. Similarly, Ciresan et al. [8] used convolutional neural networks (CNN) to detect mitosis in each of the patches extracted from H&E stained breast biopsy slides. The advantage of this deep learning model is to detect nuclei in different sizes. Cruz-Roa et al. [9] also divided the breast histology slide into 100×100 patches in which his deep learning network could detect invasive carcinoma regions. Han et al. [20] used a deep learning model to identify eight classes of breast cancer. With the application of deep learning models, Euclidean distances among samples in the same class were minimized whereas the Euclidean distance between two arbitrary samples that belonged to two different classes was maximized. Similarly, Song et al. [36] combined a convolution neural network with a Fisher feature layer to encode the local features of breast cancer tumors in a higher discriminative space where breast cancer types were distinguished effectively. Interestingly, Chen et al. [7] proposed an architecture of deep cascaded networks to quickly retrieve the mitosis candidates while preserving a high sensitivity. The retrieved candidates were then classified by the second deep convolutional neural network which can discriminate mitoses from hard mimics more precisely. Therefore, this approach achieved high performance of classifying breast cancer images.

By using boosting algorithms, some researchers [13], [35], [21] can theoretically design a strong learning model by combining multiple weak learners. These boosting algorithms collect and optimize decision trees, which are considered weak learners. Among the available boosting algorithms, AdaBoost could advantageously use recurrent neural networks as weak learners, which are applied for analyzing text or time series, as demonstrated by Assaad et al. [2] and Buabin et al. [6]). In addition, Karianakis et al. [24] tried to use the AdaBoost decision trees method to combine CNNs for classifying images. Friedman et al.

ACCEPTED MANUSCRIPT

[14] introduced gradient boosting machines (GBM) which can be applied for training neural networks and can be used as an effective gradient descent optimization algorithm. In turn, Gao et al. [16] developed a boosting algorithm that selected weak CNN learners among a set of CNNs to build a stronger learner for video recognition. Wang et al. [50] and Han et al. [19] built a boosting algorithm to regularize the CNN loss function.

3. Gradient Boosting Trees

The algorithm for boosting trees was developed to solve regression and classification problems and to produce a strong prediction model by ensembling weak prediction models. This algorithm can be considered as an optimization algorithm that computes a sequence of successive trees. Every single tree is used to predict the residuals of the preceding tree. In general, the gradient boosting trees algorithm aims to combine weak classifiers into a single strong classifier.

3.1. Regularized Learning Objective

Like any supervised learning algorithm, the objective of the gradient boosting algorithm is to classify objects by minimizing a loss function. To identify the solution for this optimization problem, boosting trees are trained by using gradient descent algorithms, and predictions of new boosting trees are updated and improved based on a learning rate.

First, we denote the set of n_1 examples and n_2 features by $X = \{(u_i, v_i)\}$ ($|X| = n_1$, $u_i \in R^{n_2}$, $v_i \in R$). We aim to build a tree ensemble model that combines M single functions to classify the examples in this dataset.

$$\hat{v_i} = \phi(u_i) = \sum_{k=1}^{M} f_k(u_i),\ f_k \in F \tag{1}$$

where $F = \left\{f(u) = \delta_{h(u))}\right\}$ ($h : R^{n_2} \rightarrow D, \delta \in R^D$), v_i is the i^{th} target value, and $\hat{v_i}$ is the i^{th} prediction. Herein, $h(u)$ represents a single tree that can classify training data u into the corresponding group. Each single tree $h(u)$ consists of D leaves. Thus, each f_k is a function of a regression tree h and corresponding

ACCEPTED MANUSCRIPT

leaf weights w. We denote the score on the i^{th} leaf of the regression tree h by δ_i. Each example is classified in accordance with the leaves of the regression tree h based on a set of decision rules. As a result, we can compute the last prediction for this example by summing up the scores in the corresponding leaves. The set of functions can be found by solving the following minimization problem

$$G(\phi) = \sum_{i=1} g(\hat{v}_i, v_i) + \sum_{k} \delta(f_k) \tag{2}$$

where $\delta(f) = \gamma D + \frac{1}{2}\lambda \|\delta\|^2$. In this formula, the difference between the prediction $\hat{v}_i$ and the target v_i is computed by the differentiable convex loss function g and the complexity of the model is measured by the additional regularization term δ. The additional regularization term is added to this equation to avoid the overfitting problem and to smooth the learning weights. This is because the model of boosting trees tends to select simple functions for the prediction task. Accordingly, the training parameters are minimized. Ideally, if the loss function $G(\phi)$ can be minimized, the sum of its residuals should be approximately zero. In fact, each training data point has one residual equal to the difference between the observed value v_i and the predicted value $\hat{v}_i$. The idea of the gradient boosting algorithm is the repetitive updating of the new boosting trees to reduce residuals and strengthen the model. The training process can stop if the sum of these residuals is less than a predefined threshold.

3.2. Gradient Tree Boosting

The optimal solution for Eq. (2) cannot be directly found by conventional optimization methods. Instead, we apply an effective approximation method to optimize parameters and functions in this equation. We denote the prediction of the i^{th} example at the t^{th} iteration by $\hat{v}_i^{(t)}$. The solution for Eq. (2) can be found by solving the following minimization problem:

$$G^{(t)} = \sum_{i=1}^{n_1} g(v_i, \hat{v}_i^{(t-1)} + f_t(u_i)) + \delta(f_t) \tag{3}$$

ACCEPTED MANUSCRIPT

By adding f_t to Eq. (3), we can apply a second-order approximation to optimize the objective in this equation quickly, as follows:

$$G^{(t)} \simeq \sum_{i=1}^{n_1}[g(v_i, \hat{v}_i^{(t-1)} + l_i f_t(u_i) + \frac{1}{2}k_i f_t^2(u_i)] + \delta(f_t) \tag{4}$$

where $l_i = \partial_{\hat{v}^{(t-1))}} g(v_i, \hat{v}_i^{(t-1)})$ is the first order gradient statistic on the loss function, and $k_i = \partial^2_{\hat{v}^{(t-1))}} l(v_i, \hat{v}_i^{(t-1)})$ is second order gradient statistic value. We can simplify Eq. (4) by the following equation:

$$\tilde{G}^{(t)} \simeq \sum_{i=1}^{n_1}[l_i f_t(u_i) + \frac{1}{2}k_i f_t^2(u_i)] + \delta(f_t) \tag{5}$$

