

Classification of Histopathological Breast cancer images using Deep Learning

Rishab Kumar

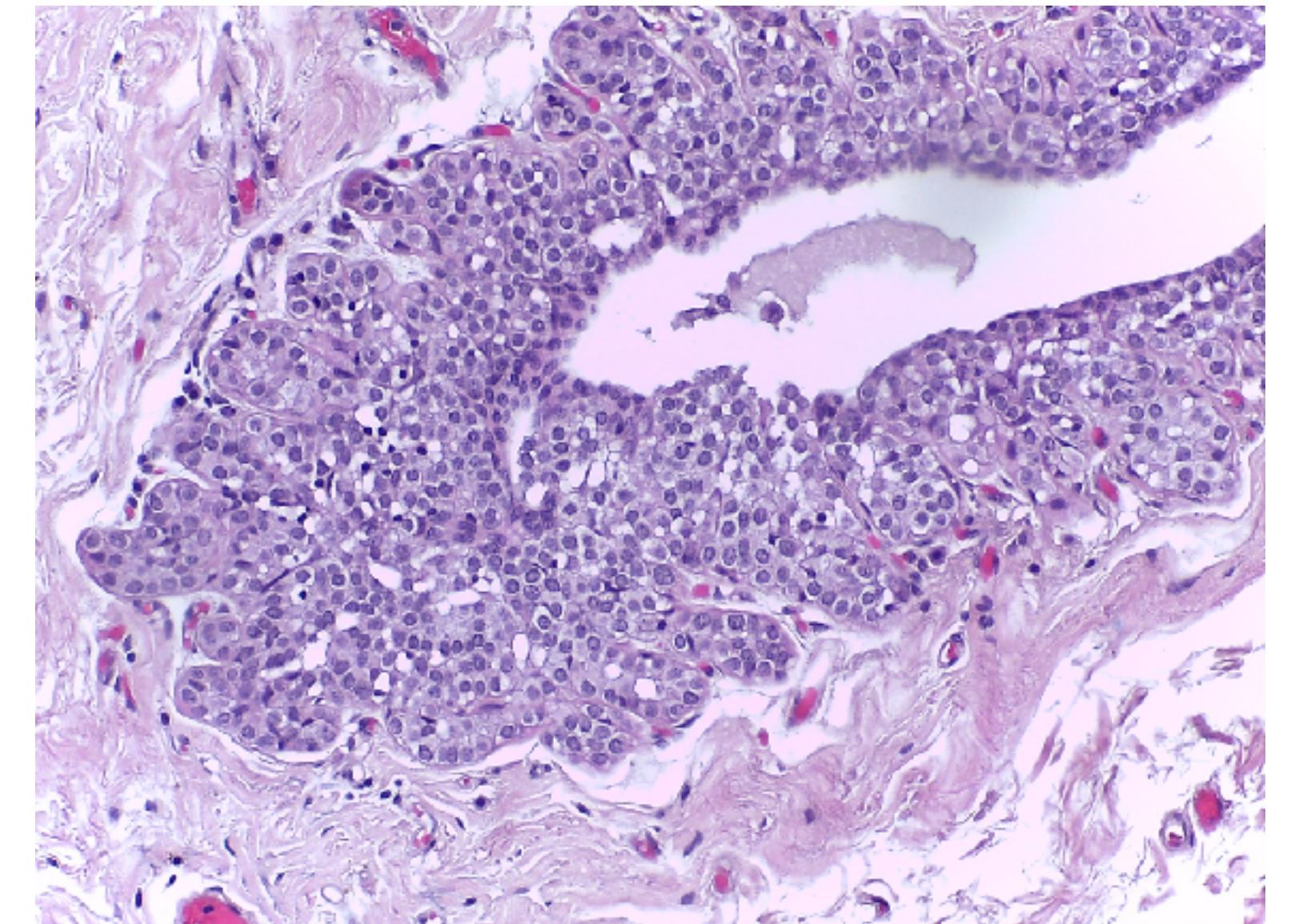
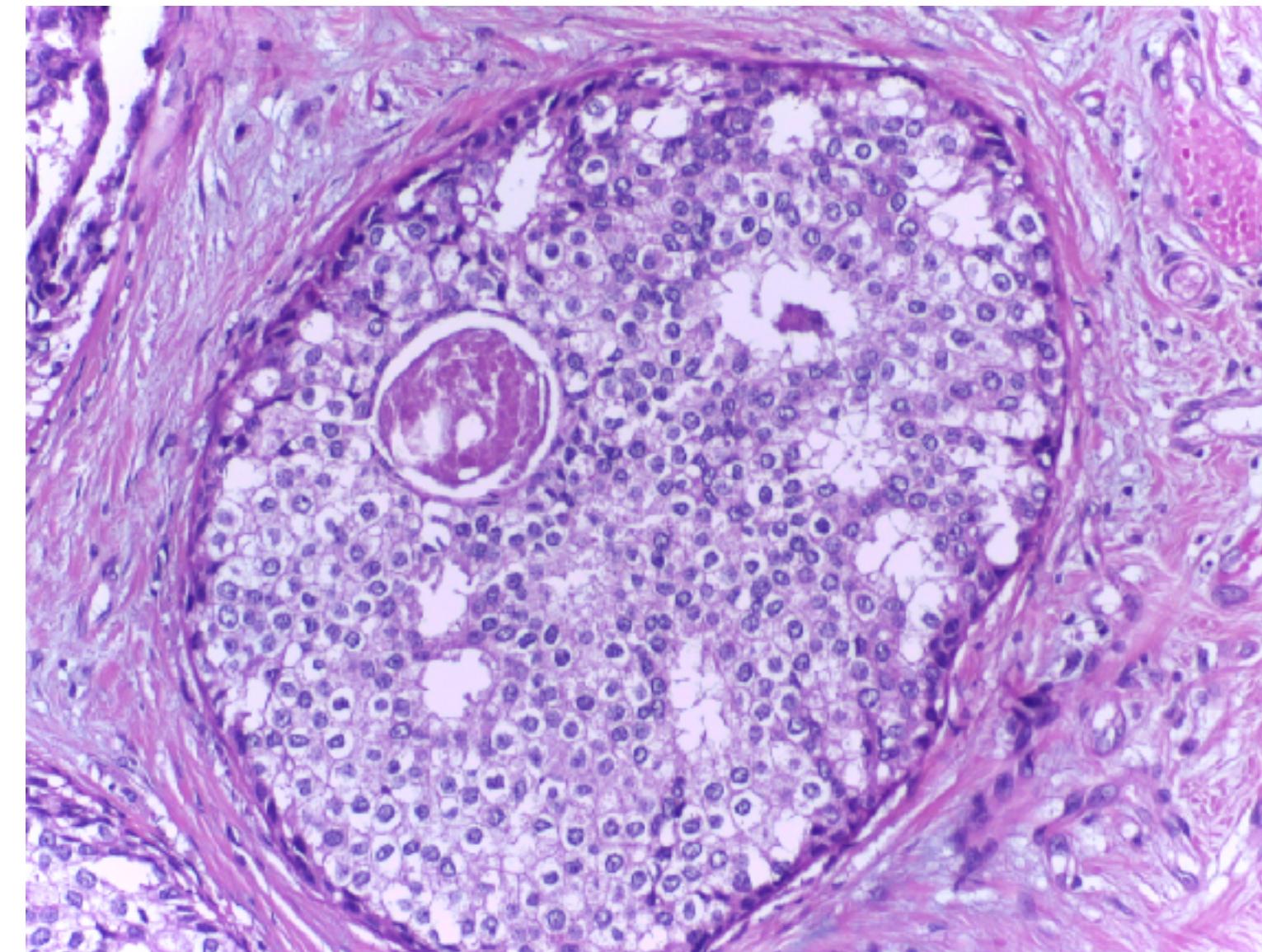
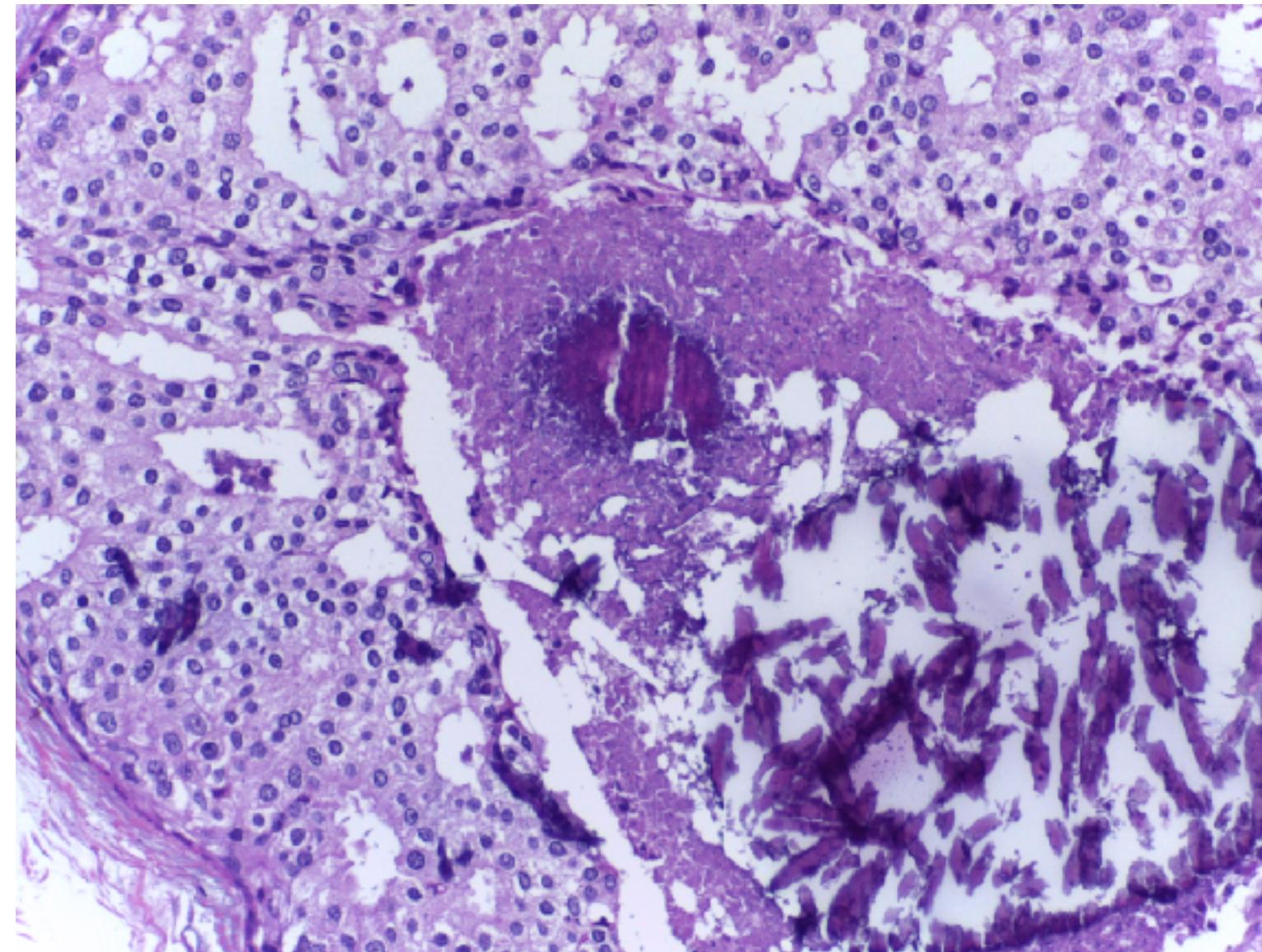
1701EE39

