bangkit

conclusion

documentation

improvement

result

background

CNN Approach for Breast Cancer Diagnosis in Histopathological Dataset

Hansen Idden
Nadhifa Sofia
Reza Anugrah Prakasa

Highest Mortality

Breast cancer has the second highest mortality rate after Lung & Bronchial cancer, and about 30% of newly diagnosed cases are of breast cancer only.

Human Error

Manual detection is a tedious, tiring task and most likely to comprise human error, as most parts of the cell are frequently part of irregular random and arbitrary visual angles.

Histopathology

Images are acquired by histopathology, which generally includes biopsy of the affected tissue.

Benign/Malignant Classification

The goal is to identify whether a tumor is benign or of a malignant in nature, as malignant tumors are cancerous and should be treated as soon as possible to reduce and prevent further complications.

conclusion

documentation

improvement

result

What is the potential of computational pathology?

background

conclusion

documentation

improvement

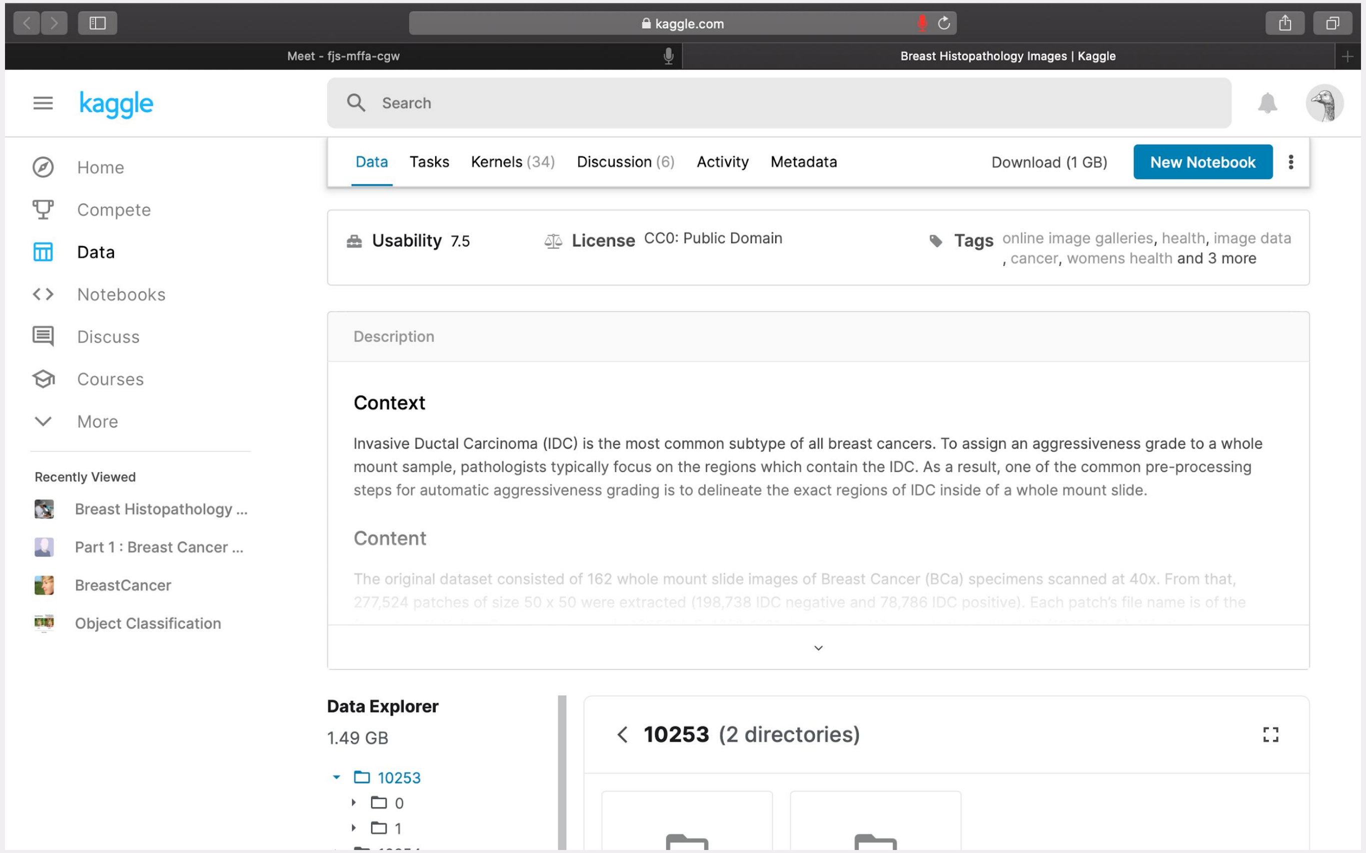
result

background

What is the potential of computational pathology?

Why did we choose this dataset?

