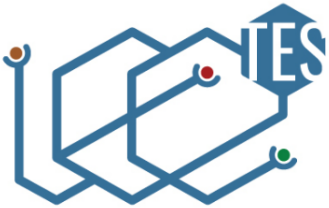
UNIVERSITAS KRISTEN MARANATHA

Breast Cancer Histopathological Image Classification using Progressive Resizing Approach



1ST INTERNATIONAL CONFERENCE ON EMERGING ISSUES IN TECHNOLOGY, ENGINEERING AND SCIENCE

1-2 JULY 2021

Hendra Bunyamin, Hapnes Toba, Meyliana, and Roro Wahyudianingsih

1 July 2021

Introduction (1/3)

Introduction (1/3)

- Breast cancer (BC) is the cause of the second largest deaths among women (Spanhol et al., 2016; Bray et al., 2018; McKinney et al., 2020) in the world.

Introduction (1/3)

- Breast cancer (BC) is the cause of the second largest deaths among women (Spanhol et al., 2016; Bray et al., 2018; McKinney et al., 2020) in the world.
- Data Global Cancer Observatory 2018 from World Health Organization (WHO) showed that the most number of cancer cases in Indonesia is BC with 58,256 of 348,809 patients or 16.7% (World Health Organization, 2018)

Introduction (1/3)

- Breast cancer (BC) is the cause of the second largest deaths among women (Spanhol et al., 2016; Bray et al., 2018; McKinney et al., 2020) in the world.
- Data Global Cancer Observatory 2018 from World Health Organization (WHO) showed that the most number of cancer cases in Indonesia is BC with 58,256 of 348,809 patients or 16.7% (World Health Organization, 2018)
- Patients with suspected breast cancer need to have a *biopsy* which is frequently used to confirm the diagnosis before treatment is planned (Millis, 1984).

Introduction (2/3)

Biopsy diagnosis from histopathological images is considered a *gold standard* for determining almost all types of cancer, specifically BC (International Agency for Research on Cancer, 2012; David S. Strayer, 2014).

(a) (b) (c)

Figure 1:

Introduction (2/3)

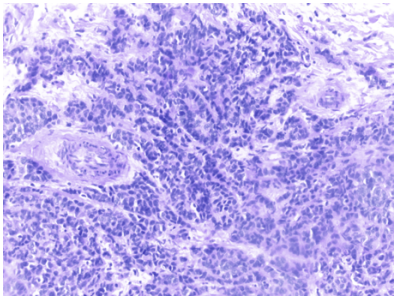
Biopsy diagnosis from histopathological images is considered a *gold standard* for determining almost all types of cancer, specifically BC (International Agency for Research on Cancer, 2012; David S. Strayer, 2014).

(a) (b) (c)

Figure 1: Three images of sample cell images

Introduction (2/3)

Biopsy diagnosis from histopathological images is considered a *gold standard* for determining almost all types of cancer, specifically BC (International Agency for Research on Cancer, 2012; David S. Strayer, 2014).



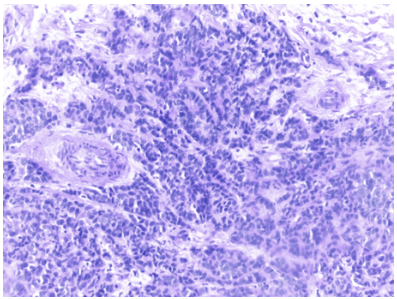
(a) Malignant cells

(b) (c)

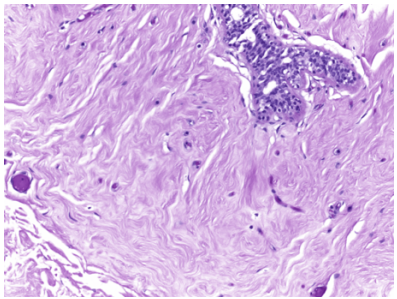
Figure 1: Three images of sample cell images

Introduction (2/3)

Biopsy diagnosis from histopathological images is considered a *gold standard* for determining almost all types of cancer, specifically BC (International Agency for Research on Cancer, 2012; David S. Strayer, 2014).



