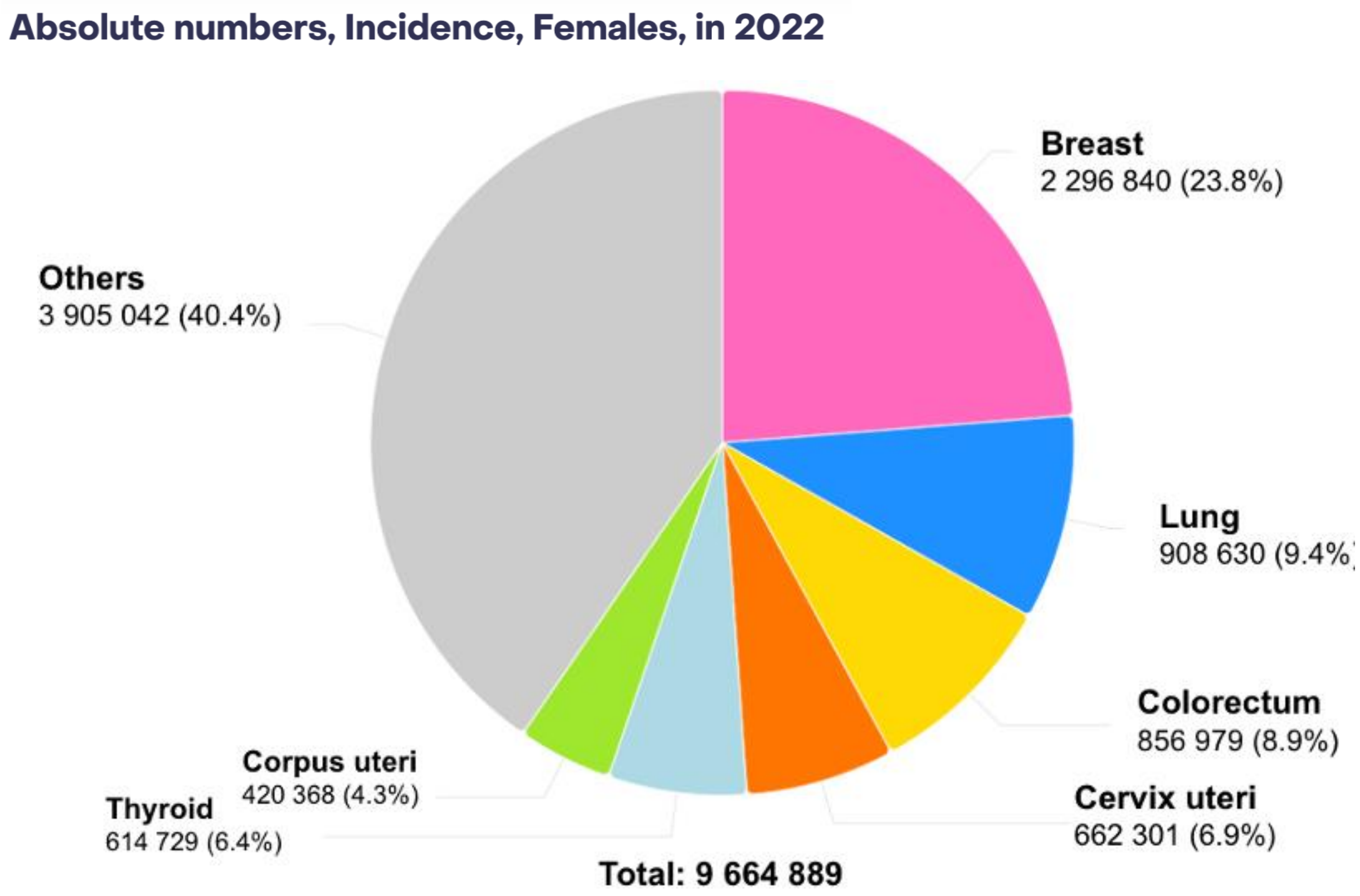
Magnification Independent Multi-Classification of Breast Cancer in Histopathology Images Using Deep Learning

Jason Nishio[1], Nelson Nishio[2], Alexander Nishio[3]