The original dataset consisted of 162 whole mount slide images of Breast Cancer (BCa) specimens scanned at 40x. From that, **277,524 patches of size 50 x 50** were extracted (198,738 IDC negative and 78,786 IDC positive). Each patch's file name is of the format: uxXyYclassC.png — > example 10253idx5x1351y1101class0.png . Where u is the patient ID (10253idx5), X is the x-coordinate of where this patch was cropped from, Y is the y-coordinate of where this patch was cropped from, and C indicates the class where 0 is non-IDC and 1 is IDC.



conclusion

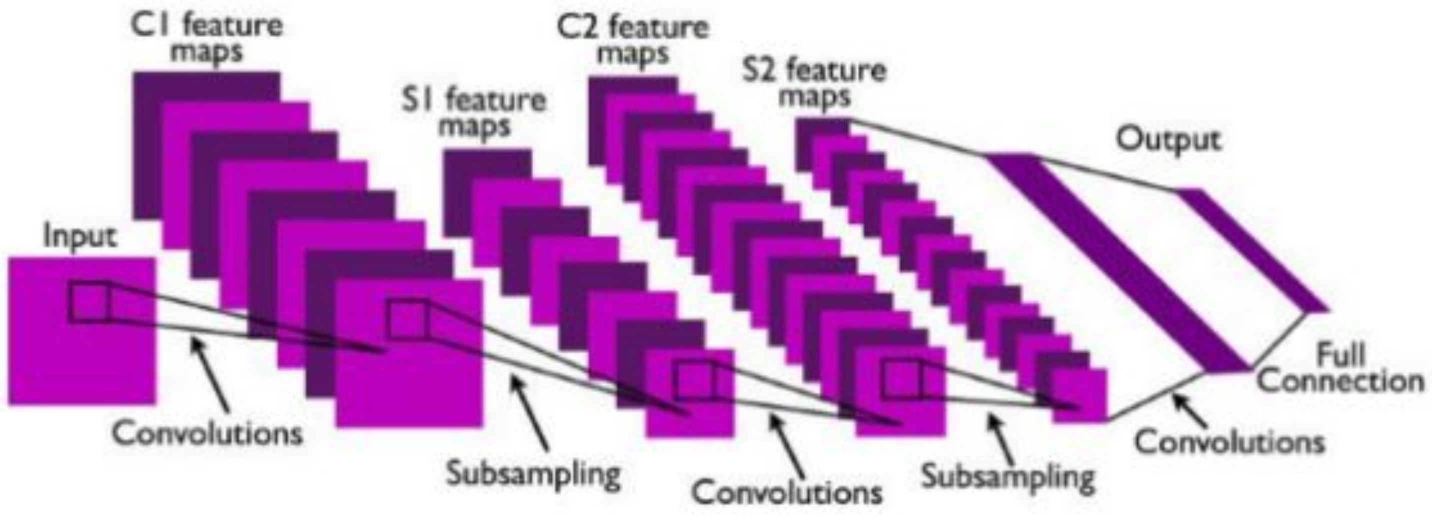
documentation

