IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 23, Issue 4, Ser. II (Jul. –Aug. 2021), PP 30-42
www.iosrjournals.org

Breast Cancer Histological Image Classification using EnsembleConvolutional Neural Networks and Triplet Loss

Omar Ghoneim[a], Ghada Soliman[a, b, c], Amr Galal[a], Heba Mahgoub[a]

[a]*Orange Labs Egypt, Smart Village, Km 28 Cairo-Alexandria Desert Road, Sixth of October City, Giza, Egypt, 12577*

[b]*PhD in Environmental Engineering, Institute of Environmental Studies and Research, Ain Shams University, Cairo, Egypt*

[c]*Corresponding Author Ghada Soliman, ghada.soliman@gmail.com*

Abstract.

Purpose: *Breast cancer is the most common type of cancer in women worldwide, accounting for 25.4% of the total number of new cases diagnosed in 2018. This study investigates the problem of classifying Breast Cancer histology images dataset, which is considered a challenging task. The challenge lies in detecting the predominant cancertype in the whole-slide images (WSI). We propose training an ensemble architecture that consists of three convolutional neural networks: Inception-v3, Densenet-121 and ResneXt-101 optimized using triplet loss. The three networksshowed great performance on medical image processing and analysis. Due to the limited number of images in the training dataset, we used transfer-learning to initialize the weights of our networks. We opted for using the triplet lossduring training due to the similarity of some cell types of histology images to produce good representations for the medical images and better separate the different classes. The proposed model with triplet loss proved to be effective with the small dataset due to the huge number of unique triplet samples that are produced for the training step. We trained the ensemble network using the ICIAR 2018 grand challenge on Breast Cancer Histology (BACH) dataset. Images are classified into four classes, normal tissue, benign lesion, in-situ carcinoma and invasive carcinoma.*

Results: *The proposed model achieves 92% accuracy on the test dataset compared to the previously reported 87% accuracy rate in the BACH literature.*

Conclusions: *In this paper we present our ensemble model for the challenge and show how the triplet loss helped the model converge faster and improved the model accuracy in the classification task and how the visualization of thepredicted embeddings can help us display the model behavior and figure out the challenging diagnoses and how it can be addressed.*

Keywords: *Bach challenge, Breast cancer, Deep learning, Histology.*

Date of Submission: 07-08-2021 Date of Acceptance: 21-08-2021

I. Introduction

Breast cancer is ranked as the top cause of death for women.[1] In 2017, studies showed that approximately 252,000 new patients with invasive breast cancer and 63,000 cases of in-situ breast cancer are expected to be diagnosed.[2] Therefore, early detection is crucial to increase the chancesof survival.

A biopsy of breast tissue allows the pathologists to histologically evaluate the structure and elements of the tissue. The goal of Histopathology is to differentiate between normal, benign, in-situ, and invasive breast cancer tissues. Different types of tissues are seen in Fig 1 The fundamental diagnosis is based on the examination of Hematoxylin and Eosin (H&E) stained tissue samples. The manual analysis of tissues is a time-consuming, difficult, and subjective process, which leads to interobserver variations. In order to handle the subjectivity problem and reduce workload for the specialists, automatic computer aided diagnoses have become essential to improve diagnosis efficiency.

Classification of histology images in whole-slide images is considered a challenging task. Recently, multiple research papers were published for the sake of automating breast cancer detection by analyzing microscopic biopsy images using Computer Vision techniques,[3].[4] Two approaches were followed to tackle this problem. The first approach is based on handcrafted features such astexture extraction[5] or nuclei shape and size measurements, which can be calculated using color segmentation as in.[6] The second approach is based on deep learning, as deep learning has outperformed handcrafted features in multiple image analysis tasks such as image segmentation,[7] satelliteimagery, and feature detection.[8]

Convolutional neural network (CNN) has shown an excellent performance in natural image classification tasks as proposed in A. Krizhevsky et al.[9] and K. He et al.[10] Therefore, there is an increasing tendency to adapt CNN for medical images. Several works that use CNN for breast cancer histological image classification have been reported. The reported works done on classification are divided into two approaches. Some studies are based on patch-wise classification, while others are based on image-wise classification. Different patch sizes were reported in the literature (e.g. from 32x32 to 512x512). Most of the related works were based on CNNs trained from scratch onsmall size datasets[11] ,[12] . Models based on CNNs architectures like DenseNet and ResNet have proven to be among the highest performers on the BACH challenge.

In our work, we investigate the efficiency of triplet loss optimization with the significance of concatenating features from different pretrained models such as Inception-v3,[13] Densenet,[14] and Resnext[15] to produce suitable embeddings for the H&E-stained breast cancer histology images. Transfer learning is used to initialize the weights of the proposed model using the weights ofthe ImageNet dataset. Transfer-learning allows the network to learn progressively and adapt new features that represent the deep anatomic information of the medical image. The rest of the paper is organized as follows. Section 2 gives a summary of the related work. Section 3 gives a detaileddescription of the data set, the preprocessing steps, the approach, and the models in which we used to perform the classification task. Following, Section 4 shows the performance of our approach. Finally, Section 5 summarizes our findings.