Under the guidance of Dr. M H Kolekar

INTRODUCTION

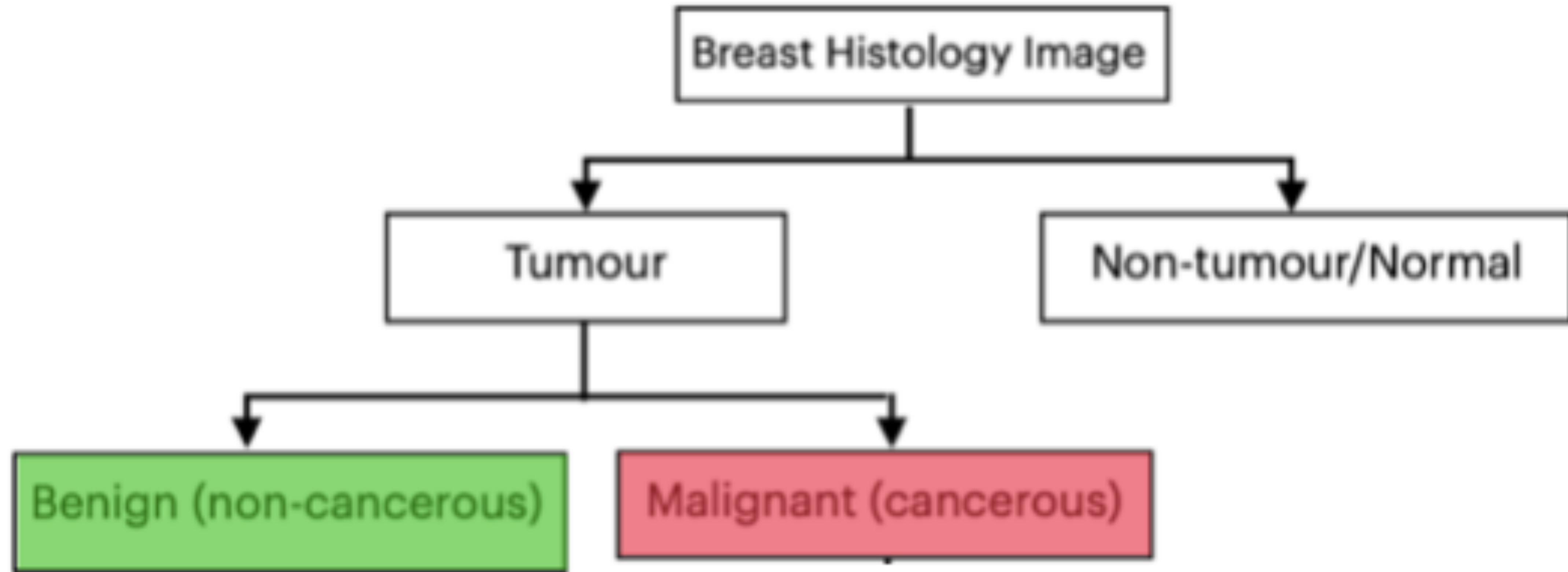
Certain types of breast cancer images

- Histopathological images
- Mammogram images
- Extracted feature based data
- Ultrasounds
- MRI images



Histopathological images from ICIAR2018 dataset

CLASSIFICATION OF BREAST HISTOLOGY IMAGES



DATASETS

Three standard datasets were chosen for training and testing purposes

1. Breakhis Dataset - 7,909 images
2. ICIAR2018 Dataset - 400 images
3. Kaggle Breast Histology Images Dataset - 277,524 images

Breast Cancer Histopathological Database (BreakHis)

The Breast Cancer Histopathological image Classification (BreakHis) is composed of 9,109 microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). To date, it contains 2,430 benign and 5,429 malignant samples (700x460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format). This database has been built in collaboration with the P&D Laboratory – Pathological Anatomy and Cytopathology, Paraná, Brazil (<http://www.prevencaodediagnose.com.br>). We believe that researchers will find this database a useful tool since it makes future benchmarking and evaluation possible.