improvement

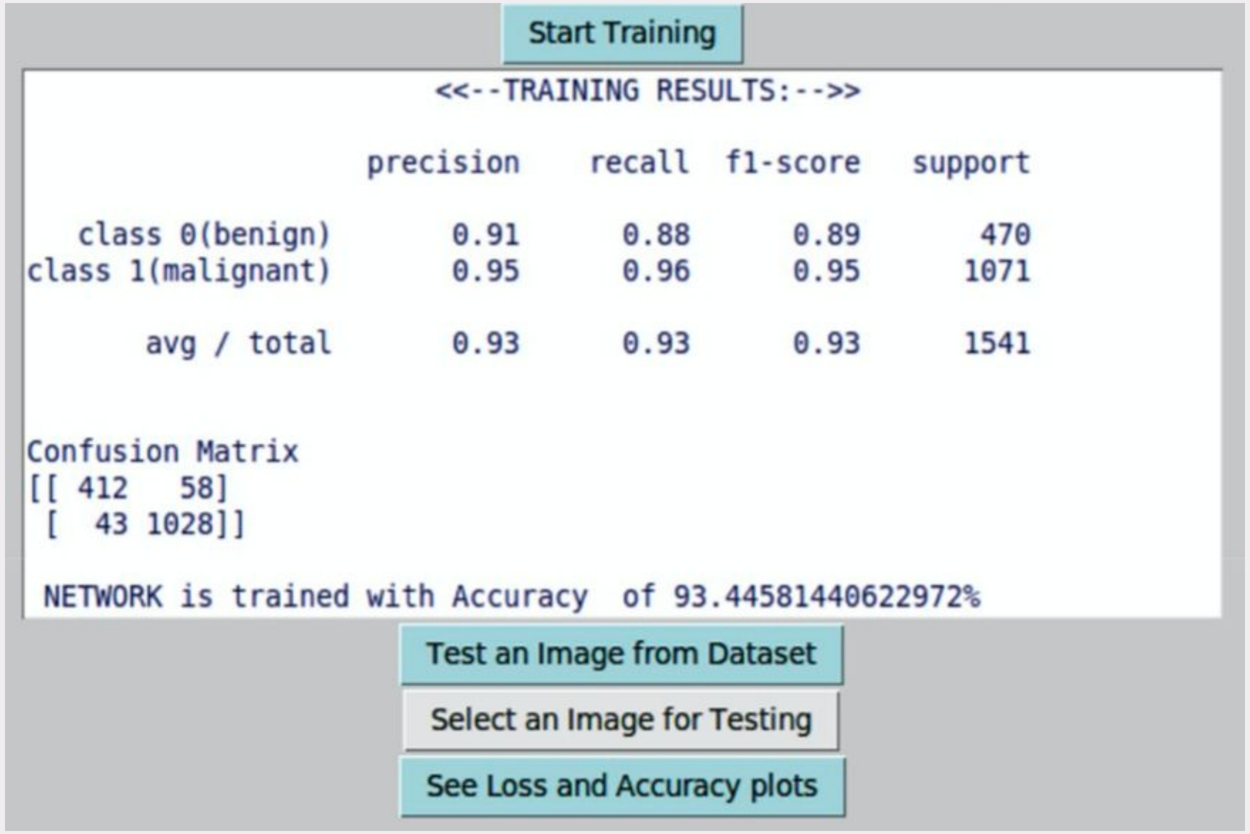
result

background

CNN Implementation Baseline



- **Convolution Layer**
- **Pooling Layer (Subsampling)**
- **Full Connection Layer**


Start Training
<<--TRAINING RESULTS:-->>
precision recall f1-score support
class 0(benign) 0.91 0.88 0.89 470
class 1(malignant) 0.95 0.96 0.95 1071
avg / total 0.93 0.93 0.93 1541
Confusion Matrix
[[412 58]
[43 1028]]
NETWORK is trained with Accuracy of 93.44581440622972%
Test an Image from Dataset
Select an Image for Testing
See Loss and Accuracy plots

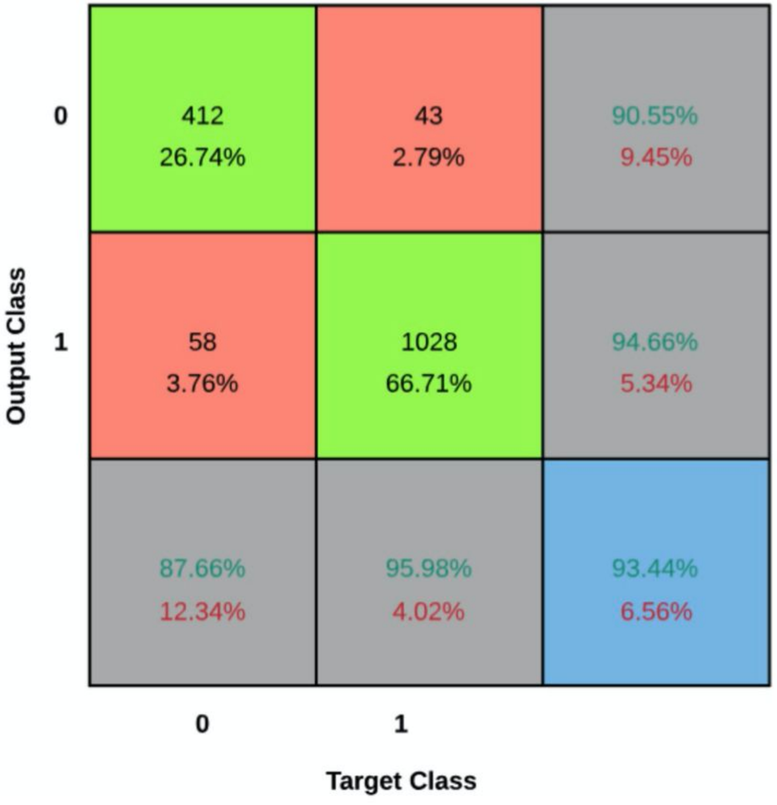


Fig. 5. Confusion matrix.