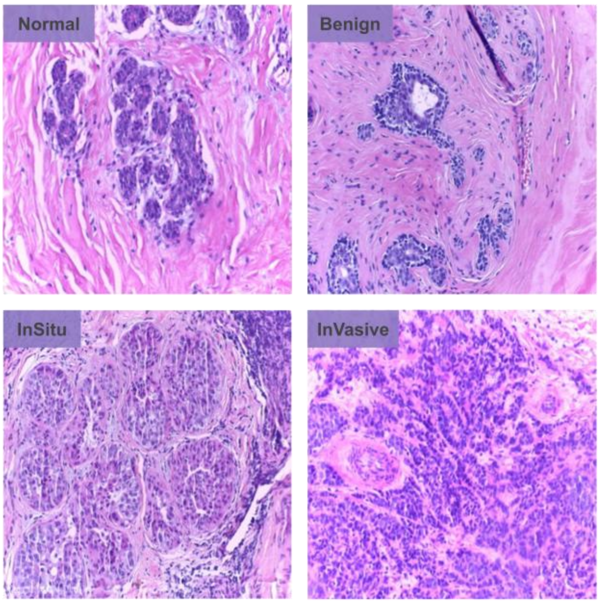


Fig 1 Example of different types of breast tissues from BACH dataset.

II. Related Work

Models based on CNNs are so far the highest performers on the BACH challenge. Using DenseNet which was pretrained on ImageNet,[14] Kohl et al.[16] were able to obtain 83% on the training data.

In their work, they normalized the resulting images of downsampling by a factor of 10% to unit standard deviation and zero mean. The network weights were initialized by freezing the convolution layers and training the network for 25 epochs. The convolution layers were then unfreezed andtraining was resumed for 250 more epochs. Brancati et al.[17] proposed an ensemble model which is composed of 3 ResNet models[10] with 34, 50 and 101 layers achieving 86% accuracy. In their work, images were downsampled by a factor of 80% and only the center batch of size m x m wasconsidered in training, where m is the minimum of the width and height of the resized image. Each ResNet model was trained only on the center batches and at inference time test images get classified according to the class having the highest probability among all the output classes probabilities.Another ensemble model was proposed by Vladimir Iglovikov and Alexey Shvets[18] and obtaineda 87% accuracy. The proposed model can be viewed as a combination of the above mentioned models with slight modifications. In their work, the authors composed one ResNet-101[10] and twoDenseNet-121[14] to form their ensemble model. Prior to training, they resized the images to 224 x 224 pixels. In addition to that, all images were normalized to unit standard deviation and zero mean. In order to prevent overfitting, they used both ImageNet and BACH dataset to tune the threemodels. Given a test image, they classify it by taking the majority vote of the three models.

T. Shakeel et al.[19] proposed a multi-scale input and multi-feature network (MSI-MFNet) model that takes the concatenation of the image patches processed from four different resolutions (1×, 0.5×, 0.33×, and 0.25×) as an input. The multi-feature output is created by the concatenation of the global average pooling of four depth blocks exploring the four different-scale features. The authors achieved a multi-class classification accuracy as low as 60% for the BACH dataset due tothe difficulty to distinguish between normal and benign classes.

W. Mi et al.[20] adopted a two-stage architecture based on deep learning and machine learning methods for the multi-class classification on BACH dataset. They designed a patch-level classifier using trained CNN, Inception V3, in the first stage. In the second stage, the classification results of all patches were combined into one heatmap, and statistical features were extracted from the generated heatmaps at WSI-level to make the final diagnosis through XGBoost. They achieved 87.2% accuracy on the BACH test set.

Most of the submitted work done on Bach challenges used deep learning. Deep learning is thetrend in medical image analysis, where deep learning approach has replaced feature engineering approach. Deep learning requires a large amount of data in order to produce a generalized model.However, large amount of data is not available in medical image analysis field. So a common approach is to train the model on natural image datasets, such as ImageNet, then fine-tune this model on the medical image dataset. A summary of different deep learning approaches is shown in Table 1

Table 1 Summary of the performance of different deep learning approaches on BACH dataset.

Team	Acc.	Approach	Pre-trained	Ensemble	External sets	Input size	Normalization
Chennamsetty et al	0.87	Resnet-101; Densenet-161	Y	3	N	224x224	N
Kwok, 2018	0.87	Inception-Resnet-v2	Y	N	Y	299x299	N
Brancati et al., 2018	0.86	Resnet-34, 50, 101	Y	3	N	308x308	N
Marami et al., 2018	0.84	Inception-v3	Y	4	Y	512x512	WSIC
T. Shakeel et al., 2020	0.68	MSI-MFNet	N	N	Y	224x224	HECN
W. Mi et al., 2021	0.87	Inception-v3-XGBoost	N	N	Y	1024x1024	N

WSIC is the whole-slide image color standardizer proposed in[21]
HECN stands for H&E color normalization[22]

2.1 Contributions and Organization

We participated in the BACH challenge for our interest in detecting cancers from Hematoxylin and Eosin (H&E) stained breast-cancer histopathology images, which remains an important research stream in medical image processing. In our work, we used the microscopy images of part A for the training that are of size 2048 x 1536 pixels without splitting them into patches whilst the majorityof the contributions on this competition deal only with small regions of interest (ROIs), or split the images into patches from either the whole slide or microscopy images as conducted by W. Mi et al.[20] and Spanhol et al.[23] respectively. The patch-based approach introduces challenges, as it is computationally expensive due to the estimation of a high number of parameters.[24] It also has limitations, as the patch may only contain normal tissue and be labeled with a different class. Hence, we assessed the performance of the ensemble of the features of Densely Connected Convolutional Networks (DenseNet) developed by Huang et al.,[14] ResNeXt proposed by Xie et al.[15] and Inception proposed by Szegedy et al.[29] to deal with the labeled microscopy image dataset available at ICIAR 2018 BACH Grand Challenge without the need for ROI selection or WSI patch-based approach. This ensemble yielded a good performance with triplet loss, achievingan accuracy of 92% on the test set prediction submission on ICIAR 2018 challenge, and increasingthe computation efficiency by finishing the training in reasonably short time. Inception, DenseNetand ResNeXt are described in section 3.3 and the ensemble of these pre-trained models is explainedin section 3.4. The evaluation of our approach is described in section 4 before we conclude our paper with Section 5.