(a) Malignant cells



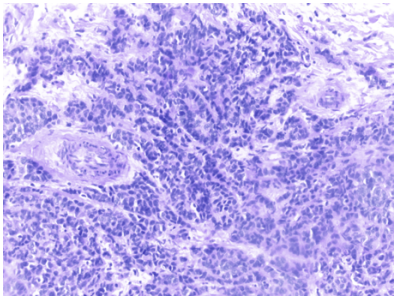
(b) Benign cells

(c)

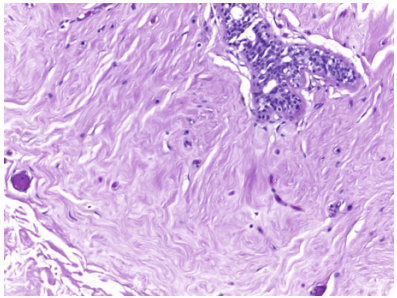
Figure 1: Three images of sample cell images

Introduction (2/3)

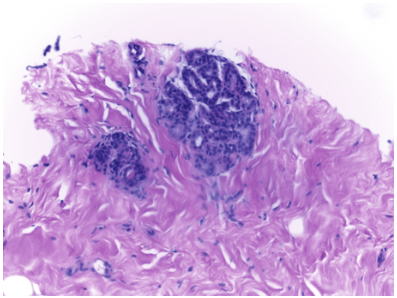
Biopsy diagnosis from histopathological images is considered a *gold standard* for determining almost all types of cancer, specifically BC (International Agency for Research on Cancer, 2012; David S. Strayer, 2014).



(a) Malignant cells



(b) Benign cells



(c) Normal cells

Figure 1: Three images of sample cell images

Introduction (3/3)

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.
- Fortunately, international research organizations such as

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.
- Fortunately, international research organizations such as
  - Pathological Anatomy and Cytopathology (P&D) Laboratorium in Parana, Brazil (Spanhol et al., 2016),

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.
- Fortunately, international research organizations such as
  - Pathological Anatomy and Cytopathology (P&D) Laboratorium in Parana, Brazil (Spanhol et al., 2016),
  - the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) in Porto, Portugal (Aresta et al., 2019), and

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.
- Fortunately, international research organizations such as
  - Pathological Anatomy and Cytopathology (P&D) Laboratorium in Parana, Brazil (Spanhol et al., 2016),
  - the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) in Porto, Portugal (Aresta et al., 2019), and
  - Center for Bio-Image Informatics, University of California in Santa Barbara, USA (Gelasca et al., 2008)

Introduction (3/3)

- *Deep learning* can help classify types of BC with some degrees of automation provided there are datasets.
- Fortunately, international research organizations such as
  - Pathological Anatomy and Cytopathology (P&D) Laboratorium in Parana, Brazil (Spanhol et al., 2016),
  - the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) in Porto, Portugal (Aresta et al., 2019), and
  - Center for Bio-Image Informatics, University of California in Santa Barbara, USA (Gelasca et al., 2008)

  have released publicly available BC histopathological image datasets.

Motivation

Motivation

- Combining different characteristics of the three datasets into a unified dataset will make a rich learning dataset for deep learning models $\Longrightarrow$ better in generalization on unseen data than the one learns from only one dataset (Géron, 2019).

Objective #1

Motivation

- Combining different characteristics of the three datasets into a unified dataset will make a rich learning dataset for deep learning models $\Longrightarrow$ better in generalization on unseen data than the one learns from only one dataset (Géron, 2019).

Objective #1

Firstly this paper also provides a unified dataset as a new publicly available dataset to help this research field progress.

Motivation

- Combining different characteristics of the three datasets into a unified dataset will make a rich learning dataset for deep learning models $\Longrightarrow$ better in generalization on unseen data than the one learns from only one dataset (Géron, 2019).

Objective #1

Firstly this paper also provides a unified dataset as a new publicly available dataset to help this research field progress.

- We train a deep learning which employs a progressive resizing approach (Howard and Gugger, 2020) with the unified dataset preprocessed by the Vahadane preprocessing technique (Vahadane et al., 2016).

Objective #2

Motivation

- Combining different characteristics of the three datasets into a unified dataset will make a rich learning dataset for deep learning models $\Longrightarrow$ better in generalization on unseen data than the one learns from only one dataset (Géron, 2019).