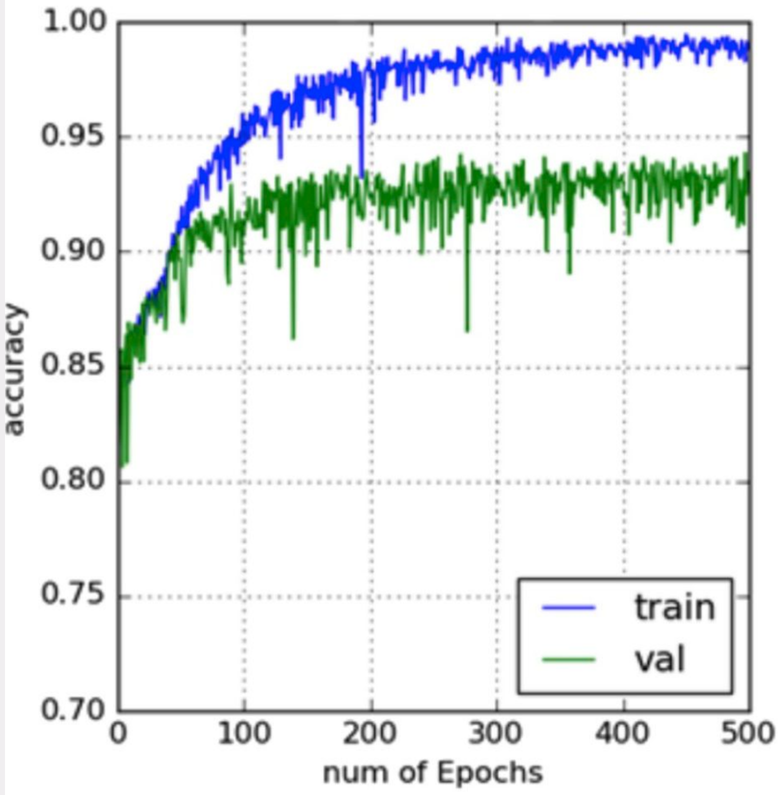
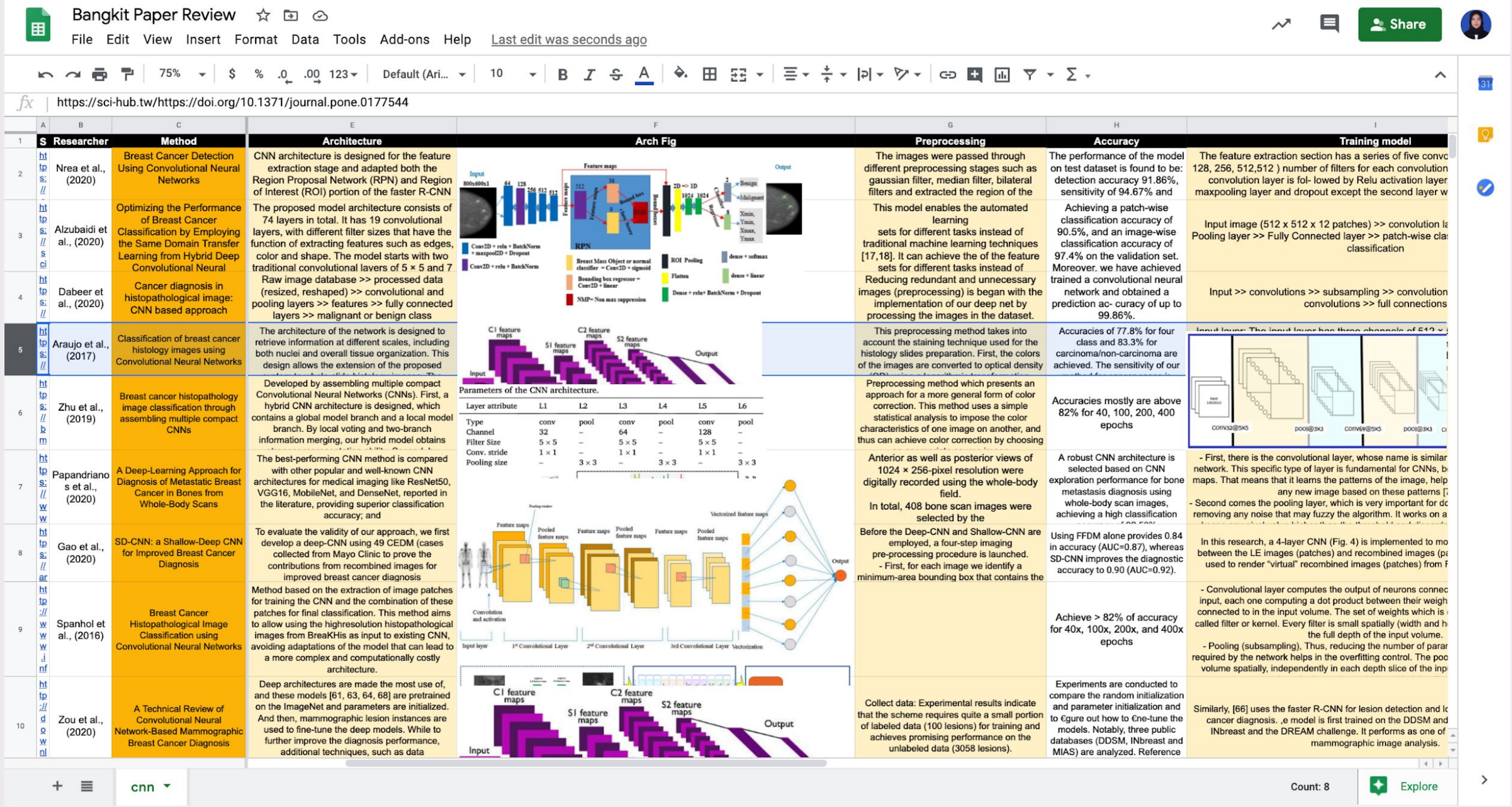