III. Materials and Methods

3.1. Dataset

The BACH challenge made available two labeled training datasets for the registered participants.The first dataset is composed of microscopy images annotated image-wise by two expert pathologists from the Institute of Molecular Pathology and Immunology of the University of Porto (IPA- TIMUP) and from the Institute for Research and Innovation in Health (i3S). The second dataset contains pixel-wise annotated and unannotated WSI images. For the WSI, annotations were performed by a pathologist and revised by a second expert.[24] The focus of this paper is on part A, microscopy images dataset, the dataset is composed of high-resolution (2040 x 1536 pixels) images, uncompressed, and annotated H&E stain images from the Bioimaging 2015 breast histology classification challenge.[25] All the images are digitized with the same acquisition conditions, with magnification of 200x and pixel size of 0.42µm x 0.42µm. Each image is labeled with one of four classes: normal tissue, benign lesion, in-situ carcinoma, and invasive carcinoma. The labeling was performed by two pathologists, who only provided a diagnostic from the image contents,without specifying the area of interest for the classification. Cases of disagreement between specialists were discarded. The goal of the challenge is to provide automatic classification of each input image.[12]

The microscopy dataset is composed of 400 training and 100 test images, with the four classes equally represented and balanced. The images were selected so that the pathology classification can be objectively determined from the image contents. The training and test datasets are publiclyavailable at ICIAR 2018 BACH Grand Challenge under the CC BY-NC-ND license.

3.2 Preprocessing

The training dataset Images are passed through a pipeline of random transformations: All imagesare resized to 299 x 299 while maintaining the 3-channels RGB image data format.

1. The images are randomly rotated by a range of [-20, 20] degrees while duplicating the pixelson the edges to fill the missing pixels resulting from rotation.
2. The images are randomly shifted vertically and horizontally by a range of [-60, 60] pixels while duplicating the pixels on the edges to fill in the missing pixels resulting from shifting.
3. The images are randomly zoomed to a range of [0.8, 1.2] of the original images while maintaining the image size and duplicating the pixels on the edges to fill in the missing pixels incase of zooming out.
4. The images are randomly flipped vertically and horizontally with a probability of 0.5 for each.
5. Finally, the images are mapped to a range of [0, 1] by min-max normalization technique.

Considering the original dataset range is [0, 255], the normalization function is as follows:

$$F(X) = X * \frac{1}{255.0} \qquad (1)$$

Similarly, the test dataset images are resized to 299 x 299 pixels and normalized to a range of [0, 1] without random transformation. Fig 2 shows four training images after applying these transfor mations. In this study, the dataset of 400 images was mainly used for training without splitting it totraining and validation. It is difficult to perform proper data splitting, such as train-validation-test splitting or cross-validation without yielding overly optimistic results.[24]

3.3 Backbone Models

3.3.1 Inception

The term “Inception” was first introduced by Szegedy et al[26] proposing GoogLeNet model that was responsible for setting the new state of the art level in ImageNet LSVRC 2014 challenge in both classification and detection. The Inception architecture is based mainly on the idea of concatenating a set of parallel convolutional feature maps into a single vector for each inception module with the output vector forming the input of the next stage.

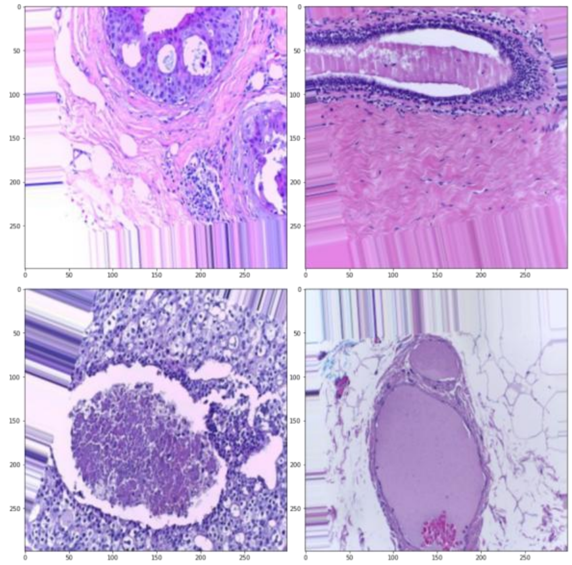


Fig 2 Example of training images after transformation.

A sequence of inception modules can form a full inception feature extractor followed by a fully connected classifier as in inception-v3 modelarchitecture illustrated in Fig 3. This concept was developed and introduced in various incarnations like BN-GoogLeNet, BN-Inception,[27] Inception-v2 and inception-v3.[13] The one we are concernedabout in this paper is Inception-v3 which conducted a new state-of-the-art results on the ILSVRC-2012 validation set and the ensemble of 4 inception-v3 models represented almost half of the error of GoogLeNet ensemble. In the Inception v3 model, several techniques for optimizing the networkhave been suggested to loosen the constraints for easier model adaptation. The techniques include factorized convolutions, regularization, dimension reduction, and parallelized computations. The factorized convolutions help in increasing the computational efficiency as it reduces the number ofparameters involved in a network.