Characteristics

The dataset BreakHis is divided into two main groups: benign tumors and malignant tumors. Histologically benign is a term referring to a lesion that does not match any criteria of malignancy – e.g., marked cellular atypia, mitosis, disruption of basement membranes, metastasize, etc. Normally, benign tumors are relatively “innocent”, presents slow growing and remains localized. Malignant tumor is a synonym for cancer: lesion can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastasize) to cause death.

In current version, samples present in dataset were collected by SB method, also named partial mastectomy or excisional biopsy. This type of procedure, compared to any methods of needle biopsy, removes the larger size of tissue sample and is done in a hospital with general anesthetic.



ICIAIR 2018 Grand Challenge on **Breast Cancer Histology** images

BACH's dataset is publicly available under the CC BY-NC-ND license. The citation should refer to [10.1101/medRxiv.2019.05.010](https://doi.org/10.1101/medRxiv.2019.05.010). Teams must notify the organisers of the challenge about any publication that is (partly) based on the results or data published on this site, in order to the challenge organization being able to maintain a list of publications associated with the challenge.

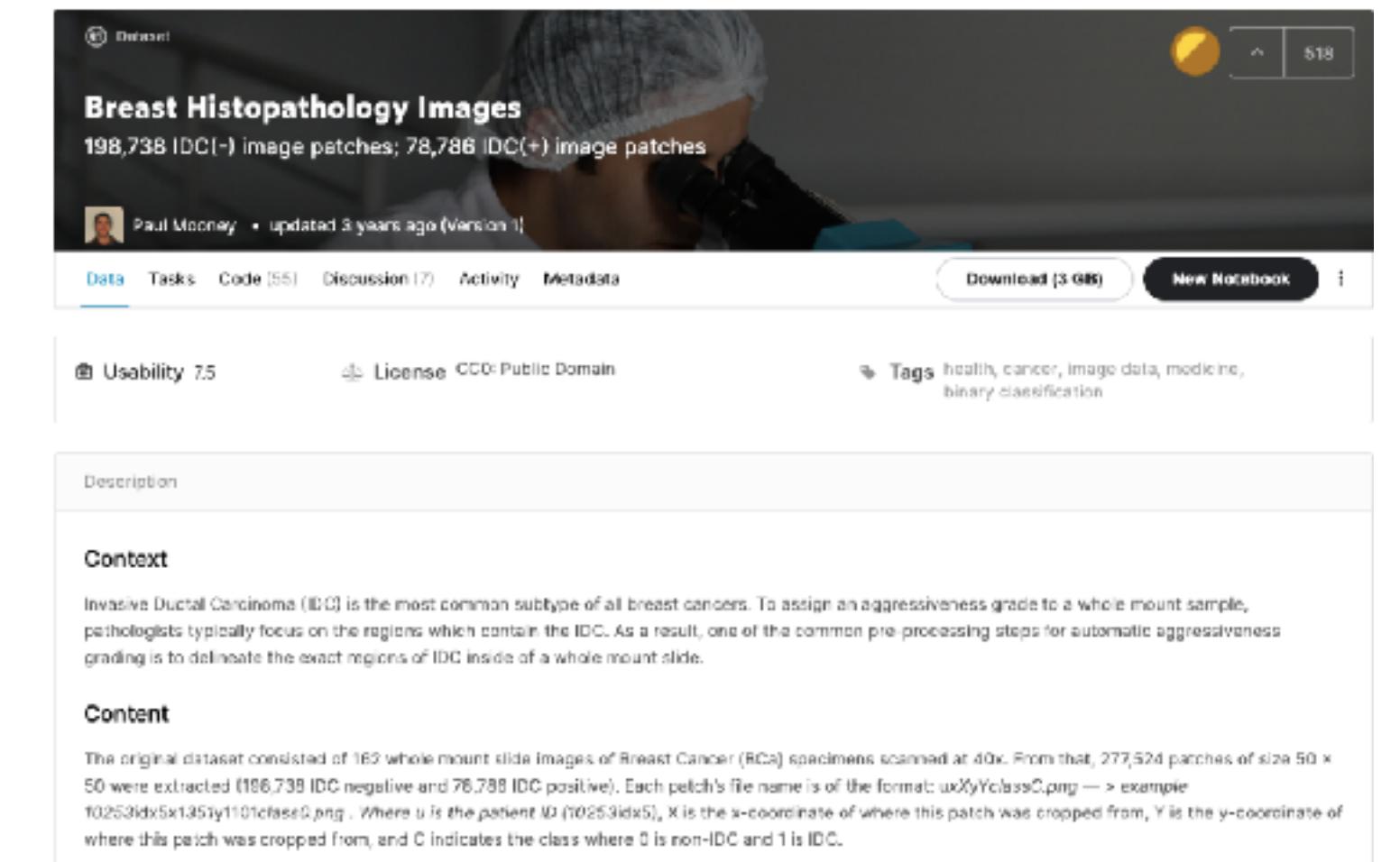
Dataset

The dataset is composed of hematoxylin and eosin (H&E) stained breast histology microscopy and whole-slide images. Challenge participants should evaluate the performance of their method on either/both sets of images.