[1,2] BASIS Independent Silicon Valley, San José, California, USA

[3] Department of Computer Science, Purdue University, West Lafayette, USA

Introduction



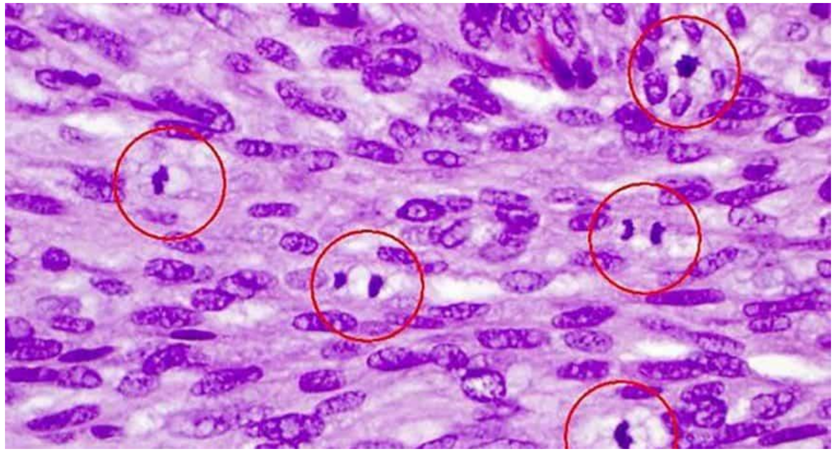
Source: International Agency for Research on Cancer, World Health Organization

- Breast cancer (BC) is the second leading cause of death for women within cancers globally
- No current cure - many deaths can be prevented if detected as early as possible
- Diagnosis methods
  - Mammogram and ultrasound: cheap, low-dose radiation, low diagnostic accuracy due to resolution limitations
  - MRI: expensive; not often used for BC diagnosis
  - **Biopsy (histopathology images):** provides informative details for the detection of many types of malignancies

Research Motivation

- **Challenges with Manual Biopsy Diagnosis**
  - Manually finding mitotic counts (indicators of cell proliferation rates/cancer aggressiveness) is labor-intensive & challenging
  - Accuracy of counts can be tainted by factors such as the pathologist's expertise, sampling biases, fatigue

- **Challenges in Automated Mitosis Detection**
  - Significant variations in cell shape and size
  - Histopathology images may contain structures that look like mitotic events but are not
  - Variations in staining colors, cell density, magnification scales and image quality across histology slides—due to differences in staining processes, scanners and microscopes—can impact the accuracy of deep learning models



Source: [Pathology News]

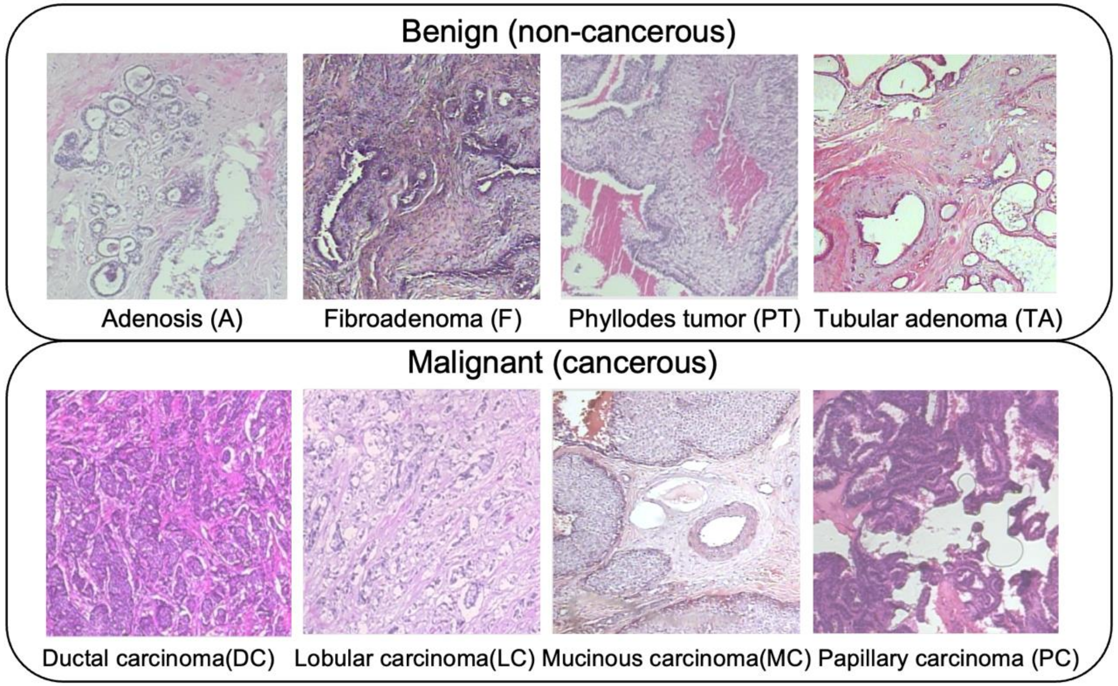
Charles Sturt University · MELBOURNE INSTITUTE OF TECHNOLOGY · University of Rajshahi · THE UNIVERSITY OF NOTRE DAME AUSTRALIA · WESTERN SYDNEY UNIVERSITY · GCSTMR