Objective #1

Firstly this paper also provides a unified dataset as a new publicly available dataset to help this research field progress.

- We train a deep learning which employs a progressive resizing approach (Howard and Gugger, 2020) with the unified dataset preprocessed by the Vahadane preprocessing technique (Vahadane et al., 2016).

Objective #2

Secondly, the paper initializes a deep learning model which functions as a baseline on the new unified dataset.

Histopathological Image Datasets: BreakHis

Histopathological Image Datasets: BreakHis

- The BreaKHis consists of $7,909$ BC histopathological images taken from 82 patients from January 2014 to December 2014 with different magnifying factors (40, 100, 200, and 400) (Spanhol et al., 2016).

Histopathological Image Datasets: BreakHis

- The BreaKHis consists of $7,909$ BC histopathological images taken from 82 patients from January 2014 to December 2014 with different magnifying factors (40, 100, 200, and 400) (Spanhol et al., 2016).
- Moreover, the number of benign and malignant cases are $2,480$ and $5,429$ respectively.

Histopathological Image Datasets: BreakHis

- The BreaKHis consists of 7,909 BC histopathological images taken from 82 patients from January 2014 to December 2014 with different magnifying factors (40, 100, 200, and 400) (Spanhol et al., 2016).
- Moreover, the number of benign and malignant cases are 2,480 and 5,429 respectively.
- The benign category includes adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA).

Histopathological Image Datasets: BreakHis

- The BreaKHis consists of 7,909 BC histopathological images taken from 82 patients from January 2014 to December 2014 with different magnifying factors (40, 100, 200, and 400) (Spanhol et al., 2016).
- Moreover, the number of benign and malignant cases are 2,480 and 5,429 respectively.
- The benign category includes adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA).
- Additionally, the malignant category covers ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC).

Histopathological Image Datasets: BACH

Histopathological Image Datasets: BACH

- The BACH was released as a grand challenge on BC histopathological images which is a part of the 15th International Conference on Image Analysis and Recognition (ICIAR 2018).

Histopathological Image Datasets: BACH

- The BACH was released as a grand challenge on BC histopathological images which is a part of the 15th International Conference on Image Analysis and Recognition (ICIAR 2018).
- The annotation for the dataset was done by pathologists from the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) (Aresta et al., 2019).

Histopathological Image Datasets: BACH

- The BACH was released as a grand challenge on BC histopathological images which is a part of the 15th International Conference on Image Analysis and Recognition (ICIAR 2018).
- The annotation for the dataset was done by pathologists from the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) (Aresta et al., 2019).
- Our research employs the microscopy dataset which categorizes BC cells into 1) benign, 2) malignant, 3) in situ carcinoma, and 4) invasive carcinoma.

Histopathological Image Datasets: BACH

- The BACH was released as a grand challenge on BC histopathological images which is a part of the 15th International Conference on Image Analysis and Recognition (ICIAR 2018).
- The annotation for the dataset was done by pathologists from the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Institute for Research and Innovation in Health (i3S) (Aresta et al., 2019).
- Our research employs the microscopy dataset which categorizes BC cells into 1) benign, 2) malignant, 3) in situ carcinoma, and 4) invasive carcinoma.
- Furthermore, the dataset is composed of 400 training images and 100 test images with equal number of images in each category.

Histopathological Image Datasets: UCSB

Histopathological Image Datasets: UCSB

- The third dataset is University of California, Santa Barbara (UCSB) benchmark dataset for bioimaging application (Gelasca et al., 2008).

Histopathological Image Datasets: UCSB

- The third dataset is University of California, Santa Barbara (UCSB) benchmark dataset for bioimaging application (Gelasca et al., 2008).
- The dataset comprises of 58 histopathological images which are used for BC histopathological image classification task with associated ground truth data available.

Research Methodology

Our research methodology consists of

Research Methodology

Our research methodology consists of

1. merging the dataset,

Research Methodology

Our research methodology consists of

1. merging the dataset,
2. preprocessing the dataset,

Research Methodology

Our research methodology consists of

1. merging the dataset,
2. preprocessing the dataset,
3. building the baseline, and constructing progressive resizing approach.