[CLICK HERE TO REQUEST THE DATASET](#)

1. Microscopy images

Microscopy images are labelled as normal, benign, in situ carcinoma or invasive carcinoma according to the predominant cancer type in each image. The annotation was performed by two medical experts and images where there was disagreement were discarded.



A screenshot of a Kaggle dataset page titled "Breast Histopathology Images". The page shows a thumbnail of a pathologist working on a slide. Below the thumbnail, the dataset title is displayed along with the number of IDC(-) and IDC(+) patches. The page includes a user profile for Paul Mooney, update information, and navigation links for Data, Tasks, Code (SS), Discussion (7), Activity, and Metadata. A detailed description of the dataset is provided, mentioning its availability under a CC0 Public Domain license and its use in the ICIAR 2018 challenge. The page also lists Usability (7.5), License (CC0: Public Domain), and Tags (health, cancer, image data, medicine, binary classification).

Breakhis

ICIAIR2018

Kaggle BHI

METHODOLOGY

Trained 5 models for the dataset and obtained following accuracies

DenseNet201: 95%

ResNet152: 94.5%

VGG19: 93.5%

Xception: 93%

Inception_v3: 94%

METHODOLOGY

Combining the features extracted from 5 best performing classification architectures

Chance that an important feature will be missed is very less if we are using 5 independent classification architectures in concatenation

DenseNet201: consists of 201 layers

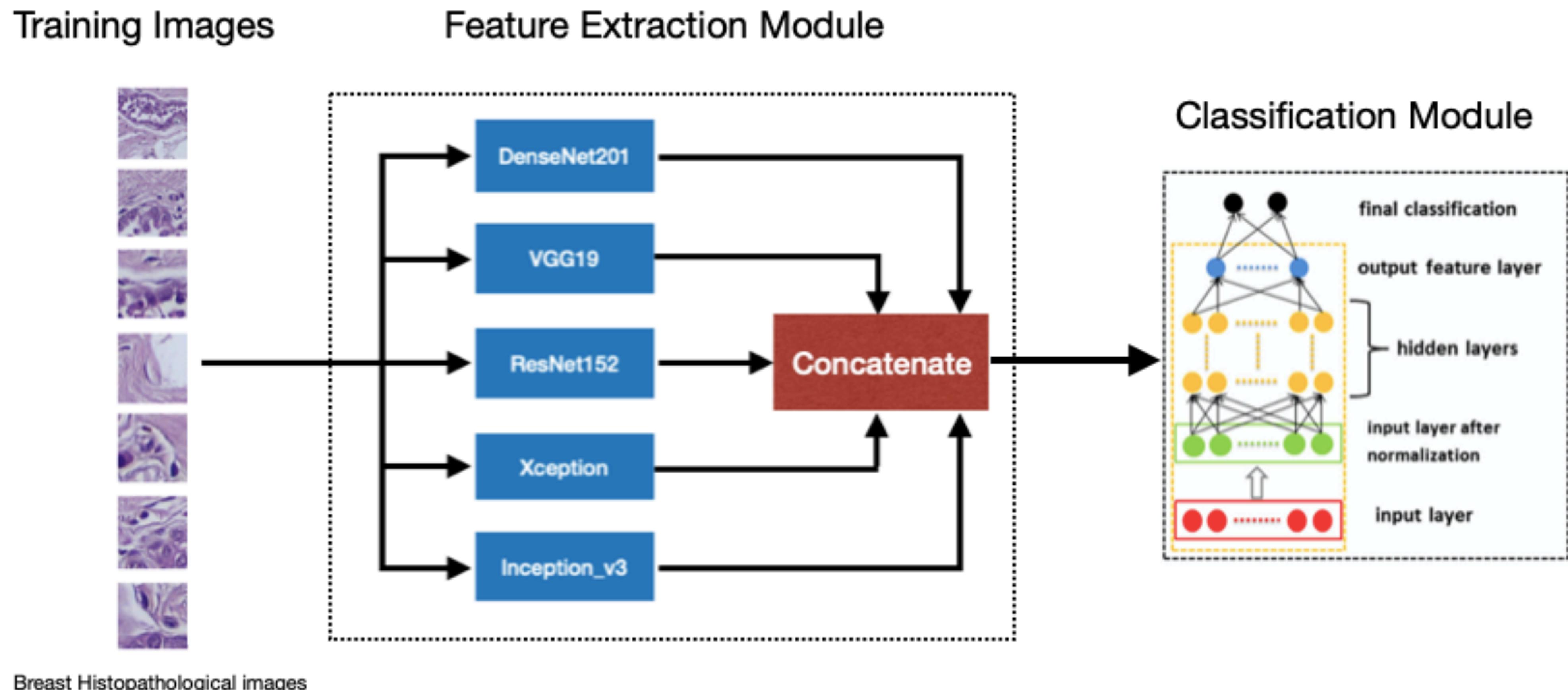
ResNet152: consists of 152 layers

VGG19: consists of 19 layers

Xception: consists of 71 layers

Inception_v3: consists of 48 layers

MODEL FLOWCHART



TRAINING

80% training
20% testing

Batch size : 16
Epochs: 100

Break his	Dataset Name	TRAIN		TEST	
		Benign	Malignant	Benign	Malignant
	40x	501	1097	126	275
	100x	517	1151	129	288
	200x	500	1113	125	279
	400x	472	987	118	247
	ICCIAR 2018	652	712	148	88
	KAGGLE BHI	1149	791	280	181