We denote the instance set of leaf j by $T_j = \{j|h(u_i) = j\}$. Eq (3) can be rewritten by expanding δ as follows:

$$\begin{aligned} \tilde{G}^{(t)} &\simeq \sum_{i=1}^{n_1}[l_i f_t(u_i) + \frac{1}{2}k_i f_t^2(u_i)] + \gamma D + \frac{1}{2}\lambda \sum_{j=1}^{D} \delta_j^2 \\ &= \sum_{j=1}^{D}[(\sum_{i \in T_j} g_j)\delta_j + \frac{1}{2}(\sum_{i \in T_j} h_i + \lambda)\delta_j^2] + \gamma D \end{aligned} \tag{6}$$

We can fix the structure $h(u)$ and optimize the weight δ_j^* of leaf j by:

$$\delta_j^* = -\frac{\sum_{i \in T_j} l_j}{\sum_{i \in T_j} k_i + \lambda} \tag{7}$$

The corresponding optimal value can be computed by:

$$\tilde{G}^{(t)}(h) = -\frac{1}{2}\sum_{j=1}^{D} \frac{(\sum_{i \in T_j} l_j)^2}{\sum_{i \in T_j} k_i + \lambda} + \gamma D \tag{8}$$

By computing the scores in Eq (8), we can evaluate the quality of a tree structure h. To enumerate all the possible tree structures h, we applied a greedy algorithm to search every single leaf, and branches were iteratively added to the tree. After the split, the instance sets of the nodes on the left and right sides become available and are denoted by T_L and T_R, respectively. The loss reduction

ACCEPTED MANUSCRIPT

is then computed based on the instance sets T_L and T_R as follows:

$$G_{split} = -\frac{1}{2}\left[\frac{(\sum_{i\in T_L} l_i)^2}{\sum_{i\in T_L} k_i + \lambda} + \frac{(\sum_{i\in T_R} l_i)^2}{\sum_{i\in T_R} k_i + \lambda} + \frac{(\sum_{i\in T} l_i)^2}{\sum_{i\in T} k_i + \lambda}\right] \tag{9}$$

Eq. (9) is used to evaluate the split candidates. These split candidates are added to minimize the residuals in the predictions. Thus, choosing the best split points is required to minimize the loss function. To prevent the overfitting problem, we applied two different techniques: shrinkage and column subsampling. The shrinkage technique [15] aimed to improve the model quality by reducing the influence among individual trees. New weights were rescaled by a factor η after each boosting step. The shrinkage factor η is referred to as the learning rate. If the learning rate is reduced, more trees are added to the model, and the training process is more effective. The column subsampling technique [4] is used to prevent the overfitting problem and speed up computations of the boosting trees algorithm. Moreover, this technique is used to reduce correlations between every successive part of boosting trees in the model.

As indicated by Eq. (9), finding the best split is also very important for improving the classification performance. For this purpose, we use the exact greedy algorithm as shown in Alg. 1. The basic idea of this algorithm is to enumerate all the possible splits and then sort the data based on feature values. In this sorted order, gradient statistics are accumulated and the structure score is computed based on Eq. (9). Finally, the best split corresponds to the highest structural score.

4. Inception Networks

Inception modules [39] aim to improve the improvement of the training performance by using many tiny convolution kernels to simultaneously map cross-channel and spatial correlations. The key advantage of these tiny kernels is that they are not only used to design a CNN with considerably fewer parameters than larger kernels but can also be used to extract effectively the detailed

ACCEPTED MANUSCRIPT

Algorithm 1: Exact Greedy Algorithm

procedure PROCEDURE(T, d)
  Input: T, set of the current node
  Input: d, feature dimension
  $gain \leftarrow 0$
  $P \leftarrow \sum_{i \in T} l_i$, $Q \leftarrow \sum_{i \in T} k_i$
  for x=1 **to** m **do**
    $P_L \leftarrow 0$, $Q_L \leftarrow 0$
    for y in sorted(T, by u_{xy}) **do**
      $P_L \leftarrow P_L + l_y$, $Q_L \leftarrow Q_L + k_y$
      $P_R \leftarrow P - P_L$, $Q_R \leftarrow Q - Q_L$
      $score \leftarrow max(score, \frac{P_L^2}{Q_L+\lambda} + \frac{P_R^2}{Q_R+\lambda} - \frac{P^2}{Q+\lambda})$
    end
  Output: Split with highest scores

information in images. Thus, the network training procedure can be more effective and efficient and the overfitting problem can be avoided. In particular, Inception networks use convolutions of different sizes to extract effectively the detailed information at different scales, as shown in Figure 2. In this figure, we can see that each Inception layer employs 1×1, 3×3 and 5×5 convolutions to represent larger convolutions. Therefore, the computational cost of the network can be reduced significantly when the network goes deeper with the use of more convolution layers. In addition, in the network, the fully-connected layers are replaced by a global average-pooling layer that averages the channel values. This change drastically reduces the overall number of network parameters without affecting the accuracy. Both key changes are also helpful in avoiding the overfitting problem when the number of training images is limited, which is a common problem in medical image analysis research. Moreover, combining the Inception architecture with residual connections can lead to dramatically improved training speeds, and can significantly improve the recognition performance, as

ACCEPTED MANUSCRIPT

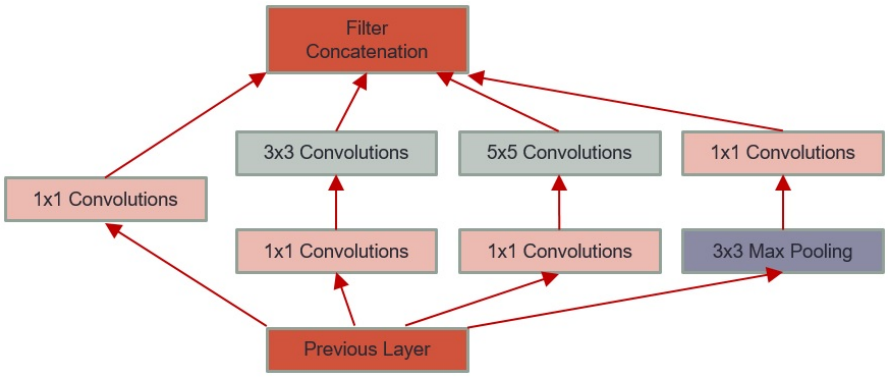


Figure 2: Example of an Inception module.

explained in [39].

Among the Inception networks, the Inception-ResNet-v2 network [39] outperforms similarly expensive networks without residual connections. This is because the residual learning framework plays a key role in the improvement of training speed for the Inception architecture. Furthermore, an Inception network using residual connections has deeper convolution layers for effectively extracting high-level features from images.