Merging the Dataset

- Three different datasets (Spanhol et al., 2016; Aresta et al., 2019; Gelasca et al., 2008) have different classes; therefore, in order to merge those different classes, those labels need adjusting.
- Firstly, the classes in BreaKHis dataset are benign and malignant. The benign class has subclasses: adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma, while the malignant class consists of adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma.
- Secondly, BACH dataset have four classes, such as normal, benign, in situ carcinoma, and invasive carcinoma.
- Lastly, UCSB benchmark dataset has benign and malignant classes.

Vahadane Color Normalization (1/3)

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.
Let $H \in \mathbb{R}^{r \times n}$ be the stain density maps, where the rows of which represent the concentration of each stain.

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.
Let $H \in \mathbb{R}^{r \times n}$ be the stain density maps, where the rows of which represent the concentration of each stain. Then, I can be written as follows (Gavrilovic et al., 2013; Vahadane et al., 2016):

(1)

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.
Let $H \in \mathbb{R}^{r \times n}$ be the stain density maps, where the rows of which represent the concentration of each stain. Then, I can be written as follows (Gavrilovic et al., 2013; Vahadane et al., 2016):

$$I = I_0 \exp\left(-WH\right) \tag{1}$$

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.
Let $H \in \mathbb{R}^{r \times n}$ be the stain density maps, where the rows of which represent the concentration of each stain. Then, I can be written as follows (Gavrilovic et al., 2013; Vahadane et al., 2016):

$$I = I_0 \exp\left(-WH\right) \tag{1}$$

Let V be the relative optical density then

(2)

Vahadane Color Normalization (1/3)

Let $I \in \mathbb{R}^{m \times n}$ be a matrix of RGB intensities, where $m = 3$ for RGB channels, and $n =$ number of pixels, and let I_0 be the illuminating light intensity on the sample (usually 255 for 8 bit images).
Let $W \in \mathbb{R}^{m \times r}$ be the stain color appearance matrix whose columns represent color basis of each stain such that r is the number of stains.
Let $H \in \mathbb{R}^{r \times n}$ be the stain density maps, where the rows of which represent the concentration of each stain. Then, I can be written as follows (Gavrilovic et al., 2013; Vahadane et al., 2016):

$$I = I_0 \exp(-WH) \tag{1}$$

Let V be the relative optical density then

$$V = \log \frac{I_0}{I_1}. \tag{2}$$

Vahadane Color Normalization (2/3)

(3)

The Goal

Vahadane Color Normalization (2/3)

Subsitute Equation (1) into Equation (2) results

(3)

The Goal

Vahadane Color Normalization (2/3)

Subsitute Equation (1) into Equation (2) results

$$V = WH \tag{3}$$

The Goal

Vahadane Color Normalization (2/3)

Subsitute Equation (1) into Equation (2) results

$$V = WH \tag{3}$$

The Goal

So, given an observation matrix V, the goal is to find stain color appearance matrix W and stain density map matrix H.

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

(4)

(5)

(6)

(7)

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

(5)

(6)

(7)

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

where λ is sparsity and regularization parameter.

(5)

(6)

(7)

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

where λ is sparsity and regularization parameter. Then,

(5)

(6)

(7)

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

where λ is sparsity and regularization parameter. Then,

$$H_s^{\text{norm}}(j,:) = \frac{H_s(j,:)}{H^{\text{RM}}(j,:)}, \; j = 1, \ldots, r \tag{5}$$

(6)

(7)

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

where λ is sparsity and regularization parameter. Then,

$$H_s^{\text{norm}}(j,:) = \frac{H_s(j,:)}{H^{\text{RM}}(j,:)}, \; j = 1, \ldots, r \tag{5}$$

$$V_s^{\text{norm}} = W_t H_s^{\text{norm}} \tag{6}$$

$$\tag{7}$$

Vahadane Color Normalization (3/3)