PERFORMANCE MEASURES

Positive is taken as MALIGNANT and negative is taken as BENIGN

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$Sensitivity = \frac{TP}{(TP + FN)}$$

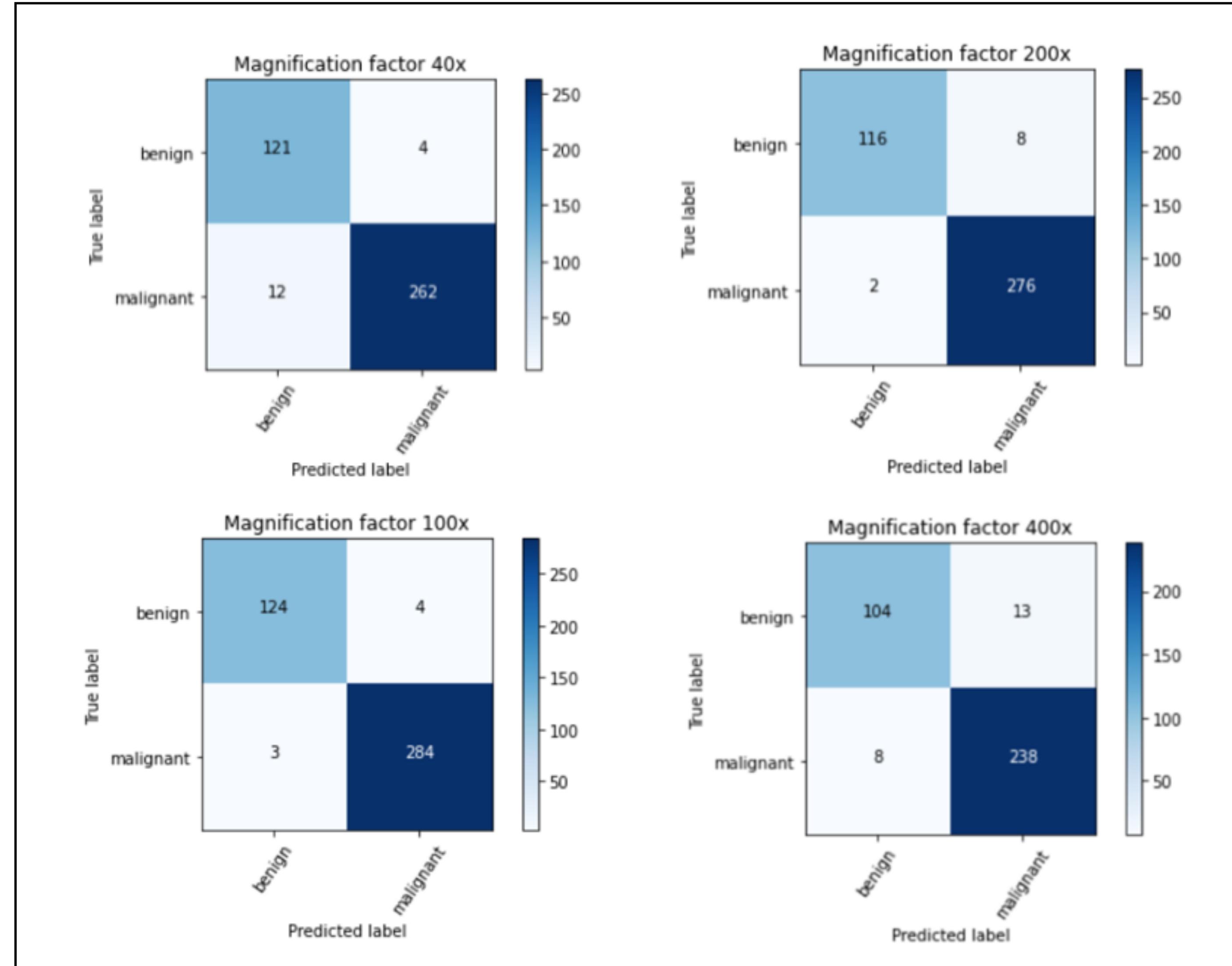
$$F1 - Score = 2 * \frac{Recall * Precision}{Recall + Precision}$$

RESULT

<i>Dataset Name</i>	<i>Accuracy</i>	<i>Sensitivity/Recall</i>	<i>Specificity</i>	<i>Precision</i>	<i>F1-score</i>
BREAKHIS	40x	95.98%	98.49%	90.97%	95.62%
	100x	98.31%	98.61%	97.63%	98.95%
	200x	97.51%	97.18%	98.30%	99.28%
	400x	94.21%	94.82%	92.85%	96.74%
ICIAR 2018	91.52%	81.48%	100%	100%	0.90
KAGGLE BHI	96.74%	99.40%	95.22%	92.26%	0.96