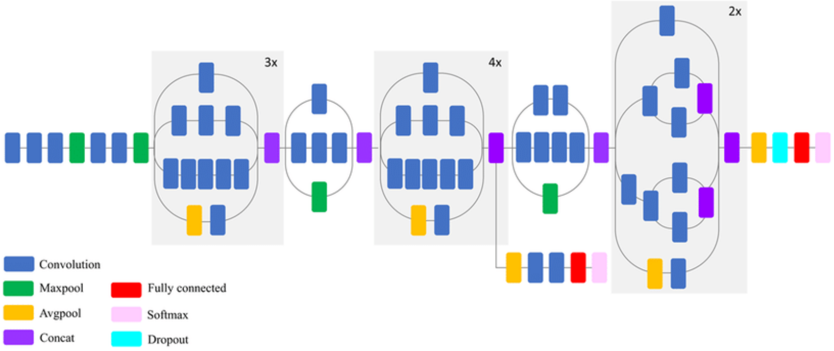


Fig 3 Inception-v3 architecture. Reprinted from Mahdianpari et al.[28]

3.3.2 *DenseNet*

Dense convolutional networks (DenseNet) developed by Huang, Liu and Maaten[29] had the best classification performance on publicly available image datasets such as CIFAR-10 and ImageNet in 2017. It builds on the idea of ResNet, but instead of summing the residuals as in ResNet, DenseNet concatenates all the feature maps. It has high computational and memory efficiency.It was built with a structure that connects each layer to later layers. Each layer obtains additional inputs from all preceding layers and passes on its own feature maps to all subsequent layers. DenseNet architecture uses the residual mechanism to its maximum by making every layer of the same dense block connect to its subsequent layers as shown in Fig 4. This model's compactness makes the learned features non-redundant as they are all shared through a common knowledge. DenseNets performs well when training data is insufficient since DenseNet uses features of all complexity levels. Hence Densenet is particularly well-suited for smaller datasets outperforming many previous models on datasets like Cifar-10 and Cifar-100.

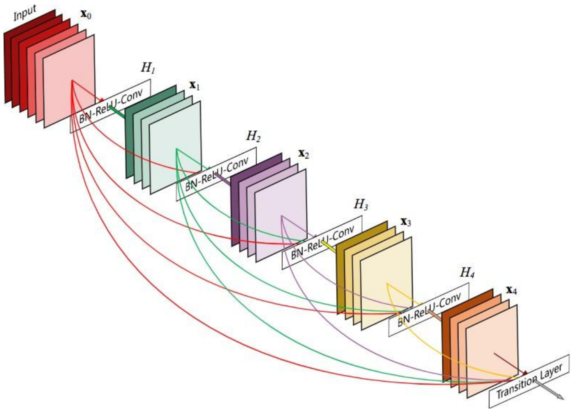


Fig 4 A 5-layer dense block from DenseNet architecture. Reprinted from DenseNet paper[14]

3.3.3 *ResNeXt*

Facebook AI Research proposed a variant of ResNet that is codenamed ResNeXt.[15] It is an extension of the deep residual network which replaces the standard residual block with the one that leverages a ”split-transform-merge” strategy that has been adopted in the Inception models.[26] In an Inception module, the input is split into a few lower- dimensional embedding (by 1x1 convolutions), transformed by a set of specialized filters (3x3, 5x5, etc.), and merged by concatenation. Although Inception model achieves good accuracy, it has been accompanied by a series of complications. For instance, the filter numbers and sizes were tailored for each individual transformation, and the modules were customized stage-by-stage. ResNeXt exploits the split-transform-merge strategy in an easy and extensible way. In ResNeXt module, the input is split into a few lower-dimensional embeddings with the same set of transformations in which outputs are aggregated bysummation as illustrated in Fig 5. For these transformations to be aggregated, they all must be of the same topology. This design allows the extension to any large number of transformations without specialized designs compared to Inception model, which makes it easy to adapt to new datasets or tasks. ResNeXt introduced a new hyper-parameter called cardinality which refers to the numberof independent paths or branches. The aim of cardinality is to provide a new way of adjusting the model capacity. ResNeXt showed that increasing cardinality is more effective at benefiting modelperformance than increasing the width or depth of the network.

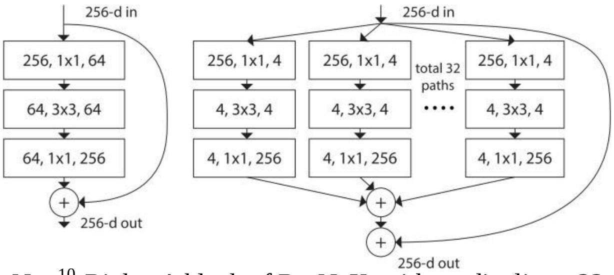


Fig 5 Left: A block of ResNet.[10] Right: A block of ResNeXt with cardinality = 32. Reprinted from ResNeXt paper.[15]

3.4 Ensemble of Pre-trained CNN Models

Classification of breast cancer histology images is a very complex task as these images have normal and abnormal biological structures, morphological and architectural characteristics. These images exhibit relevant patterns that generally have high visual appearance variability.[30] These characteristics make it difficult to learn the discriminative features among the classes. Convolutional neural networks have proven to be effective in building classifiers to extract these discriminative featuresof the classes. In this study, we investigated transfer-learning to address the challenge of limited training data of BACH images dataset, to reduce the need of a large training set. There are many ways to use pre-trained models in which the selection generally depends on the size of the datasetand the extent of computational resources available. Our proposed architecture relies mainly on the concatenation of the features derived from the pre-trained deep CNN models: Inception-v3, ResNeXt-101 and DenseNet-121. The motivation behind using more than one pre-trained model isthe possibility to provide some non-overlapping information, allowing superior performance whencombined. Using the triplet loss function, our model optimizes the image embeddings for better representation of images, where embeddings of images belonging to the same class are forced to be close in the Euclidean space while those of different classes are pushed away from one another.We run these steps for fine-tuning of each of these CNN models in the proposed architecture as shown in Fig 6:

1. Initializing the CNN model with the weights of a pre-trained network that is trained on the 1000-class Imagenet dataset.