Given a source image s and a pathologist-preferred target image t, we estimate their color appearances and stain density maps by factorizing V_s into $W_s H_s$ and V_t into $W_t H_t$ using

$$\min_{W,H} \frac{1}{2}\|V - WH\|_F^2 + \lambda \sum_{j=1}^{r} \|H(j,:)\|_1, \quad W, H \geq 0, \|W(:,j)\|_2^2 = 1 \tag{4}$$

where λ is sparsity and regularization parameter. Then,

$$H_s^{\text{norm}}(j,:) = \frac{H_s(j,:)}{H^{\text{RM}}(j,:)}, \; j = 1, \ldots, r \tag{5}$$

$$V_s^{\text{norm}} = W_t H_s^{\text{norm}} \tag{6}$$

$$I_s^{\text{norm}} = I_0 \exp\left(-V_s^{\text{norm}}\right). \tag{7}$$

Building the Baseline

Building the Baseline

- ResNet-34 (He et al., 2016) is chosen to be the baseline for our experiment because ResNet architecture which relies on residual connections is the most widely used architecture and proven to be a strong baseline among Convolutional Neural Network (CNN) architectures;

Building the Baseline

- ResNet-34 (He et al., 2016) is chosen to be the baseline for our experiment because ResNet architecture which relies on residual connections is the most widely used architecture and proven to be a strong baseline among Convolutional Neural Network (CNN) architectures;
- Recent development in image classfication models is getting more and more on using the same trick of residual connections or tweaking the original ResNet architecture (Howard and Gugger, 2020).

Constructing Progressive Resizing Approach

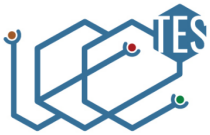
Constructing Progressive Resizing Approach

- Progressive resizing approach was one of the most important innovations when `fast.ai` and its team won the DAWNBench Competition in 2018 (Howard and Gugger, 2020).

Constructing Progressive Resizing Approach

- Progressive resizing approach was one of the most important innovations when `fast.ai` and its team won the DAWNBench Competition in 2018 (Howard and Gugger, 2020).
- The idea is very simple, that is to start training using small images and finally end the training using large images.

Constructing Progressive Resizing Approach



- Progressive resizing approach was one of the most important innovations when `fast.ai` and its team won the DAWNBench Competition in 2018 (Howard and Gugger, 2020).
- The idea is very simple, that is to start training using small images and finally end the training using large images.
- Training with small images for most of the epochs helps finishing the training much faster. Additionally, completing training with large images achieves a much higher final accuracy.

Constructing Progressive Resizing Approach

- Progressive resizing approach was one of the most important innovations when `fast.ai` and its team won the DAWNBench Competition in 2018 (Howard and Gugger, 2020).
- The idea is very simple, that is to start training using small images and finally end the training using large images.
- Training with small images for most of the epochs helps finishing the training much faster. Additionally, completing training with large images achieves a much higher final accuracy.
- Progressive resizing is also another strategy of *data augmentation*. Accordingly, better generalization of our models should be expected when they are trained with progressive resizing.

Results: The Unified Dataset

The unified dataset has three classes such as normal, benign, and malignant.

Table 1: Statistics of image sizes which consist of mean, standard deviation, minimum, Q_1, Q_2, Q_3, and maximum (measurement unit: pixel)

Statistics	**Width**	**Height**
Count	8,367.000	8,367.000
Mean	513.540	765.802
Standard deviation	230.537	287.778
Minimum	456.000	700.000
Q_1	460.000	700.000
Q_2 (median)	460.000	700.000
Q_3	460.000	700.000
Maximum	1,536.000	2,048.000

Results: The Unified Dataset

The unified dataset has three classes such as normal, benign, and malignant.

Table 1: Statistics of image sizes which consist of mean, standard deviation, minimum, Q_1, Q_2, Q_3, and maximum (measurement unit: pixel)

Statistics	**Width**	**Height**
Count	8,367.000	8,367.000
Mean	513.540	765.802
Standard deviation	230.537	287.778
Minimum	456.000	700.000
Q_1	460.000	700.000
Q_2 (median)	460.000	700.000
Q_3	460.000	700.000
Maximum	1,536.000	2,048.000

Furthermore, 70% of the dataset is chosen randomly to be training set and the rest is determined as validation set.