In this study, we used Inception networks at different input scales to capture multi-level features of breast tissues. By using these networks, we can capture detailed information pertaining to breast cell types that indicates the similarity of breast cancer cells to normal breast tissues. In addition, the Inception network can extract high-level features in the breast cancer images to evaluate the growth rate of breast cancer cells by estimating the density of breast cells in the image. Moreover, since breast cancers can grow, spread, and invade the surrounding breast tissues, we employed Inception networks with different input scales to extract multi-scale features of different breast cancer types.

5. Proposed Approach

In this section, we describe our approach of breast cancer detection in detail. The entire training process of our proposed algorithm is illustrated as a flow

ACCEPTED MANUSCRIPT

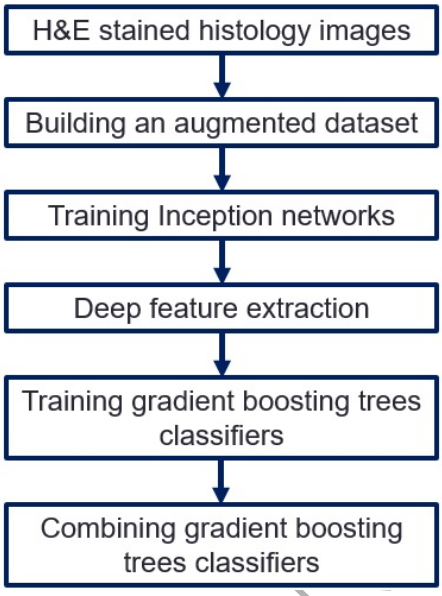


Figure 3: The whole training process of our proposed algorithm.

chart in Figure 3. Our approach include five basic steps: First, we apply a preprocessing method of stain normalization on H&E stained histology images to transform them into a common space and reduce their variances. This essential step is useful in improving the detection performance. In the second step, we present novel augmentation methods that are able to increase effectively the amount of training images based on our original limited training dataset. In the third step, we employ this augmented training dataset to train a set of Inception networks with multi-scale input images. After these training processes, the most discriminative deep features of breast tissues can be extracted from these Inception models. Similar to the third step, in the fourth step, the discriminative deep features extracted from the training dataset can be used again to train a set of gradient boosting trees classifiers to improve the classification performance. In the last step, we employ a novel strategy for combining the gradient boosting trees into a stronger boosting classifier that is able to detect breast cancer clues precisely on histology images.

ACCEPTED MANUSCRIPT

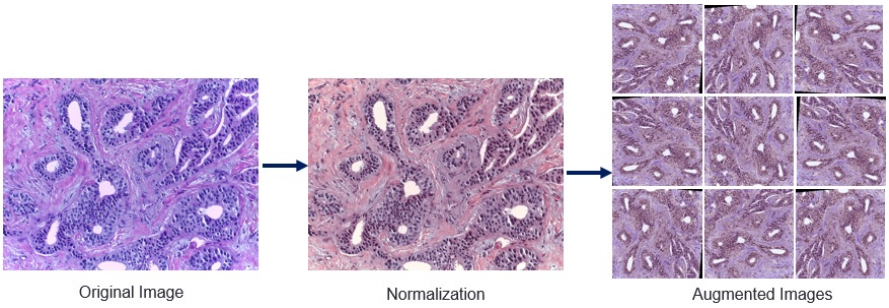


Figure 4: Preprocessing steps of our proposed algorithm.

5.1. Preprocessing

For an automated breast cancer classifier using deep learning networks, stain normalization is an essential step in improving the detection performance. Because the procedures of tissue staining, fixation and cutting are not consistent, the appearance of H&E stained histology images significantly changes across different laboratories. The preprocessing step of stain normalization transforms the histology images into a common space and reduce their variances. In this study, we used the method of stain normalization proposed in [28]. This method uses a logarithmic transformation to compute the optical density of each histology image. The singular value decomposition method was applied to this optical density image to estimate the relevant degrees of freedom and construct a 2D projection matrix with a higher variance. We then calculated the intensity histograms for all pixels, whereby the dynamic range of elicited intensities covered the lower 90% of the data.

Since access to data is limited owing to privacy concerns, breast cancer detectors were often trained with insufficient training datasets. Consequently, the cancer classification performance is hindered by this lack of training data. Recent work has demonstrated the effectiveness of data augmentation in increasing the amount of training data based on our original training dataset that consisted of limited data. By augmenting training data, we can also reduce the overfitting problem on the training models. In this study, we mainly performed geometric augmentations including reflecting, randomly cropping, rotating and translating

ACCEPTED MANUSCRIPT

the image. Since the color of H&E stained histology images significantly varies across laboratories due to various technical skills, we applied an effective color constancy method, namely gray world that assumes the scene in an image, on average, is a neutral gray and the average reflected color is from the color of the light. For this reason, the illuminant color cast can be well estimated by computing the average color and comparing it to gray values. In this algorithm, the illuminant colors are computed by the mean of each channel of the image. Figure 4 shows all the basic preprocessing steps of our proposed algorithms.

5.2. Feature Extraction

We aim at building an ensemble of DCNNs to train the most discriminative features for the breast cancer detection task. A DCNN ensemble is a learning paradigm that many DCNNs are jointly employed to solve a specific problem. Our previous work [44] has demonstrated that an ensemble of multiple deep convolutional networks significantly outperforms single deep convolutional networks. This is because the DCNN ensemble has some competitive advantages that are useful for increasing the prediction accuracy rate. First, we can apply multi-scale input images to our network ensemble in which each scale is passed through at least one CNN. This ensemble not only expands the receptive field in the original image to cover better global features but also extracts better multi-scale local features. Second, our ensemble of DCNNs is a reliable technique to increase the classification performance. Because each training deep model presents several local minima, multiple training processes of different DCNNs can improve the distribution of errors in each class. Thus, combining their outputs leads to the improved performance on the overall task. Third, we use the inception network architecture that has been shown to achieve very good performance at relatively low computational costs. In this study, we employed the state-of-the-art Inception network model namely the Inception-ResNet-v2 model. Szegedy et al. [39] demonstrated that the Inception-ResNet-v2 model yields state-of-the-art performance in the ImageNet large-scale visual recognition challenge (ILSVRC) [34] because it gets advantage from residual connec-