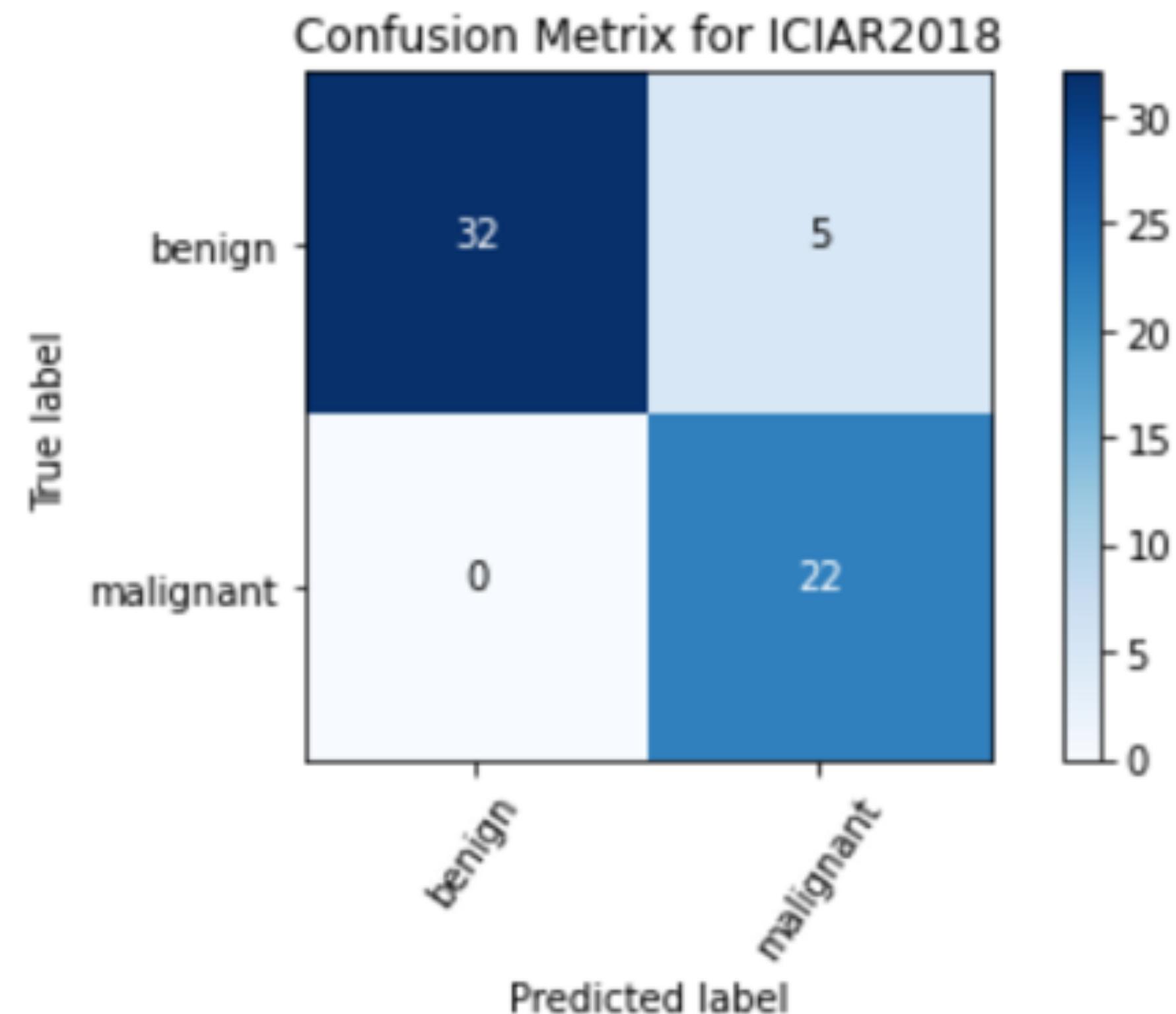
RESULT

Breakhis Dataset

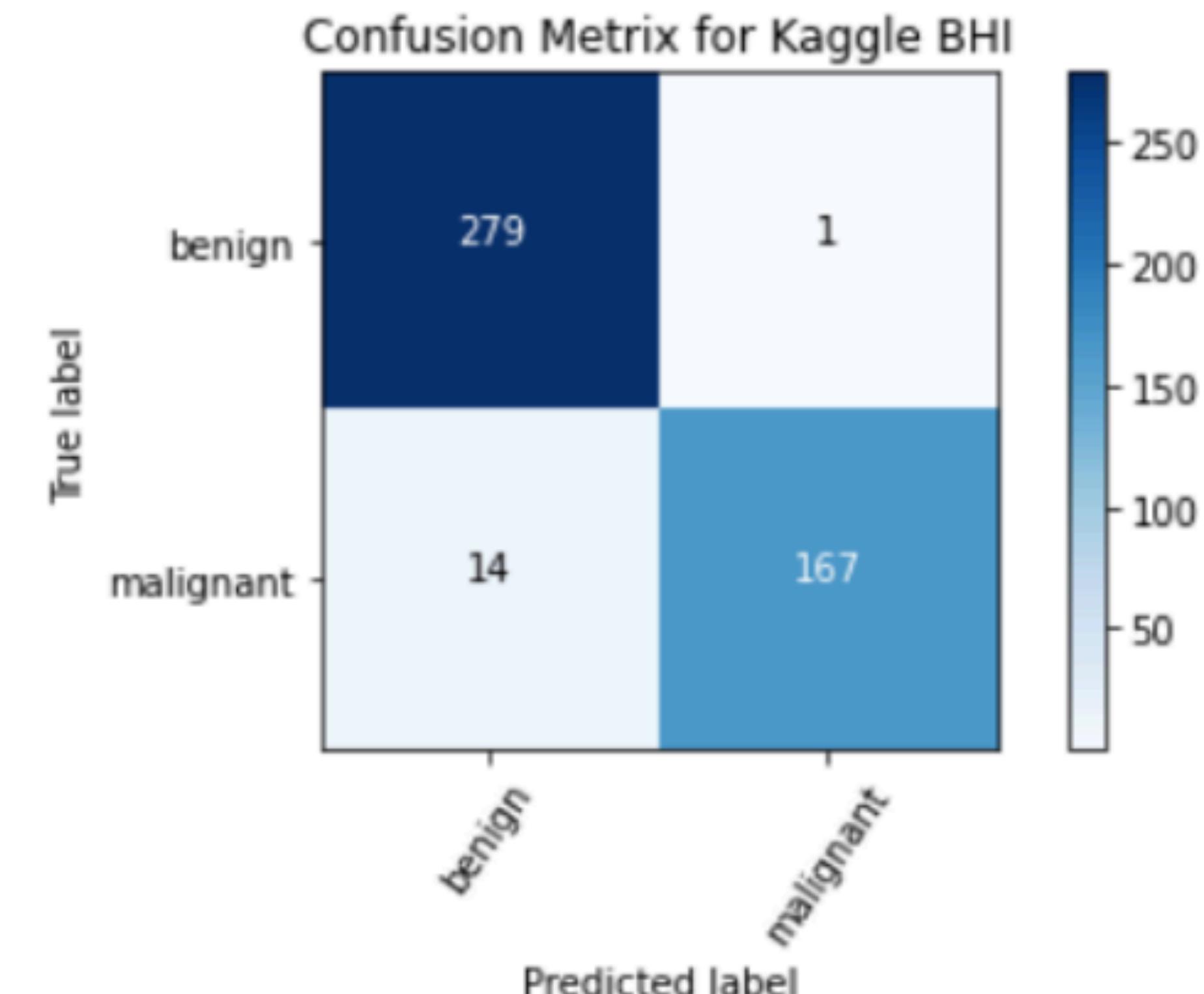


RESULT

ICIAR2018



Kaggle BHI



COMPARISON FROM OTHER MODELS ON BREAKHIS DATASET

References	40x	100x	200x	400x
Kahya et al. (2017)	94.97	93.62	94.54	94.42
Wei et al. (2017)	97.89	97.64	97.50	97.97
Pratiher et al. (2018)	96.8	98.1	98.2	97.5
Bardou et al. (2018)	96.82	96.96	96.36	95.97
Bardou et al. (2018)	92.71	93.75	92.72	92.12
Present Work	95.98	98.31	97.51	94.21

CONCLUSIONS

- The used model achieved an average accuracy of 96.5% on the breakhis dataset, 91.52% on the ICIAR2018 dataset and 96.74% on Kaggle BHI dataset
- The used model has performed good as comparable to the existing models.
- The parallel model has proved to be more efficient in terms of the accuracy as it combines all the extracted features from 5 different models and then uses all those features to classify the images.
- It is better than using a single CNN network as any single model misses some features.

REFERENCES

1. Classification of breast cancer histopathology images based on adaptive sparse support vector machine, 2017
2. Deep learning model based breast cancer histopathological image classification, 2017
3. Manifold Learning & Stacked Sparse Autoencoder for Robust Breast Cancer Classification from Histopathological Images, 2018
4. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks, 2018
5. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges, 2020.
6. Breast Tumor Classification Using an Ensemble Machine Learning Method, 2020.
7. Prediction of benign and malignant breast cancer using data mining techniques, 2018.
8. Histopathological Classification of Breast Cancer Images Using a Multi-Scale Input and Multi-Feature Network, 2020

REFERENCES

9. Dataset for Breast Cancer Histopathological Image Classification, 2016.
- 10.Breast Histopathology Images, 198,738 IDC(-) image patches; 78,786 IDC(+) image patches, 2018.
- 11.BACH: Grand challenge on breast cancer histology images, Medical Image Analysis, 2019.
- 12.Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering, 2018.
- 13.Breast cancer histopathological image classification using convolutional neural networks, 2016.
- 14.Breast cancer classification using digital biopsy histopathology images through transfer learning., 2019.
- 15.Breast Cancer Multi-classification through Deep Neural Network and Hierarchical Classification Approach, 2020.

THANK YOU