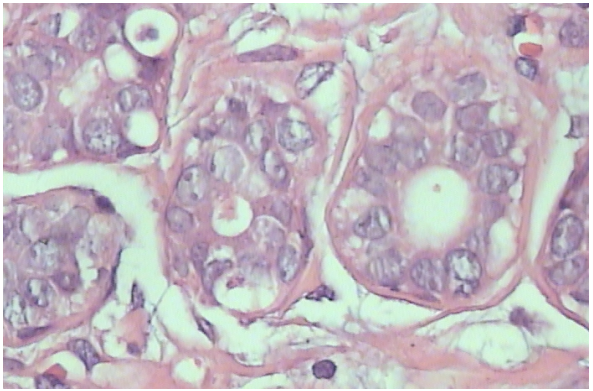
Results: Normalizing Colors of Images

We utilize `StainTools` library to do the normalization. The library can be found at `https://github.com/Peter554/StainTools`.

(a) (b)

Results: Normalizing Colors of Images

We utilize `StainTools` library to do the normalization. The library can be found at `https://github.com/Peter554/StainTools`.

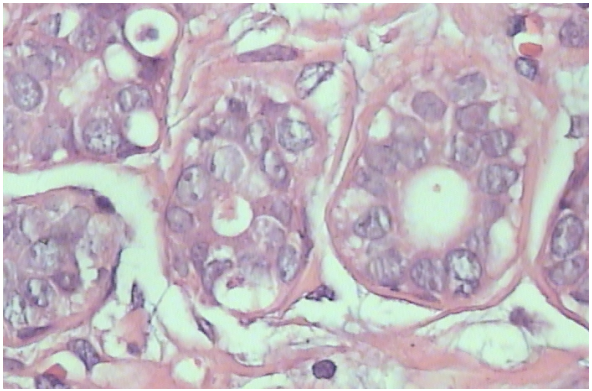


(a) Before normalization

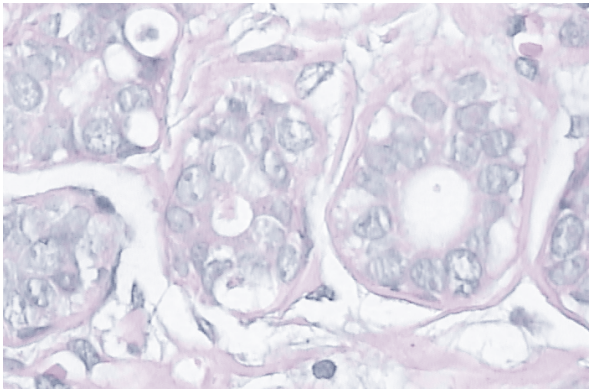
(b)

Results: Normalizing Colors of Images

We utilize `StainTools` library to do the normalization. The library can be found at `https://github.com/Peter554/StainTools`.



(a) Before normalization



(b) After normalization

Results: Performance of the First Baseline (1/2)

Results: Performance of the First Baseline (1/2)

- ResNet-34 model as a baseline is created by employing `fast.ai` library.

Results: Performance of the First Baseline (1/2)

- ResNet-34 model as a baseline is created by employing `fast.ai` library.
- Before the model is trained with the unified dataset, all images are resized into the mean of the width and height of the images that are 514 pixels by 766 pixels respectively.

Results: Performance of the First Baseline (1/2)

- ResNet-34 model as a baseline is created by employing `fast.ai` library.
- Before the model is trained with the unified dataset, all images are resized into the mean of the width and height of the images that are 514 pixels by 766 pixels respectively.
- In addition, the standard of one epoch is used to do a fine-tuning process on the pre-trained ResNet-34 (Howard and Gugger, 2020).

Results: Performance of the First Baseline (2/2)

Table 2: The F_1 scores of the the baseline by fine-tuning ResNet-34 pre-trained model (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.636	0.516	85.547%	19:06
Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.104	0.741	**83.733%**	05:13

Results: Performance of the First Baseline (2/2)

Table 2: The F_1 scores of the the baseline by fine-tuning ResNet-34 pre-trained model (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.636	0.516	85.547%	19:06
Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.104	0.741	**83.733%**	05:13

We opt to use F_1 score as our performance metric since the number of instances in each class of our dataset are imbalanced and F_1 is the best choice for measuring performance on imbalanced datasets (Sokolova and Lapalme, 2009).

Results: Performance of the Second Baseline

Results: Performance of the Second Baseline

- To assess the effect of progressive resizing, we construct another baseline (ResNet-50) based on a data augmentation technique, the so-called *presizing trick*.