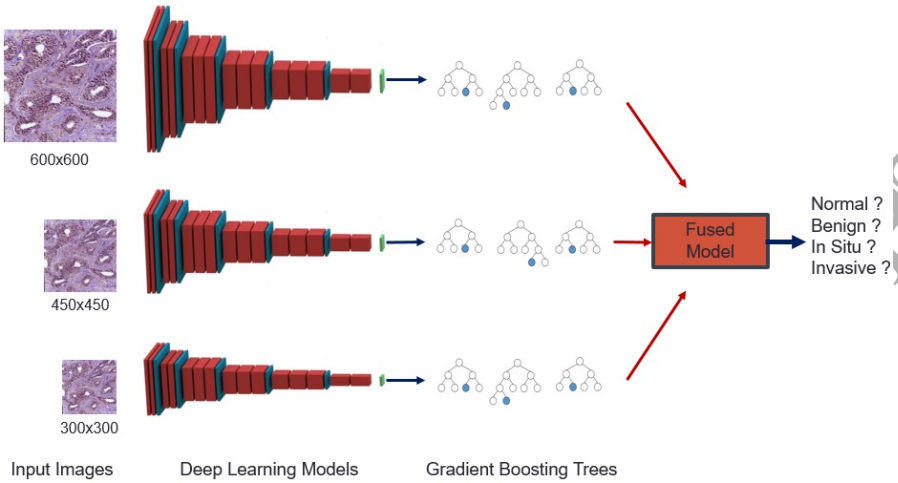


Figure 5: The entire architecture of our feature extraction model.

tions. Unlike our previous research [46] that used pre-trained deep learning models to extract discriminated features, we can train every single DCNN using its pre-trained model and our augmented training dataset. This is because in this study we can successfully generate enough augmented images for training these networks and avoiding the overfitting problem. After these training processes, our Inception models can extract far much better features than their pre-trained models. The training steps are explained in detail in the next section.

5.3. Training Inception Models

In the data augmentation step, randomly cropping, rotating and translating, each original training image generates hundreds of new augmented training images. This new dataset is used for training each Inception network with differently scaled input images. Each Inception-ResNet-v2 model is trained with one of three kinds of scaled input images that are 600×600, 450×450 and 300×300 images. The fully connected layers from each model are disconnected so that each network can accept input images with an arbitrary size. Each input image is a rescaled image from the augmented dataset. The entire architecture of our

ACCEPTED MANUSCRIPT

feature extraction model is illustrated in Figure 5. We train each network with the stochastic gradient running on a GeForce GTX 1080 Ti, utilizing the TensorFlow [29] distributed machine learning system. We fine-tune the networks by using the corresponding Inception-ResNet-v2 models using RMSProp [42] with a decay of 0.9. We use a learning rate of 0.007, decayed every two epochs using an exponential rate of 0.94. The training process for each Inception network takes 50 epochs. We convert the last convolutional layer of each model into a one-dimensional feature vector with a length of 1056. Finally, all the one-dimensional feature vectors are used as the inputs to gradient boosting trees classifiers.

5.4. Training Gradient Boosting Trees Classifiers

For each Inception-ResNet-v2 model, we build a training dataset of deep feature vectors, extracted from training images in our augmented dataset. Each dataset of deep feature vectors is used to train a gradient boosting classifier to improve further the classification accuracy rate. We use a learning rate of 0.05 in every training model. The number of leaves and the tree depth, which are the main parameters to control the complexity of the tree model, are set to 191 and 6, respectively. In the training processes, each training data is split into classes based on the highest score.

After the training processes, all the gradient boosting trees classifiers are combined using the majority voting strategy to create a stronger classifier. In our previous paper [44], we proved that although DCNN is one of state-of-the-art object detection methods, it is not able to capture fully multi-scale context information of different objects. In this study, an Inception network can achieve the highest accuracies in detecting one or two breast cancer types but other Inception networks can be better at recognizing the remaining types. Thus, this combination is a better solution for utilizing the advantages of multi-resolution images and multi-scale feature descriptors and extracting both global and local information of different breast cancer tumors.

ACCEPTED MANUSCRIPT

6. Experimental Results and Analysis

6.1. Dataset

In this section, we demonstrate the effectiveness of our proposed methods of breast cancer detection by evaluating them on two datasets: the Bioimaging 2015 breast histology classification challenge [32] and the BreaKHis dataset [38].

We used the challenging database (Bioimaging-2015) from the Bioimaging 2015 breast histology classification challenge to evaluate the accuracy and sensitivity of our methods and compared our results with those obtained from state-of-the-art algorithms. This dataset consisted of H&E stained histology images that were digitized with the same acquisition conditions, with the same acquisition matrix of 2040×1536 pixels, and with an in-plane pixel size of $0.42\mu m \times 0.42\mu m$. Each image was labeled by two pathologists who provided a diagnostic from the image contents. They classified each image into four classes: normal tissues, benign lesions, in situ carcinomas and invasive carcinomas. There were 249 microscopy training images and 36 microscopy testing images in total. Each image was also classified into two groups, non-carcinomas, and carcinomas, by grouping the normal and benign classes into the non-carcinoma class and grouping the in situ and invasive classes into the carcinoma class.

We trained our Inception models using differently scaled challenging input images from this challenging dataset. Images had matrix sizes of 300×300, 450×450 and 600×600. We also used features extracted from these models to train three corresponding gradient boosting models. These models were then compared with the state-of-the-art models [1] [9] [37]. Each proposed model was denoted by the pre-trained model name, the image input size, and the classifier name. For example, Inception-450x450+GBT denotes the method that uses the pre-trained Inception-ResNet-v2 model with a 450×450 image input and a gradient boosting trees classifier. Inception-600x600 represents the method that uses the pre-trained Inception-ResNet-v2 model with a 600×600 image input and a Softmax classifier.

The proposed method was also tested on the BreaKHis dataset that included

ACCEPTED MANUSCRIPT

Table 1: Recognition rate on the challenging database of H&E stained histological breast cancer images.

Method	Four Classes	Two Classes
CNN+SVM	77.8 %	83.3 %
CNN	77.8 %	80.6 %
Inception-300x300	82.0 %	90.9 %
Inception-300x300+GBT	90.8 %	97.2 %
Inception-450x450	88.1 %	92.2 %
Inception-450x450+GBT	91.9 %	95.0 %
Inception-600x600	88.9 %	95.3 %
Inception-600x600+GBT	92.2 %	98.6 %
Model Fusion	**96.4** %	**99.5** %

7909 images collected from 82 anonymous patients. These images were digitized with the same resolution of 700×460 pixels. The BreaKHis dataset was divided into benign and malignant tumors, each of which included four magnification factors, namely, 40X, 100X, 200X, and 400X. Particularly, pathologists classified benign tumors into four subclasses that included adenosis (A), tubular adenoma (TA), phyllodes tumor (PT), and fibroadenoma (F). They also divided malignant tumors into four other subclasses, which were ductal carcinoma (DC), mucinous carcinoma (MC), lobular carcinoma (LC), and papillary carcinoma (PC). To ensure a higher level of reliability, 70% of available data were randomly chosen to train deep learning models and gradient boosting classifiers. We used the remaining 30% of the data for performance evaluation. We also trained our Inception models and gradient boosting classifiers on this challenging dataset, using differently scaled input images with the sizes of 300×300, 450×450 and 600×600. These models were then compared with the state-of-the-art algorithms [1] [9] [37].