Fig. 9. Training accuracy and validation accuracy.

conclusion

documentation

improvement



Bangkit Paper Review

https://sci-hub.tw/https://doi.org/10.1371/journal.pone.0177544

S	Researcher	Method	Architecture	Arch Fig	Preprocessing	Accuracy	Training model
https://	Nrea et al., (2020)	Breast Cancer Detection Using Convolutional Neural Networks	CNN architecture is designed for the feature extraction stage and adapted both the Region Proposal Network (RPN) and Region of Interest (ROI) portion of the faster R-CNN		The images were passed through different preprocessing stages such as gaussian filter, median filter, bilateral filters and extracted the region of the	The performance of the model on test dataset is found to be: detection accuracy 91.86%, sensitivity of 94.67% and	The feature extraction section has a series of five convo 128, 256, 512,512) number of filters for each convolution convolution layer is fol- lowed by Relu activation layer maxpooling layer and dropout except the second layer w
https://sci	Alzubaidi et al., (2020)	Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural	The proposed model architecture consists of 74 layers in total. It has 19 convolutional layers, with different filter sizes that have the function of extracting features such as edges, color and shape. The model starts with two traditional convolutional layers of 5 × 5 and 7		This model enables the automated learning sets for different tasks instead of traditional machine learning techniques [17,18]. It can achieve the of the feature sets for different tasks instead of	Achieving a patch-wise classification accuracy of 90.5%, and an image-wise classification accuracy of 97.4% on the validation set. Moreover, we have achieved	Input image (512 x 512 x 12 patches) >> convolution la Pooling layer >> Fully Connected layer >> patch-wise cla classification
https://	Dabeer et al., (2020)	Cancer diagnosis in histopathological image: CNN based approach	Raw image database >> processed data (resized, reshaped) >> convolutional and pooling layers >> features >> fully connected layers >> malignant or benign class		Reducing redundant and unnecessary images (preprocessing) is began with the implementation of our deep net by processing the images in the dataset.	trained a convolutional neural network and obtained a prediction ac- curacy of up to 99.86%.	Input >> convolutions >> subsampling >> convolution convolutions >> full connections
https://	Araujo et al., (2017)	Classification of breast cancer histology images using Convolutional Neural Networks	The architecture of the network is designed to retrieve information at different scales, including both nuclei and overall tissue organization. This design allows the extension of the proposed		This preprocessing method takes into account the staining technique used for the histology slides preparation. First, the colors of the images are converted to optical density	Accuracies of 77.8% for four class and 83.3% for carcinoma/non-carcinoma are achieved. The sensitivity of our	conv32@5X5 pool@3X3 conv64@5X5 pool@3X3
https://bm	Zhu et al., (2019)	Breast cancer histopathology image classification through assembling multiple compact CNNs	Developed by assembling multiple compact Convolutional Neural Networks (CNNs). First, a hybrid CNN architecture is designed, which contains a global model branch and a local model branch. By local voting and two-branch information merging, our hybrid model obtains		Preprocessing method which presents an approach for a more general form of color correction. This method uses a simple statistical analysis to impose the color characteristics of one image on another, and thus can achieve color correction by choosing	Accuracies mostly are above 82% for 40, 100, 200, 400 epochs	
https://ww	Papandrianos et al., (2020)	A Deep-Learning Approach for Diagnosis of Metastatic Breast Cancer in Bones from Whole-Body Scans	The best-performing CNN method is compared with other popular and well-known CNN architectures for medical imaging like ResNet50, VGG16, MobileNet, and DenseNet, reported in the literature, providing superior classification accuracy; and		Anterior as well as posterior views of 1024 × 256-pixel resolution were digitally recorded using the whole-body field. In total, 408 bone scan images were selected by the	A robust CNN architecture is selected based on CNN exploration performance for bone metastasis diagnosis using whole-body scan images, achieving a high classification	- First, there is the convolutional layer, whose name is similar network. This specific type of layer is fundamental for CNNs, b maps. That means that it learns the patterns of the image, help any new image based on these patterns - Second comes the pooling layer, which is very important for d removing any noise that may fuzzy the algorithm. It works on a
https://ar	Gao et al., (2020)	SD-CNN: a Shallow-Deep CNN for Improved Breast Cancer Diagnosis	To evaluate the validity of our approach, we first develop a deep-CNN using 49 CEDM (cases collected from Mayo Clinic to prove the contributions from recombined images for improved breast cancer diagnosis		Before the Deep-CNN and Shallow-CNN are employed, a four-step imaging pre-processing procedure is launched. - First, for each image we identify a minimum-area bounding box that contains the	Using FFDM alone provides 0.84 in accuracy (AUC=0.87), whereas SD-CNN improves the diagnostic accuracy to 0.90 (AUC=0.92).	In this research, a 4-layer CNN (Fig. 4) is implemented to mo between the LE images (patches) and recombined images (p used to render "virtual" recombined images (patches) from F
http://www.inf	Spanhol et al., (2016)	Breast Cancer Histopathological Image Classification using Convolutional Neural Networks	Method based on the extraction of image patches for training the CNN and the combination of these patches for final classification. This method aims to allow using the highresolution histopathological images from BreaKHis as input to existing CNN, avoiding adaptations of the model that can lead to a more complex and computationally costly architecture.			Achieve > 82% of accuracy for 40x, 100x, 200x, and 400x epochs	- Convolutional layer computes the output of neurons connec input, each one computing a dot product between their weigh connected to in the input volume. The set of weights which is called filter or kernel. Every filter is small spatially (width and h the full depth of the input volume. - Pooling (subsampling), Thus, reducing the number of para required by the network helps in the overfitting control. The poo volume spatially, independently in each depth slice of the inp
http://downl	Zou et al., (2020)	A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis	Deep architectures are made the most use of, and these models [61, 63, 64, 68] are pretrained on the ImageNet and parameters are initialized. And then, mammographic lesion instances are used to fine-tune the deep models. While to further improve the diagnosis performance, additional techniques, such as data		Collect data: Experimental results indicate that the scheme requires quite a small portion of labeled data (100 lesions) for training and achieves promising performance on the unlabeled data (3058 lesions).	Experiments are conducted to compare the random initialization and parameter initialization and to €gure out how to €ne-tune the models. Notably, three public databases (DDSM, INbreast and MIAS) are analyzed. Reference	Similarly, [66] uses the faster R-CNN for lesion detection and l cancer diagnosis. ,e model is first trained on the DDSM and INbreast and the DREAM challenge. It performs as one of mammographic image analysis.