2. Adding a convolution layer to each CNN model to reduce the output feature maps as follows:
   a. InceptionV3 feature map is reduced from 2048 channels to 64 by adding a 2D convolutional layer with 64 filters, kernel size of 2x2, 1 stride and same padding.
   b. ResNeXt-101 feature map is reduced from 2048 channels to 32 by adding a 2D convolutional layer with 32 filters, kernel size of 2x2, 1 stride and same padding.
   c. DenseNet-121 feature map is reduced from 1024 channels to 32 by adding a 2D convolutional layer with 32 filters, kernel size of 2x2, 1 stride and same padding.

   This method of reducing the number of output features has proven to achieve a better test accuracy than average and max pooling as demonstrated in Table 3. All the reduced outputsare activated with Rectified Linear Unit (ReLU).

3. The concatenated reduced features of DenseNet-121 and ResNeXt-101 are encoded with an L2-normalized fully-connected layer with 256 neurons. A similar encoding layer is added tothe reduced features of InceptionV3.

4. The encodings of the ensemble model are concatenated and connected to a fully-connected layer with L2-normalization to output an embedding representing the input image in a 256-dimensional space.

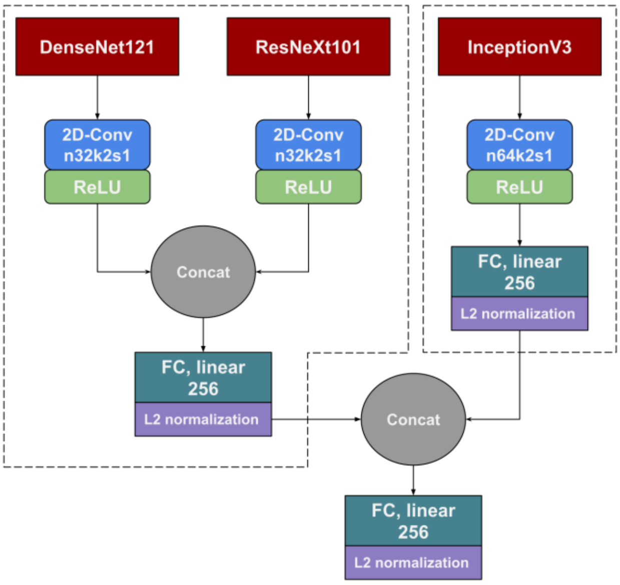


Fig 6 Full model architecture

3.5 Loss Function

Introduced by Kilian Q. Weinberger, John Blitzer and Lawrence K. Saul[31] for nearest neighbor classification, and is what FaceNet relies on for face verification, recognition and clustering.[32] Thetriplet loss is used for learning efficient encoding for data points, by adjusting the distances between points from similar and different classes. The loss operates on triplets: a pivot embedding (anchor), an embedding of the same class as the anchor (positive) and an embedding from a different class (negative). The distance between the anchor and the positive example is then minimized, while the distance between the anchor and the negative example is maximized. In order to represent the error term accordingly, the loss function is described as an Euclidean distance function 2. Where A is the anchor, P is the positive input, N is the negative input, α is the margin between positive and negative pairs, and f is the image embedding vector outputted by the model.

$$l(A, P, N) = max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0) \qquad (2)$$

The selection of triplets is crucial for training efficiency, convergence time and model accuracy. To achieve the best possible performance, we used the semi-hard negative selection approach as mentioned in FaceNet paper.[32] According to this approach, the semi-hard negative example wouldbe selected such that it is farther away from the anchor than the positive example within a specificmargin as demonstrated in Fig. 7.

Online learning approach is necessary for the training process in order to fit in the limitations of the hardware resources making the triplet selection stage operates efficiently within reasonablylow space and time complexity. In order to utilize online learning, we divide our dataset into mini- batches, where the semi-hard triplets are selected from each batch independently for every trainingstep regardless of the rest of the dataset distribution. This online triplet mining approach is provento boost training efficiency as selecting triplets from the whole dataset at once is more time and space consuming.

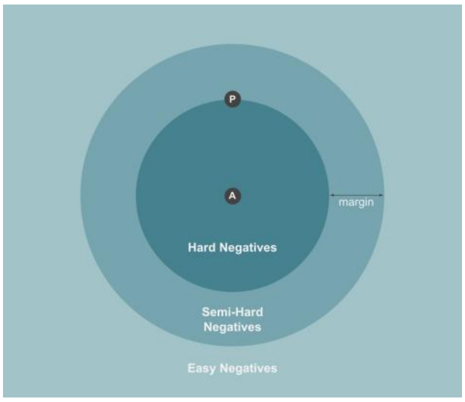


Fig 7 Hard, semi-hard and easy negative examples area. (A) for anchor and (P) for positive example.

3.6 Data Visualization and Analysis

In order to monitor the training iterations and observe changes to the distribution of the predicted clusters from the dataset, we need to visualize the two dimensional representation of the predictedembeddings from the training and test examples. We used Singular Value Decomposition (SVD) to reduce the dimension of the image embedding for the sake of visualization. The whole dataset predicted embeddings are visualized for each epoch to evaluate the performance of the training process and measure the consistency among the embeddings of the training and test image.