BreakHis Dataset

- BreakHis public data set (7909 images)
  - 8 tumor types – 4 benign, 4 malignant
  - 82 total patients

Binary Type	8 Tumor Types	Magnification 40x	100x	200x	400x	Total	# of Patients
Benign	Adenosis (A)	114	113	111	106	444	24
	Fibroadenoma (F)	253	260	264	237	1014	
	Phyllodes tumor (PT)	109	121	108	115	453	
	Tubular adenoma (TA)	149	150	140	130	569	
Malignant	Ductal Carcinoma (DC)	864	903	896	788	3451	58
	Lobular carcinoma (LC)	156	170	163	137	626	
	Mucinous carcinoma (MC)	205	222	196	169	792	
	Papillary carcinoma (PC)	145	142	135	138	560	
	Total	1995	2081	2013	1820	7909	

Eight Breast Cancer Tumour Types

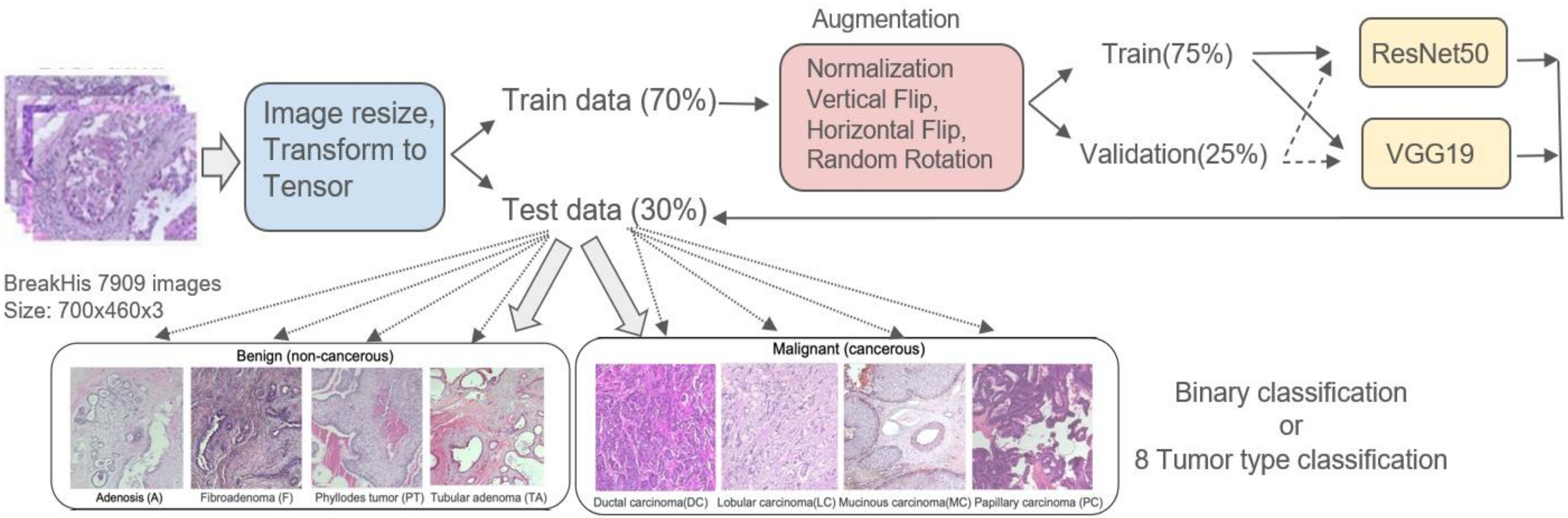


Research Gaps and Contributions

- Research Gaps of previous work using the BreakHis dataset for BC classification:
  - Predominantly focused on **binary classification**
  - Separated images by magnification factor to develop **magnification-dependent** feature extractors

- Our Contributions:
  - A **magnification-independent** deep learning framework, offering a scalable solution across various magnification levels
  - Improved accuracy for both binary & multi-class BC classification of histopathological images
  - Z-score normalization procedure introduced for histopathological images to help reduce variation in color and brightness and improving deep learning model performance.