ACCEPTED MANUSCRIPT

Table 2: Sensitivity for four-class classification on the challenging database of H&E stained histological breast cancer images.

Method	Normal	Benign	In situ	Invasive
CNN	77.8 %	55.6 %	88.9 %	88.9 %
CNN+SVM	77.8 %	66.7 %	77.8 %	88.9 %
Inception-300x300	80.0 %	87.8 %	86.7 %	73.3 %
Inception-300x300+GBT	100 %	90.0 %	88.9 %	84.4 %
Inception-450x450	77.8 %	96.7 %	82.2 %	95.6 %
Inception-450x450+GBT	83.3 %	100 %	85.5 %	98.9 %
Inception-600x600	82.2 %	95.6 %	86.7 %	91.1 %
Inception-600x600+GBT	95.6 %	96.7 %	88.9 %	87.7 %
Model Fusion	**97.8** %	**100** %	**88.9** %	**98.9** %

6.2. Results on the Bioimaging-2015 dataset

Table 1 shows that the classification accuracy of our Inception networks is much better than that of the competing classifiers (CNN) [1] for the four classes, with a better accuracy of at least 4.2%. These results prove that our Inception models effectively improved the performance of the breast cancer classifier. This is because these models can extract more key breast cell features compared to CNN. CNN consisted of four narrow convolution layers that were not enough to extract unique characteristics of breast cancer cells, which was not an easy task because of a wide variety of H&E stained sections. On the contrary, our Inception models can extract detailed information from breast cell types that indicate the similarities of breast cancer cells to normal breast cells. Each model was trained by a very deep network that was crucial for capturing the natural hierarchy of objects. Low-level features were captured in the first layer, and object parts were extracted at higher layers. Furthermore, the residual learning framework eased the training of these networks and enabled them to extract higher feature levels, leading to improved performance in recognition tasks.

We also evaluated the performance of the gradient boosting trees classifiers

ACCEPTED MANUSCRIPT

Table 3: Sensitivity for two-class classification on the challenging database of H&E stained histological breast cancer images.

Method	Non-carcinoma	Carcinoma
CNN	74.5 %	80.6 %
CNN+SVM	79.2 %	74.5 %
Inception-300x300	83.9 %	97.8 %
Inception-300x300+GBT	94.4 %	100 %
Inception-450x450	87.7 %	96.7 %
Inception-450x450+GBT	91.7 %	98.3 %
Inception-600x600	90.5 %	100 %
Inception-600x600+GBT	97.2 %	100 %
Model Fusion	98.9 %	100 %

using deep features from the Inception models, as shown in Table 1. Inception-300x300+GBT, Inception-450x450+GBT and Inception-600x600+GBT are more accurate than Inception-300x300, Inception-450x450 and Inception-600x600, respectively. This is because the gradient boosting trees classifier considerably improves the accuracy improvement of the classification of breast cancer features in the deep learning models.

Table 2 indicates that each Inception network has its own advantages and disadvantages in detecting breast cell types. We can observe that Inception-300x300+GBT is the best classifier for verifying normal breast cells with a sensitivity of 100% while Inception-450x450+GBT achieves the highest accuracies in detecting benign tumors and invasive carcinomas, with sensitivities of 100% and 98.9%, respectively. For the non-carcinoma/carcinoma tissue classification task, Inception-600x600+GBT achieved a higher accuracy rate than Inception-300x300+GBT and Inception-450x450+GBT. The sensitivity of Inception-600x600+GBT in detecting carcinomas was 100% and its specificity was 97.2%. This can be explained by the fact that, although the Inception-ResNet-v2 network is the state-of-the-art object detection method, it is not

ACCEPTED MANUSCRIPT

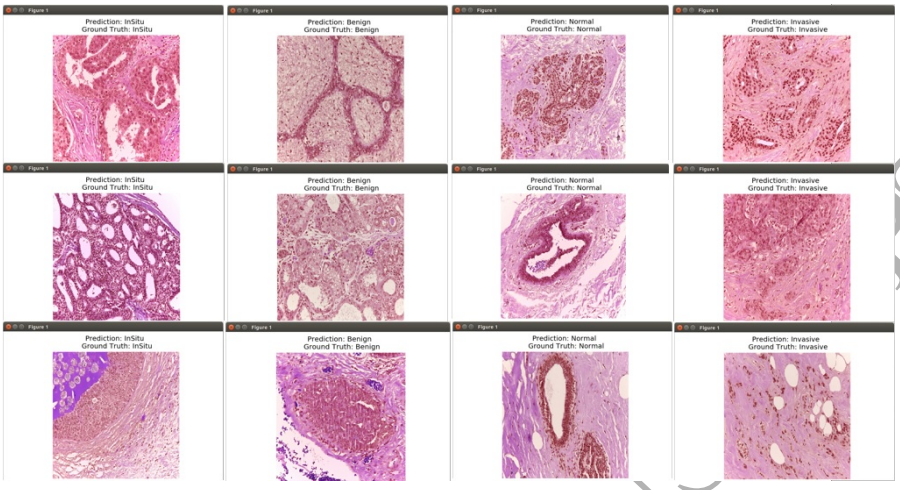


Figure 6: Qualitative results of our proposed method on the database from the Bioimaging 2015 breast histology classification challenge [32].

able to capture fully multi-scale context information of different breast cancer types. Table 1 demonstrates that the fused model can achieve a 96.4% accuracy in the cases of four class problems, which is the best among the competitive approaches using deep learning models. The fused model accuracy was at least 4.2% higher than any of the single models. This proves once again that this model can exploit the deep network architecture of multi-resolution input images to aggregate multi-scale contextual information, and can also utilize the advantages of its single models, as mentioned in our previous research [44].

Regarding binary classification, the accuracies of our classifiers considerably increased compared to the four-class problem. This is because the normal and benign classes are not much different, and the in situ class also shares similar features with those of the invasive class. The results prove that the fused model was the best in reference to the algorithms included in the experiment of binary classification, and achieved a total accuracy of 99.5%. Table 3 also demonstrates that the sensitivity of the fused model used to detect carcinomas is 100% and its specificity is 97.2%.