Data visualization analysis shows that the benign diagnosed microscopic images are the most challenging ones as they tend to be the last convergent examples in training phases for the majorityof the models as demonstrated in Fig 8. As a result, we found out that “Benign” training exampleswere eventually converged by the complexity of the model due to potential failure in the generalization of few benign test examples resulting in mis-classification to either normal or malignantthat is in line with the same results observed by the experts on the same dataset.[24] This conclusion can also be observed in the embedding visualizations of the converged data points in Fig 9. The test points reasonably exist near the centers of the training data clusters except for the benign class,despite the clear separation of the clusters.

3.7 Training

The full model is divided into two sub-models as shown in Fig 6, each sub-model is trained separately until convergence in order to produce two different set of features, different point of views, for the dataset images. After 70 epochs of each model training, we assembled the full model and resumed training updating all the weights without freezing any part of the model. We used Adam optimizer with a learning rate of 1e-4 declined after 30 epochs to 1e-5, beta 1 of 0.90, beta 2 of 0.99 and a 1e-7 epsilon.

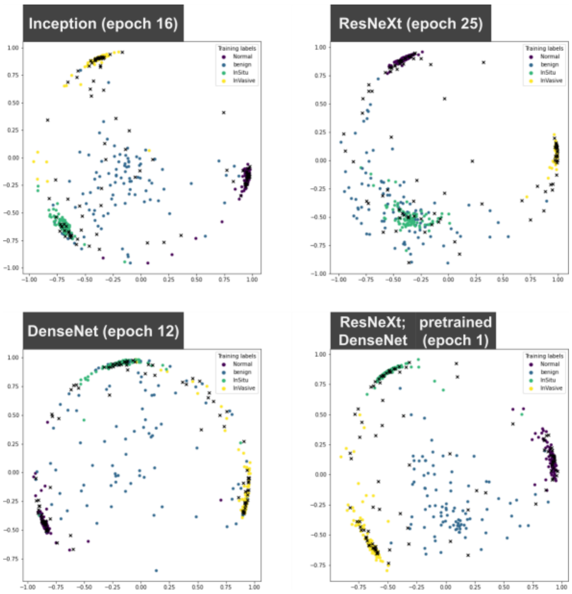


Fig 8 Embeddings visualization at early stages of training at different epochs shows the scattering of the benign datapoints. Colored dots represent labeled training images while x points represent the unlabeled test images.

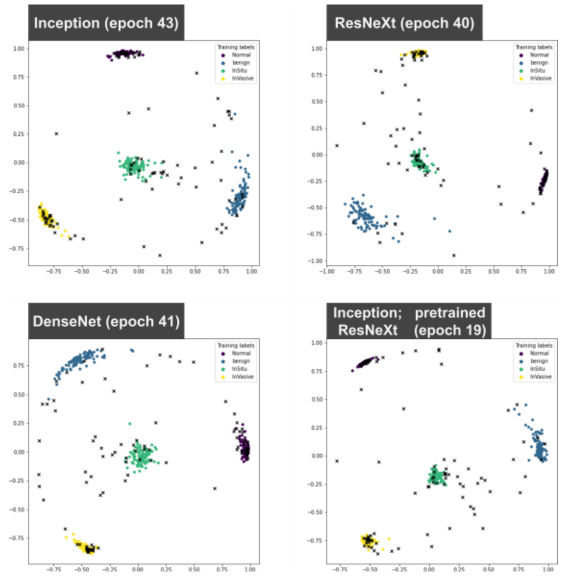
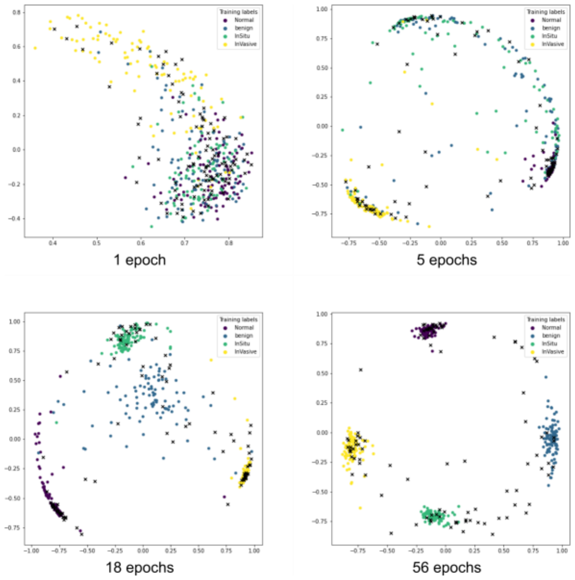


Fig 9 Embeddings visualization of the data after convergence at different epochs. The figure shows the distribution ofembeddings for the training and test images.

Due to GPU memory constraints, our model was trained with a batch size of 10. The training loss and accuracy started to converge after around 50 epochs for each sub-model. Generally, the dataset points are gradually separating into clusters over training epochs as shown in Fig 10. All models are trained on a single NVIDIA Tesla P100, 2 CPU cores, and 13Gigabytes of RAM. The full model was implemented and trained using the TensorFlow frameworkwith most of the operations done using its high level API tensorflow.keras.



1 epoch

5 epochs

18 epochs

56 epochs

Fig 10 Embeddings visualization over training epochs in one of the experiments (Inception model).

3.8 *Class Prediction*

Classifying the images requires additional post processing techniques since the model merely outputs an embedding for the image. To specify the class to which the image belongs, we need to figure out where the image embedding is located with respect to the clusters of the training set embeddings. We experimented with different clustering-based classification techniques like K-Nearest Neighbor (KNN) and Gaussian Mixture Model (GMM). We also tried estimating the predicted class using different distance calculation techniques like Euclidean and Manhattan distance. In conclusion, all the previous methods came out with similar results as shown in Table 2. We can assume that the H&E images are well represented by the embedding generated by model that any simple classification method can discriminate between the four classes depending on the distribution of the training data points resulted from the model.