cnn

Count: 8

https://bit.ly/bangkit-paper-review

results

background

Alzubaidi et al., (2020)
Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model

Araujo et al., (2020)
Classification of breast cancer histology images using Convolutional Neural Networks

Dabeer et al., (2020)
Cancer diagnosis in histopathological image: CNN based approach

Gao et al., (2020)
SD-CNN: a Shallow-Deep CNN for Improved Breast Cancer Diagnosis

Papandrianos et al., (2020)
A Deep-Learning Approach for Diagnosis of Metastatic Breast Cancer in Bones from Whole-Body Scans

Zou et al., (2020)
A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis

Zhu et al., (2019)
Breast cancer histopathology image classification through assembling multiple compact CNNs

Spanhol et al., (2016)
Classification of breast cancer histology images using Convolutional Neural Networks

Nrea et al., (2020)
Breast Cancer Detection Using Convolutional Neural Networks

https://bit.ly/bangkit-paper-review

Adjusting learning rate

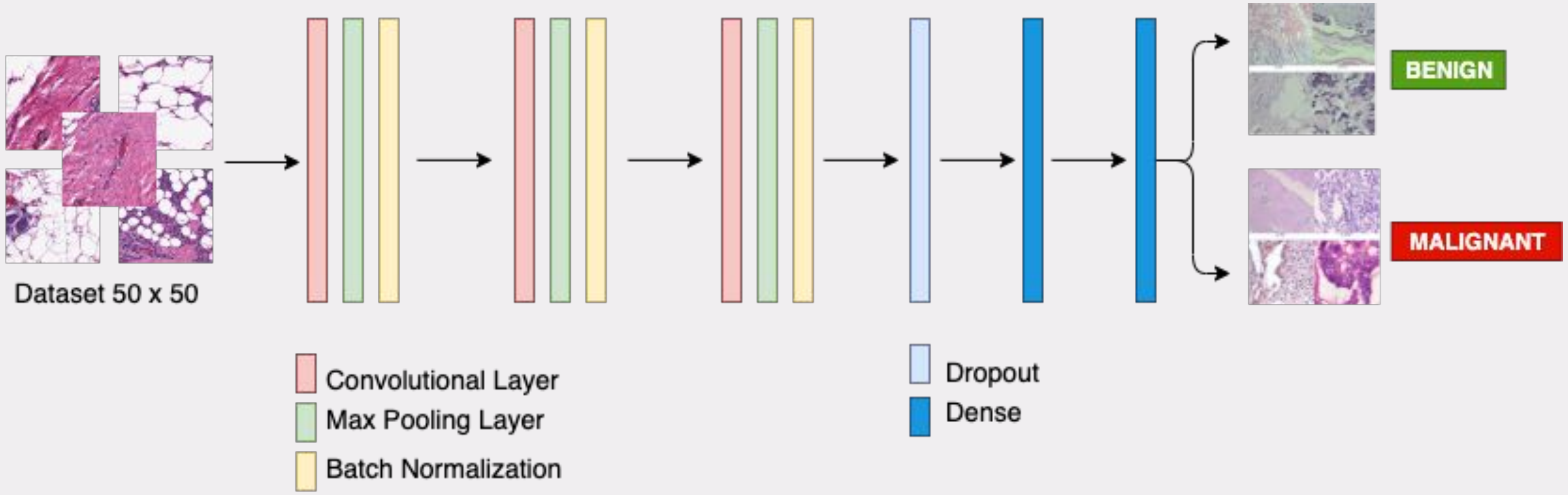
In order to enhance accuracy; 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 1×10^{-2}, 1×10^{-1}