We also compared our proposed deep learning networks with one of the state-

ACCEPTED MANUSCRIPT

of-the-art algorithms using CNN (Cruz-Roa-CNN), presented by Cruz-Roa et al. [9]. This algorithm aimed at the detection of invasive carcinomas in each high-resolution image patch. Table 3 shows that the overall sensitivity of our fused model for patch-wise classification of invasive carcinomas is better than that of this competing classifier. Our model achieved a sensitivity of 100 % for patch-wise classification of invasive carcinomas, which is 20.4% better than Cruz-Roa-CNN. Our method significantly outperformed Cruz-Roa-CNN despite the fact that our method was not a dedicated invasive carcinoma detection method. This is because our method can capture multi-level and multi-scale features and can recognize not only individual nuclei features but also the structural organization.

Spanhol et al. [37] used approximately 2000 images for training CNNs that were used for classifying benign and malign tumors. They achieved a classification accuracy rate of 84%. For the non-carcinoma/carcinoma tissue classification task, our fused model achieved 99.5 % accuracy, which was the best among the competitive approaches using deep learning models. Although we used less training images, our performance was still better because our method could generate more training images by using the proposed data augmentation technique. Moreover, our network model could learn features at different scales through its convolutional layers. Thus, our network can recognize better individual nuclei as well as nuclei structures. Qualitative results of our proposed method on the database from the Bioimaging 2015 breast histology classification challenge are shown in Figure 6.

6.3. Results on the BreaKHis dataset

Table 4 shows that our fused model achieved the highest accuracy among all the Inception models. The experimental results show again that the fused method, owing to its ability to aggregate multi-scale contextual information, outperforms all other approaches, and achieves improvements of at least 3.3%, 4.2%, 5.5% and 3.6% for images at the respective magnification factors of of 40X, 100X, 200X and 400X. In fact, our fused model can produce considerably higher accuracies compared to the state-of-the-art classifier, VLAD classifier [10], be-

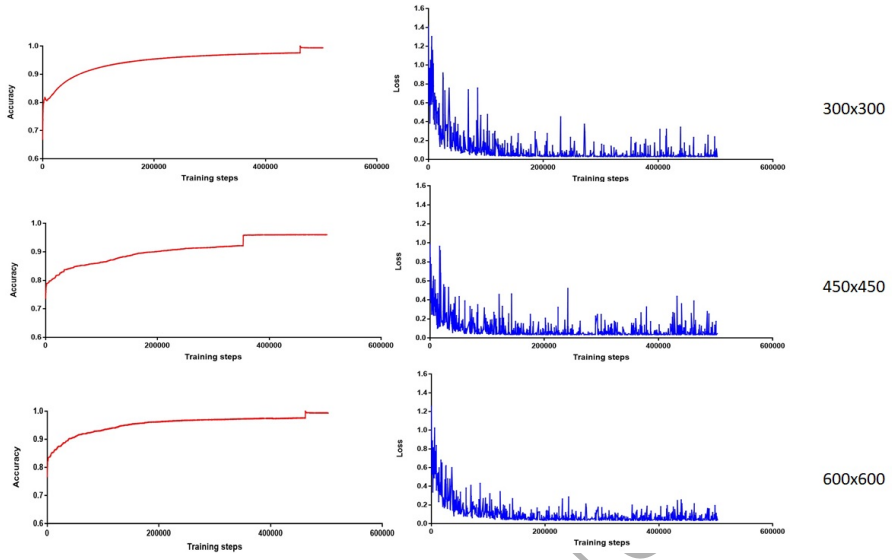


Figure 7: Variations in accuracy and loss with number of training steps.

cause it is able to aggregate multi-level local descriptors into a compact object representation. Although the VLAD method can effectively extract local feature descriptors of breast cancer tumors, our boosting framework of deep learning models still considerably outperformed their approach, which only models histological images through a higher-order linear dynamical systems analysis using a set of multidimensional spatially-evolving signals.

Table 4 also shows that features that are extracted from multi-scale image inputs and are then fused into a boosting framework can perform better than conventional deep learning networks in problems of object classification. This also proves that our boosting algorithm is more effective than deep learning networks when dealing with the problem of a limited number of available training samples. Figure 7 shows the training accuracies and losses of our deep convolutional neural networks, which were Inception-300x300, Inception-450x450, and Inception-600x600. The training images used in this experiment are the images at the magnification factors of 400X in the BreaKHis dataset. These training accuracies are considerably higher than the corresponding testing accuracies.

ACCEPTED MANUSCRIPT

Table 4: Comparison of our proposed methods with other state-of-the-art methods on the BreaKHis database.

Method	40X	100X	200X	400X
PFTAS [10]	83.80 %	82.10 %	85.10 %	82.30 %
ORB [10]	74.40 %	69.40 %	69.60 %	67.60 %
LPQ [10]	73.80 %	72.80 %	74.30 %	73.70 %
LBP [10]	75.60 %	73.20 %	72.90 %	73.10 %
GLCM [10]	74.70 %	78.60 %	83.40 %	81.70 %
CLBP [10]	77.40 %	76.40 %	70.20 %	81.80 %
Inception-300x300	90.2 %	91.9 %	93.7 %	88.9 %
Inception-300x300+GBT	93.5 %	95.3 %	96.1 %	91.1 %
Inception-450x450	90.1 %	91.5 %	93.9 %	90.5 %
Inception-450x450+GBT	93.8%	94.9 %	96.5 %	92.7 %
Inception-600x600	91.5 %	92.5 %	94.1 %	91.9 %
Inception-600x600+GBT	94.2 %	95.9 %	96.7 %	92.5 %
CNN [37]	90.40 %	87.40 %	85.00 %	83.80 %
VLAD [10]	91.80 %	92.10 %	91.40 %	90.20 %
Model Fusion	**95.1** %	**96.3** %	**96.9** %	**93.8** %

These results imply that DCNNs suffer from overfitting. In this experiment, our gradient boosting method is able to improve the performance by preventing overfitting.

We also evaluated the classification performances of each different deep learning framework, including ResNet-V1-152 [22], Inception-V3 [40], Inception-V4 [39], and Inception-300x300, as shown in Table 5. All the image inputs of these networks were images acquired with 300×300 matrices. They are the state-of-the-art convolutional neural network image classification models. Table 5 shows that the models using the residual learning framework, including Inception-300x300+GBT and ResNet-V1-152+GBT, achieve considerably increased accuracies for object classification. The residual learning framework

ACCEPTED MANUSCRIPT

Table 5: Comparison of our proposed methods with other state-of-the-art fine-tuning deep learning models on the BreaKHis database.