Table 2 Comparing the results of different prediction methods

Prediction Method	Accuracy
KNN (K=50)	0.92
GMM	0.91
Manhattan Distance	0.92
Euclidean distance	0.92

3.8.1 K-Nearest Neighbor (KNN)

KNN is a simple classification algorithm that depends only on the distribution of the labeled training data to classify any new test point according to the class of the nearest K training points. We used KNN classifier with K set to 50 to classify the H&E images according to the embeddings ofthe training data. This prediction method produced a test accuracy of 0.92 which resembles the highest accuracy proposed by this paper.

3.8.2 Gaussian Mixture Model (GMM)

GMM is a statistical model that approximate any data points with a predefined number of Gaussians to separate the unlabeled data into a set of clusters with every point having a probability of being assigned to each cluster. The gaussian mixtures can also be useful in the supervised problems given that the number of Gaussians is defined by the number of classes with each Gaussian’scenter defined by the mean of its class data points. We used GMM algorithm with the number of Gaussians set to four and the initial center of each Gaussian determined by the mean of each of thefour classes embeddings. This prediction method produced a test accuracy of 0.91 which doesn’t pass the score of the previous KNN method.

3.8.3 Similarity With Median

Another simple prediction technique is done using a similarity measure between the median of each class in the training set embeddings calculated by 3, and the test image embedding generatedby the model. We tried two different similarity measures, Manhattan distance 4 and Euclidean distance 5, to determine the predicted class where the test image is assigned to the same class as the nearest median embedding. We preferred to use the median due to its relatively low sensitivityto outliers when compared with the mean. Each of the two similarity measures produced a test accuracy of 0.92 which resembles the score of the KNN method.

$$P_i = med(f(X^j)) \tag{3}$$

Class embedding equation, where P_i is the point of class i and j ϵ [0, m] for m training examples.

$$C(X) = argmin \left[\frac{1}{n} \sum_{j=0}^{n} \left| f(X)^i - P_i^j \right\| \right]_{i=0}^{k} \tag{4}$$

Class prediction equation with Manhattan distance, Where C is the predicted class, X is the inputimage, n is the number of elements in the output embedding vector, f is the embedding output by the model and P_i is the point of class i for i ϵ [0, k].

$$C(X) = argmin_i \; || f(X) - P_i \; || \tag{5}$$

Class prediction equation with Euclidean distance, Where C is the predicted class, X is the input image, f is the embedding output by the model and P_i is the point of class i

IV. Results

Among all the experiments, the top performance hypothesis was the ensemble (Inception-v3, DenseNet-121, ResNeXt-101) reduced feature maps model reaching a training accuracy of 99.75%and a 0.035 training loss with the batch size set to 10. Due to the unavailability of the labels ofthe test set, we submitted our work to the competition ICIAR 2018 BACH Grand Challenge toassess the accuracy of our model. The prediction results of the test images for part A achieved anaccuracy of 92%. The proposed model ranked the sixth for part A, taking into account that thisresult is in line with the average pathology experts accuracy of 85±10%.[24] It is noticed that the top five participants on the Leaderboard of the competition did not make their work publicly available.

We tried different hypotheses with different losses as demonstrated in Table 3. Overall, the semi-hard triplet loss operated on L2-normalized output seems to outperform the cross-entropyloss operated on softmax activated output in standalone models, while the ensemble models seemto produce similar results for both loss functions. The test accuracy did not pass 89% for all of theexperiments except for the ensemble of the reduced feature maps of these models (Inception-v3, DenseNet-121, ResNeXt-101) with semi-hard triplet loss. Fig. 11 shows the images embeddings of the best test accuracy model that demonstrates a clear separation among the four classes.

According to Guilherme et al.,[24] most of the previous works used the majority voting scheme for inference, by assigning a test image the class with the highest number of votes among the usedmodels. However our approach relies on representing images by vectors, and during inference weuse a distance metric to measure the distance between a test image vector representation and the stored classes vectors. We then assign the closest class to the test image. The results indicate thatour approach achieves higher accuracy than the methods in the literature evaluated on the same benchmark dataset.

Table 3 Comparing experimental results of single and ensemble models with cross entropy loss and semi-hard tripletloss

Model	Cross entropy	Triplet loss
Inception-v3	0.83	0.88
DenseNet-121	0.83	0.82
DenseNet-201	0.84	0.84
ResNeXt-101	0.84	0.88
(Inception-v3; DenseNet-201; ResNeXt-101;) avg pooling	0.89	0.89
(DenseNet-121; DenseNet-201; ResNeXt-101;) avg pooling	0.85	0.83
(Inception-v3; DenseNet-121; ResNeXt-101;) reduced maps	0.88	**0.92**

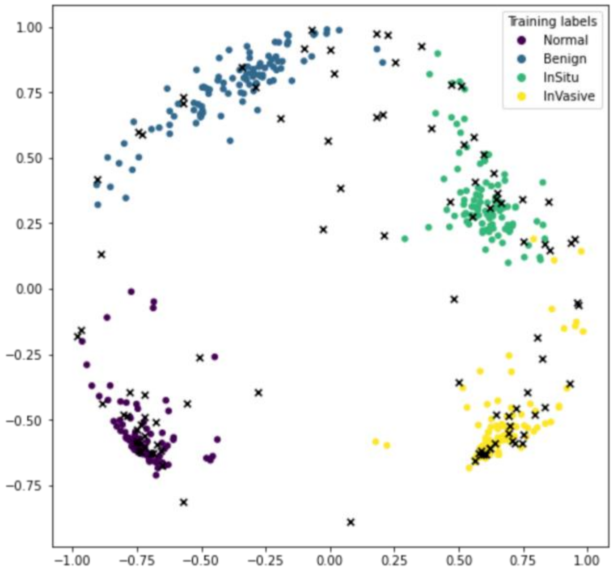


Fig 11 Scatter of embeddings from the highest test accuracy model. Colored dots represent labeled training imageswhile x points represent unlabeled test images.