Change batch

By using Stochastic Gradient Descent

Adjusting epochs

Reduce epochs from 60 to 25



conclusion

documentation

improvement

result

background

bangk!t

We propose these neural networks;

Jupyter Breast Cancer Attempt Last Checkpoint: Last Wednesday at 10:26 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

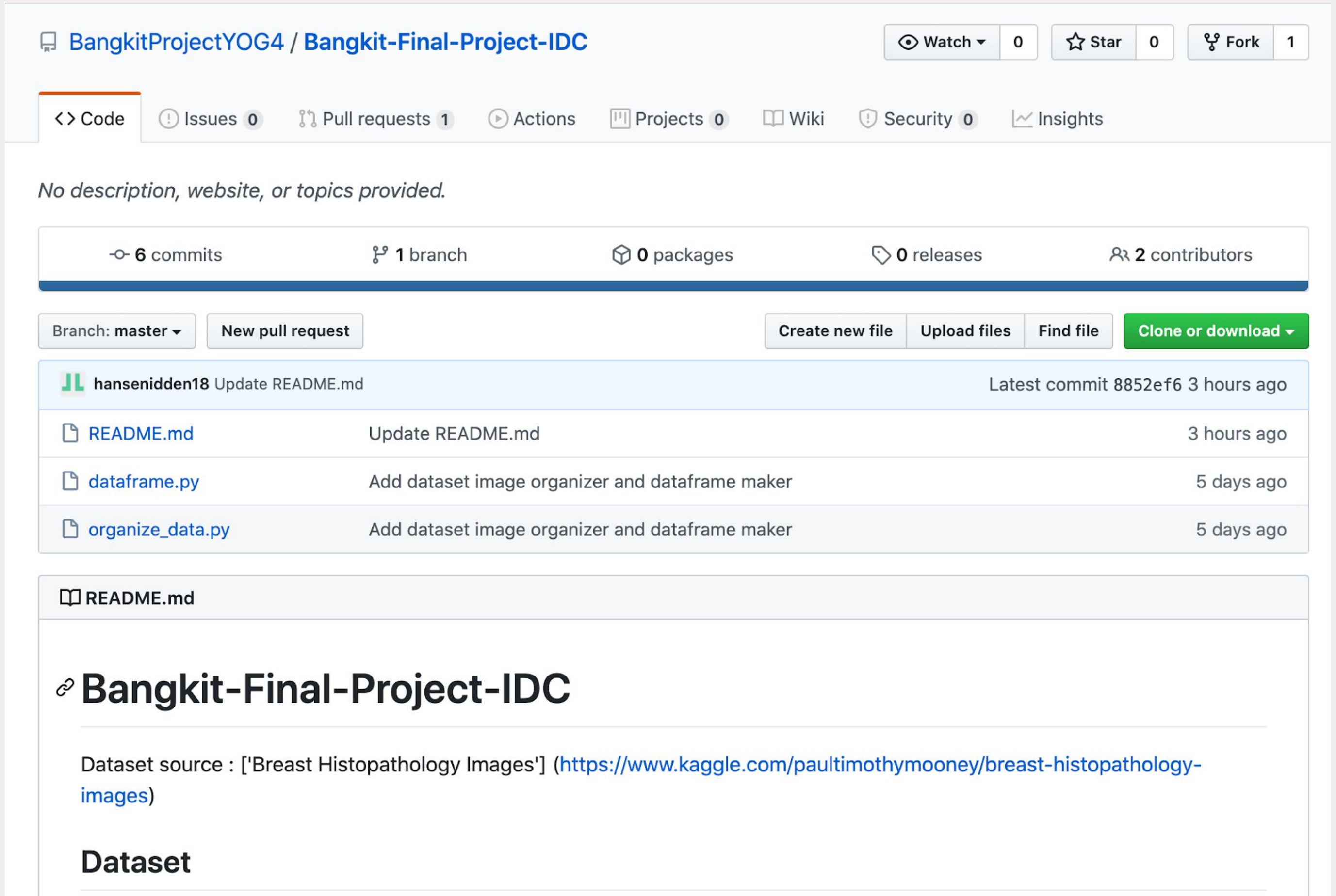
```
model.summary()

Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 50, 50, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 25, 25, 32)        0
_________________________________________________________________
batch_normalization_v2 (Batc (None, 25, 25, 32)        128
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 25, 25, 128)       36992
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 12, 12, 128)       0
_________________________________________________________________
batch_normalization_v2_1 (Ba (None, 12, 12, 128)       512
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 12, 12, 128)       147584
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 4608)              0
_________________________________________________________________
dropout (Dropout)            (None, 4608)              0
_________________________________________________________________
dense (Dense)                (None, 128)               589952
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 129
=================================================================
Total params: 776,193
Trainable params: 775,873
Non-trainable params: 320
_________________________________________________________________

In [22]: model.compile(tf.keras.optimizers.SGD(1e-3),loss = 'binary_crossentropy', metrics = ['acc'])
```

bangk!t

https://github.com/BangkitProjectYOG4/Bangkit-Final-Project-IDC



BangkitProjectYOG4 / **Bangkit-Final-Project-IDC**

Watch 0 | Star 0 | Fork 1

Code | Issues 0 | Pull requests 1 | Actions | Projects 0 | Wiki | Security 0 | Insights

No description, website, or topics provided.