Method	40X	100X	200X	400X
ResNet-V1-152+GBT	92.5 %	94.8 %	95.8 %	90.5 %
Inception-V3+GBT	88.9 %	90.1 %	89.1 %	87.5 %
Inception-V4+GBT	90.4 %	92.3 %	93.7 %	90.1 %
Inception-300x300+GBT	93.5 %	95.3 %	96.1 %	91.1 %

played an important role in training very deep networks. It eased the training of these networks and led to an improved accuracy in classification tasks. Table 5 also indicates that Inception-300x300+GBT is the best among the competitive approaches using deep learning models. This is because it benefits by combining the Inception architecture with residual connections. First, residual connections lead to dramatically improved training speeds for the Inception architecture. Second, this Inception network that used residual connections, could employ deeper convolution layers for the effective extraction of high-level features from images.

7. Conclusion

In this paper, we proposed a novel approach that used an incremental boosting convolution network for breast cancer detection. This network could classify each H&E stained histological breast cancer image into one of the four balanced classes: normal, benign, in situ carcinoma, and invasive carcinoma. Alternatively, each H&E stained histological breast cancer image could also be classified into one of two basic groups: carcinomas or non-carcinomas. The key advantages of the proposed method over existing methods are that it employs an ensemble of DCNNs trained to extract visual features from multi-scale images and then uses a boosting framework to achieve a better classification performance despite the limited number of breast cancer samples and imbalanced training data, which are the challenging problems.

ACCEPTED MANUSCRIPT

Our proposed classifier consisted of two training stages. In the first training stage, strong data augmentation methods were used to improve the classification performance of deep learning networks. We presented an ensemble of DCNNs trained to extract multi-context information from multi-scale training images. This ensemble of DCNNs was able to extract both global and local features of breast cancer tumors properly, and it could thus greatly improve the classification performance of different breast cancer types.

In the subsequent stage of execution of this method, these multi-scale breast cell features were adopted again to train the corresponding gradient boosting trees classifiers. The combination of DCNN and the boosting trees classifier was proved to elicit a better classification performance than the DCNN trained in the previous stage, despite the challenging problems of the limited number of breast cancer samples and imbalanced training data. A better training model was built by combining three gradient boosting trees, using the majority voting strategy. This fused model achieved a higher classification accuracy and higher sensitivity for carcinoma cases. The experimental results on challenging datasets demonstrated that this algorithm empirically outperformed state-of-the-art methods.

Experimental results demonstrated the superiority of the proposed method against state-of-the-art breast cancer classification approaches. In the future development of our methods, we intend to improve the performance of incremental boosting convolution networks by adopting several novel following methods:

First, to effectively overcome the problems of the limited number of available training samples and the imbalanced data, each static breast cancer histological image can be divided into a set of image patches. In addition, we can use a Euclidean distance to measure the similarity between two image patches.

Second, we aim to present a novel class-aware loss function that can be used in our proposed DCNNs to minimize the intra-class distances of image patches. This loss function is to train a center for deep features of each breast cancer class and penalize the Euclidean distances between deep features and their corresponding class centers. As a result, the inter-class features differences

ACCEPTED MANUSCRIPT

can be enlarged as much as possible while the intra-class features variations are minimized. Thus, this loss function can lead to higher classification performance for our DCNNs.

Third, by using a patch-based voting aggregation algorithm, we will be able to classify different types of breast cancer tumors precisely in whole images.

Finally, boosting trees classifiers can be used to improve classification performance. To improve classification performance of boosting trees classifiers, we plan to extract training features from the most important convolution layers of our DCNNs. This strategy is beneficial to build a strong classifier because more low-level and high-level features are collected for training boosting trees models.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (NRF-2015R1D1A1A01060422) and also supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in Software supervised by the IITP (Institute for Information & Communications Technology Promotion) (2015-0-00932).

References

[1] T. Araujo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, Classification of breast cancer histology images using convolutional neural networks,in: PLOS ONE 12(6): e0177544, https://doi.org/10.1371/journal.pone.0177544, 2017.

[2] M. Assaad, R. Boné, H. Cardot, A new boosting algorithm for improved time-series forecasting with recurrent neural networks, in: Information Fusion 9 (1), 41–55. Jan. 2008.

ACCEPTED MANUSCRIPT

[3] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, N. Meshram, Classification of breast cancer histopathology images using texture feature analysis, in: TENCON 2015—2015 IEEE Region 10 Conference, Macau: IEEE, 2015. p. 1–5.

[4] L. Breiman, Random forests. Maching Learning, 45(1):5–32, Oct. 2001.

[5] A. Brook, R. El-Yaniv, E. Issler, R. Kimmel, R. Meir, D. Peleg, Breast cancer diagnosis from biopsy images using generic features and SVMs, 2007, p. 1–16.

[6] E. Buabin, Boosted hybrid recurrent neural classifier for text document classification on the reuters news text corpus, in: International Journal Machine Learning Computing, pp. 588–592, October, 2012.

[7] H. Chen, Q. Dou, X. Wang, J. Qin, P.A. Heng, Mitosis detection in breast cancer histology images via deep cascaded networks, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI 2016, Phoenix, Arizona, pp. 1160-1166.

[8] D.C. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2013, 8150 LNCS(PART 2):411–418.

[9] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, in: Proc. SPIE. vol. 9041. San Diego, California, 2014.

[10] K. Dimitropoulos, P. Barmpoutis, C. Zioga, A. Kamas, K. Patsiaoura, N. Grammalidis, Grading of invasive breast carcinoma through Grassmannian VLAD encoding, in: PLoS ONE, vol. 12, no. 9, Article ID e0185110, 2017.

ACCEPTED MANUSCRIPT

[11] J.G. Elmore, G.M. Longton, P.A. Carney, B.M Geller, T. Onega, A.N.A Tosteson, Diagnostic concordance among pathologists interpreting breast biopsy specimens, in: JAMA 2015.

[12] P. Filipczuk, T. Fevens, A. Krzyżak, R. Monczak, Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies, in: IEEE Transactions on Medical Imaging, vol. 32, no. 12, pp. 2169–2178, 2013.

[13] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, in: Journal of Computer and System Sciences , 55 (1), Aug. 1997.

[14] J.H. Friedman, Greedy function approximation: A gradient boosting machine, in: The Annals of Statistics, 29 (5), 1189–1232, October, 2001.

[15] J. Friedman, Stochastic gradient boosting, in: Computational Statistics and Data Analysis, 38(4): 367–378, 2002.

[16] Y. Gao, W. Rong, Y. Shen, Z. Xiong, Convolutional neural net-work based sentiment analysis using adaboost combination, in: Proc IEEE IJCNN. Vancouver, Canada, pp. 1333–1338, Jul. 2016.