V. Conclusion

In this work, we tackled the problem of classifying Breast Cancer histology images dataset. We proposed a deep learning model that utilizes three of the state of the art convolution neural networks, namely Inception-v3, DenseNet-121 and ResNeXt-101 that are efficiently optimized with semi-hard triplet loss. The proposed approach yields 92% accuracy on BACH test dataset compared to 87% accuracy previously reported rates in.[24]

Disclosures

All the authors of this manuscript have no relevant financial interests and no other potential conflicts of interest related to the manuscript.

References

[1]. L. A. Torre, F. Bray, R. L. Siegel, *et al.*, "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians* **65**(2), 87–108 (2015).

[2]. C. Desantis, J. Ma, A. Goding Sauer, *et al.*, "Breast cancer statistics, 2017, racial disparity inmortality by state," *CA: A Cancer Journal for Clinicians* **67** (2017).

[3]. S. Bhattacharjee, J. Mukherjee, S. Nag, *et al.*, "Review on histopathological slide analysis using digital microscopy," *International journal of advanced science and technology* **62**, 65–96 (2014).

[4]. C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images : asystematic survey," (2005).

[5]. C. Bergmeir, M. Garcia-Silvente, and J. Ben´ıtez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a web-based software framework," *Computer methods and programs in biomedicine* **107**, 497–512 (2012).

[6]. M. Kowal, P. Filipczuk, A. Obuchowicz, *et al.*, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Computers in biology and medicine* **43**, 1563–72 (2013).

[7]. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," (2018).

[8]. V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A kaggle competition," (2017).

[9]. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems* **25** (2012).

[10]. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *2016 IEEEConference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (2016).

[11]. F. A. Spanhol, L. S. Oliveira, C. Petitjean, *et al.*, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 International Joint Conferenceon Neural Networks (IJCNN)*, 2560–2567 (2016).

[12]. T. Araújo, G. Aresta, E. Castro, *et al.*, "Classification of breast cancer histology images using convolutional neural networks," *PLOS ONE* **12**, e0177544 (2017).

[13]. C. Szegedy, V. Vanhoucke, S. Ioffe, *et al.*, "Rethinking the inception architecture for computer vision," (2015).

[14]. G. Huang, Z. Liu, L. Van Der Maaten, *et al.*, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269(2017).

[15]. S. Xie, R. Girshick, P. Dollar, *et al.*, "Aggregated residual transformations for deep neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017).

[16]. M. Kohl, C. Walz, F. Ludwig, *et al.*, "Assessment of breast cancer histology using densely connected convolutional networks," (2018).

[17]. N. Brancati, M. Frucci, and D. Riccio, *Multi-classification of Breast Cancer Histology Images by Using a Fine-Tuning Strategy*, 771–778 (2018).

[18]. S. S. Chennamsetty, M. Safwan, and V. Alex, *Classification of Breast Cancer Histology Im-age using Ensemble of Pre-trained Neural Networks*, 804–811 (2018).

[19]. T. S. Sheikh, Y. Lee, and M. Cho, "Histopathological classification of breast cancer imagesusing a multi-scale input and multi-feature network," *Cancers* **12**(8), 2031 (2020).

[20]. W. Mi, J. Li, Y. Guo, *et al.*, "Deep learning-based multi-class classification of breast digital pathology images," *Cancer Management and Research* **13**, 4605–4617 (2021).

[21]. B. Ehteshami Bejnordi, G. Litjens, N. Timofeeva, *et al.*, "Stain specific standardization of whole-slide histopathological images," *IEEE Transactions on Medical Imaging* **35**(2), 404–415 (2016).

[22]. E. Reinhard, M. Adhikhmin, B. Gooch, *et al.*, "Color transfer between images," *IEEE Computer graphics and applications* **21**(5), 34–41 (2001).

[23]. F. A. Spanhol, L. S. Oliveira, C. Petitjean, *et al.*, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 international joint conference on neural networks (IJCNN)*, 2560–2567, IEEE (2016).

[24]. G. Aresta, T. Araújo, S. Kwok, *et al.*, "Bach: Grand challenge on breast cancer histology images," *Medical Image Analysis* **56**, 122–139 (2019). Pêgo and A. Paulo *Bioimaging* (2015). Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," (2014).

[25]. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," (2015).

[26]. M. Mahdianpari, B. Salehi, M. Rezaee, *et al.*, "Very deep convolutional neural networks forcomplex land cover mapping using multispectral remote sensing imagery," *Remote Sensing***10**, 1119 (2018).

[27]. G. Huang, Z. Liu, L. van der Maaten, *et al.*, "Densely connected convolutional networks," (2016).

[28]. J. Arevalo, A. Cruz-Roa, and F. González, "Histopathology image representation for automatic analysis: A state-of-the-art review," *Revista Med* **22**, 79–91 (2014).

[29]. K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in neural information processing systems*, 1473– 1480 (2006).

[30]. F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)* (2015).

List of Figures

List of Tables