6 commits | **1** branch | **0** packages | **0** releases | **2** contributors

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

hansenidden18 Update README.md — Latest commit 8852ef6 3 hours ago

README.md	Update README.md	3 hours ago
dataframe.py	Add dataset image organizer and dataframe maker	5 days ago
organize_data.py	Add dataset image organizer and dataframe maker	5 days ago

README.md

Bangkit-Final-Project-IDC

Dataset source : ['Breast Histopathology Images'] (https://www.kaggle.com/paultimothymooney/breast-histopathology-images)

Dataset

documentation

improvement

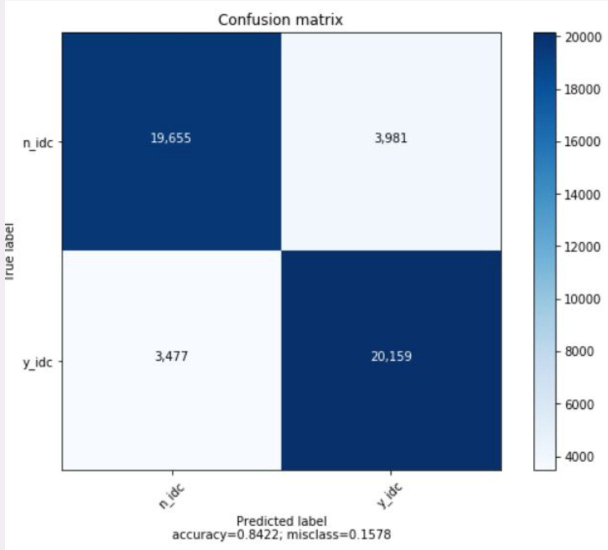
result

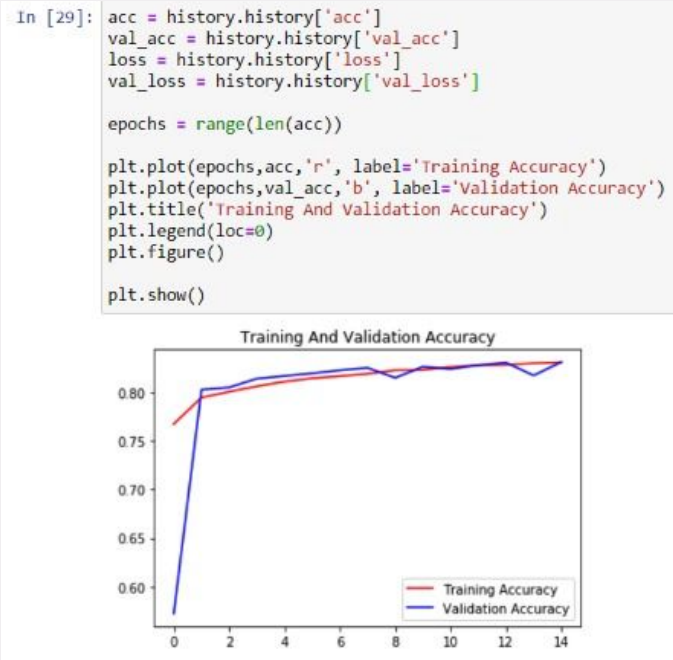
background

From this project, we could achieve as 84% of validation accuracy. It can be improved by using more HD patches as the inputs, but it also has drawback as it is computationally expensive.

Breast cancer detection by using digital/digitized histopathology images is a milestone in the field of medical pathology. It has also opened a door to new opportunities for research as there are many undiscovered areas that can be revealed by techniques and tools of machine learning and deep learning. We may obtain improved results by altering the network design and parameters. As an improvement to the proposed method, one can implement an autoencoder instead of manually reducing image size. It can compress data without losing the prominent features, because autoencoders can re-generate up to 90% of the original image. From the point of method improvement, we can incorporate spectral imaging.

Our confusion matrix :





```
In [29]: acc = history.history['acc']
         val_acc = history.history['val_acc']
         loss = history.history['loss']
         val_loss = history.history['val_loss']

         epochs = range(len(acc))

         plt.plot(epochs,acc,'r', label='Training Accuracy')
         plt.plot(epochs,val_acc,'b', label='Validation Accuracy')
         plt.title('Training And Validation Accuracy')
         plt.legend(loc=0)
         plt.figure()

         plt.show()
```