Results: Performance of the Second Baseline

- To assess the effect of progressive resizing, we construct another baseline (ResNet-50) based on a data augmentation technique, the so-called *presizing trick*.
- Firstly, we resize all images to dimensions that are significantly larger than the target training dimensions.

Results: Performance of the Second Baseline

- To assess the effect of progressive resizing, we construct another baseline (ResNet-50) based on a data augmentation technique, the so-called *presizing trick*.
- Firstly, we resize all images to dimensions that are significantly larger than the target training dimensions.
- Next, we arrange all common augmentation operations including a resize to the final target size into one big chuck of operation, and finally performing the operation on the GPU only once at the end of trick.

Results: Performance of the Second Baseline

Table 3: The F_1 scores of the the second baseline (ResNet-50) by using presizing trick (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.313	0.312	91.196%	03:40
1	0.213	0.808	74.228%	03:39
2	0.160	0.089	97.547%	03:38
3	0.116	0.048	97.976%	03:38
4	0.079	**0.0315**	**98.443%**	03:38

Results: Performance of the Second Baseline

Table 3: The F_1 scores of the the second baseline (ResNet-50) by using presizing trick (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.313	0.312	91.196%	03:40
1	0.213	0.808	74.228%	03:39
2	0.160	0.089	97.547%	03:38
3	0.116	0.048	97.976%	03:38
4	0.079	**0.0315**	**98.443%**	03:38

We can still improve the performance of the model by using the progressive resizing.

Results: Performance of the Second Baseline

Table 3: The F_1 scores of the the second baseline (ResNet-50) by using presizing trick (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.313	0.312	91.196%	03:40
1	0.213	0.808	74.228%	03:39
2	0.160	0.089	97.547%	03:38
3	0.116	0.048	97.976%	03:38
4	0.079	**0.0315**	**98.443%**	03:38

We can still improve the performance of the model by using the progressive resizing. Firstly, we normalize our input data (Z-normalization) so it has a mean of 0 and a standard deviation of 1 and verify the effect of Z-normalization on training the model.

Results: Performance of the Third Baseline

Table 4: F_1 scores of the the third baseline (ResNet-50) by using presizing trick and Z-normalization (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	$\boldsymbol{F_1}$ **score**	**Time**
0	0.578	0.376	93.182%	03:40
1	0.239	0.249	93.90%	03:40
2	0.148	0.046	98.384%	03:38
3	0.092	0.038	98.626%	03:39
4	0.072	**0.037**	**98.682%**	03:38

Results: Performance of the Third Baseline

Table 4: F_1 scores of the the third baseline (ResNet-50) by using presizing trick and Z-normalization (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.578	0.376	93.182%	03:40
1	0.239	0.249	93.90%	03:40
2	0.148	0.046	98.384%	03:38
3	0.092	0.038	98.626%	03:39
4	0.072	**0.037**	**98.682%**	03:38

Table 4 shows utilizing Z-normalization improves F_1 score a little;

Results: Performance of the Third Baseline

Table 4: F_1 scores of the the third baseline (ResNet-50) by using presizing trick and Z-normalization (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the baseline is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.578	0.376	93.182%	03:40
1	0.239	0.249	93.90%	03:40
2	0.148	0.046	98.384%	03:38
3	0.092	0.038	98.626%	03:39
4	0.072	**0.037**	**98.682%**	03:38

Table 4 shows utilizing Z-normalization improves F_1 score a little; however, Z-normalization on input data becomes a standard when working with pretrained models.

Results: Performance of Progressive Resizing

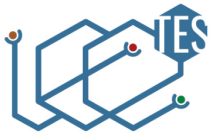
Results: Performance of Progressive Resizing

- Next, we employ the progressive resizing approach by starting a training with small images (128 pixels by 128 pixels) and ending the training using large images (the original image size).

Results: Performance of Progressive Resizing

- Next, we employ the progressive resizing approach by starting a training with small images (128 pixels by 128 pixels) and ending the training using large images (the original image size).
- This approach works because features learned by early layers of CNNs are not quite specific to the size of an image as the layers find curves and edges.