[17] Y.M. George, H.L. Zayed, M.I. Roushdy, B.M. Elbagoury, Remote computer-aided breast cancer detection and diagnosis system based on cytological images, in: IEEE Trans Med Imaging, vol. 8, no. 3, pp. 949–964, 2014.

[18] M.N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, B. Yener, Histopathological image analysis: A review, in: IEEE Reviews on Biomedical Engineering. 2009.

[19] S. Han, Z. Meng, A. Khan, Y. Tong, Incremental boosting convolutional neural network for facial action unit recognition, in: Proc NIPS. Vol. 29. Barcelona, Spain, pp. 109–117, December, 2016.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[20] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, S. Li. Breast cancer multi-classification from histopathological images with structured deep learning model, in: Scientific Reports, 7, 2017.

[21] A.H. Hassan, L. Mathieu, C. Pierre-Henri, R. Soumali, H. Xiaowei, M. Gabija, Z. Odysseas, D. Muneer, Z. Fenqiang, P. Jonas, S. Manish, G. Adrian, A. Teresa, D.M. Vo, P. Chandan, D. Navdeep, K. Satoshi, B. Zhengbing, V. Arash, Q. Gwenole, CATARACTS: Challenge on automatic tool annotation for cataRACT surgery, in: Medical Image Analysis, 2018.

[22] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition, in: CVPR, 2016.

[23] A.A. Helal, K.I. Ahmed, M.S. Rahman, S.K. Alam, Breast cancer classification from ultrasonic images based on sparse representation by exploiting redundancy, in: 16th International Conference of Computer and Information Technology, Khulna, 2014, pp. 92-97.

[24] N. Karianakis, T.J. Fuchs, S. Soatto, Boosting convolutional features for robust object proposals, in: Tech. Rep. arXiv:1503.06350, University of California Los Angeles, Mar. 2015.

[25] H. Kong, Z. Lai, X. Wang, F. Liu, Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning, in: Neurocomputing, Volume 177, 2016, pages 198-205, ISSN 0925-2312.

[26] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak, Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images, in: Computers in Biology and Medicine. 2013, 43 (10):1563–1572.

[27] C. Loukas, S. Kostopoulos, A. Tanoglidi, D. Glotsos, C. Sfikas, D. Cavouras, Breast cancer characterization based on image classification of tissue sections visualized under low magnification, in: Computational and mathematical methods in medicine, Aug 31, 2013.

ACCEPTED MANUSCRIPT

[28] M. Macenko, M. Niethammer, J.S. Marron, D. Borland, J.Y. Woosley, X. Guan, A method for normalizing histology slides for quantitative analysis, in: Proceedings—2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009. Boston, Massachusetts, 2009, p. 1107–1110.

[29] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.

[30] M.T. McCann, J.A. Ozolek, C.A. Castro, B. Parvin, J. Kovacevic, Automated histology analysis: Opportunities for signal processing, in: IEEE Signal Processing Magazine. 2015, 32(1):78.

[31] N. Nayak, H. Chang, A. Borowsky, P. Spellman, B. Parvin, Classification of tumor histopathology via sparse feature learning, in: 2013 IEEE 10th International Symposium on Biomedical Imaging, San Francisco, CA, 2013, pp. 410-413.

[32] A. Pêgo, P. Aguiar, Bioimaging 2015, available from: http://www.bioimaging2015.ineb.up.pt/dataset.html, 2015.

[33] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, M. Lamard, Deep image mining for diabetic retinopathy screening, in: Medical Image Analysis, 39, 178–193, Jul. 2017.

[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, in: International Journal Computer Vision, 115 (3), 211–252L, Apr. 2015.

ACCEPTED MANUSCRIPT

[35] H. Schwenk, Y. Bengio, Boosting neural networks, in: Neural Computing, 12 (8), 1869–1887, Aug. 2000.

[36] Y. Song, J.J. Zou, H. Chang, W. Cai, Adapting fisher vectors for histopathology image classification, in: Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pages 600–603.

[37] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, 2016.

[38] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, in: IEEE Transactions on Biomedical Engineering (TBME), 63(7):1455-1462, 2016.

[39] C. Szegedy, S. Ioffe, V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, in: CoRR , abs/1602.07261, 2016.

[40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the inception architecture for computer vision, in: CVPR, 2016.

[41] J. Tang, R.M Rangayyan, J. Xu, I.E. Naqa, Y. Yang, Computer-aided detection and diagnosis of breast cancer With mammography: recent advances, in: IEEE Transactions on Information Technology in Biomedicine, 2009, 13(2):236–251.

[42] T. Tieleman, G. Hinton, Divide the gradient by a running average of its recent magnitude, in: COURSERA: Neural Networks for Machine Learning, 4, 2012. Accessed: 2015-11-05.

[43] M. Veta, J.P. Pluim, P.J van Diest, M.A Viergever, Breast cancer histopathology image analysis: A review, Biomedical Engineering, in: IEEE Transactions on. 2014 May, 61(5):1400-11.

ACCEPTED MANUSCRIPT

[44] D.M. Vo, S.-W. Lee, Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions, in: Multimedia Tools and Applications.

[45] D.M. Vo, L. Jiang, A. Zell, Real-time multiple person detection and tracking using RGB-D images for mobile robots, in: IEEE International Conference on Robotics and Biomimetics 2014, Bali, Indonesia.

[46] D.M. Vo, S.-W. Lee, Robust face recognition using hierarchical collaborative representation, in: Information Sciences, Vol. 432, March 2018, pp. 332-346.

[47] D.M. Vo, A. Zell, Real-time face recognition using local ternary patterns with collaborative representation-based classification for mobile robots, in: Intelligent Autonomous Systems 13, 2016, pp. 781-793.

[48] B. Zhang, Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles, in: 4th International Conference on Biomedical Engineering and Informatics (BMEI). vol. 1. Shanghai: IEEE, 2011. p. 180–186.

[49] Y. Zhang, B. Zhang, F. Coenen, J. Xiao, W. Lu, One-class kernel subspace ensemble for medical image classification, in: EURASIP Journal on Advances in Signal Processing 2014, 1–13 (2014).

[50] L. Wang, B. Zhang, J. Han, L. Shen, C.S. Qian, Robust object representation by boosting-like deep learning architecture, in: Signal Process Image Commun 47, 490–499, Sep. 2016.

[51] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):210–227, 2009.

[52] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: CVPR, 2011.

ACCEPTED MANUSCRIPT

Duc My Vo

Sang-Woong Lee

ACCEPTED MANUSCRIPT