Results: Performance of Progressive Resizing



- Next, we employ the progressive resizing approach by starting a training with small images (128 pixels by 128 pixels) and ending the training using large images (the original image size).
- This approach works because features learned by early layers of CNNs are not quite specific to the size of an image as the layers find curves and edges.
- Moreover, the subsequent layers may later find shapes such as cell shapes. Therefore, changing image size in the middle of the training does not mean that the parameters of the models are completely different; it just requires the models to learn a little bit differently, that is by using transfer learning, in other words, *fine-tuning*.

Results: Training with Small Images

Table 5: F_1 scores of training on small-sized images (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the model is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.931	0.963	78.857%	03:21
1	0.397	0.109	95.593%	03:22
2	0.198	0.053	97.787%	03:21
3	0.115	0.042	98.426%	03:20

Results: Training with Small Images

Table 5: F_1 scores of training on small-sized images (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the model is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.931	0.963	78.857%	03:21
1	0.397	0.109	95.593%	03:22
2	0.198	0.053	97.787%	03:21
3	0.115	0.042	98.426%	03:20

Table 5 displays the process of training on small-sized images.

Results: Performance of Progressive Resizing

Table 6: F_1 scores of fine-tuning ResNet-50 as a part of progressive resizing approach (**Train** = train loss, **Valid** = validation loss; the higher the F_1 score is, the better the performance of the model is)

Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.109	0.049	97.117%	03:39
Epoch	**Train**	**Valid**	**F_1 score**	**Time**
0	0.081	0.044	98.501%	03:40
1	0.097	0.033	98.861%	03:39
2	0.076	0.025	98.981%	03:38
3	0.060	0.025	98.924%	03:39
4	0.050	0.022	**99.102%**	03:38

Results: Performance of Progressive Resizing

Results: Performance of Progressive Resizing

- To the best of our knowledge, the performance of our approach is among the highest BC classification model considering its nearly perfect F_1 score.

Results: Performance of Progressive Resizing

- To the best of our knowledge, the performance of our approach is among the highest BC classification model considering its nearly perfect F_1 score.
- Source codes of our approach is publicly available at `https://github.com/hbunyamin/2020-ice-tes-bc-dataset`.

Conclusion

Conclusion

- We have created a unified dataset merged from three popular datasets and propose the dataset for advancing research in BC classification field.

Conclusion

- We have created a unified dataset merged from three popular datasets and propose the dataset for advancing research in BC classification field.
- Moreover, in addition to the dataset contribution, we also provided a strong model using progressive resizing approach whose F_1 score is 99.102%. We argue that our model is comparable among other state-of-the-art models for the dataset.

References I

Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al. (2019). Bach: Grand challenge on breast cancer histology images. *Medical image analysis*.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424.

David S. Strayer, E. R. (2014). *Rubin's Pathology: Clinicopathologic Foundations of Medicine (Pathology (Rubin)) Seventh Edition*. LWW.

Gavrilovic, M., Azar, J. C., Lindblad, J., Wählby, C., Bengtsson, E., Busch, C., and Carlbom, I. B. (2013). Blind color decomposition of histological images. *IEEE Transactions on Medical Imaging*, 32(6):983–994.

Gelasca, E. D., Byun, J., Obara, B., and Manjunath, B. (2008). Evaluation and benchmark for biological image segmentation. In *2008 15th IEEE International Conference on Image Processing*, pages 1816–1819. IEEE.

References II

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems Second Edition*. O'Reilly Media Inc.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Howard, J. and Gugger, S. (2020). *Deep Learning for Coders with fastai and PyTorch*. O'Reilly Media Inc.

International Agency for Research on Cancer (2012). *WHO Classification of Tumours of the Breast [OP] (Medicine) 4th Edition*. World Health Organization.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., Ledsam, J. R., Melnick, D., Mostofi, H., Peng, L., Reicher, J. J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K. C., De Fauw, J., and Shetty, S. (2020). International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94.

References III

Millis, R. R. (1984). Needle biopsy of the breast. *Monographs in pathology*, (25):186–203.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I., and Navab, N. (2016). Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging*, 35(8):1962–1971.

World Health Organization (2018). Data global cancer observatory 2018. `https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf`. Accessed: 2020-01-04.

hendra.bunyamin@it.maranatha.edu