Image Processing and Classification Flow



Z-Score Normalization based on mean and standard deviation of pixel values of the RGB channels of training dataset

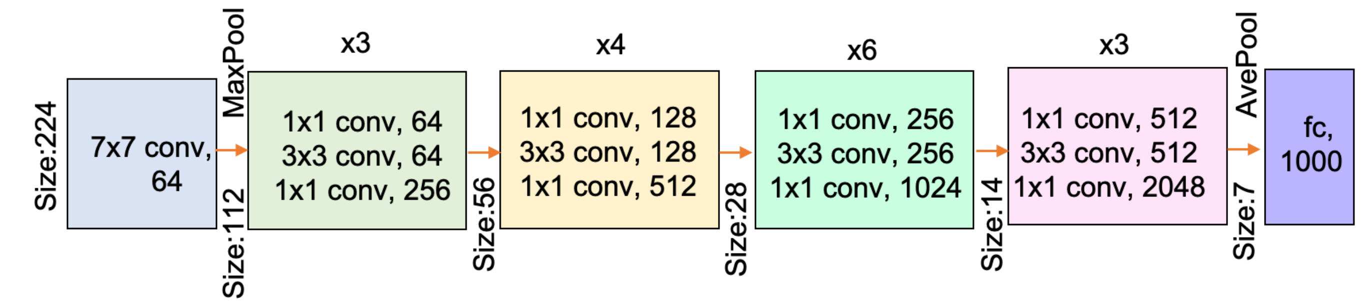
$$x' = \frac{x-\mu}{\sigma}$$

FMLDS

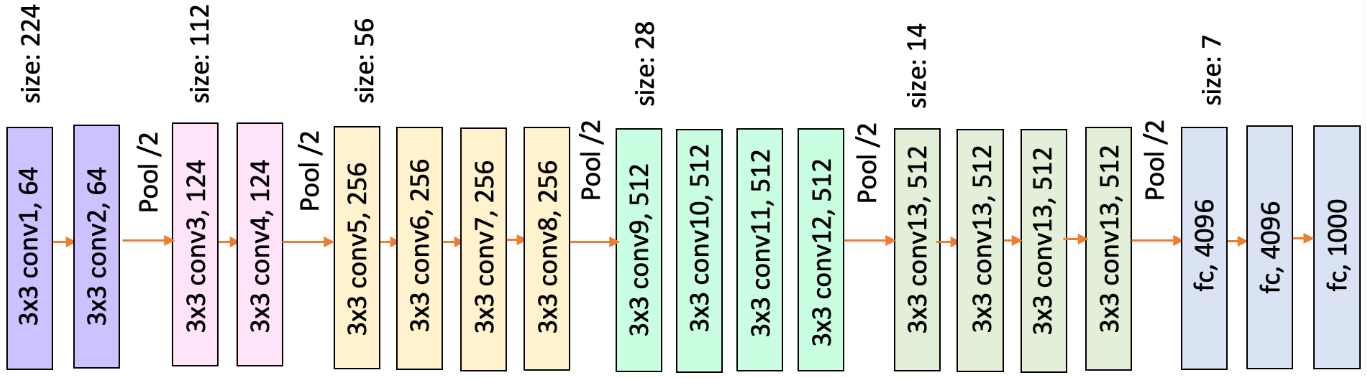
Pre-Trained Deep Learning Models

IEEE

ResNet-50 Architecture



VGG-19 Architecture

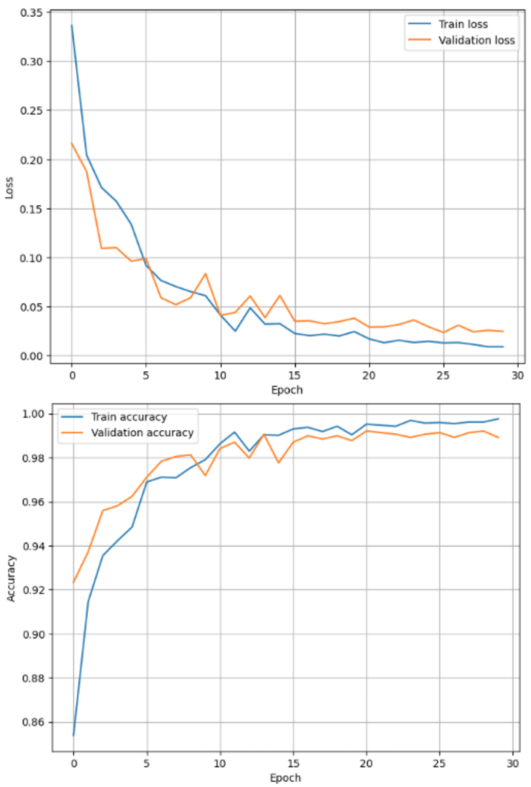


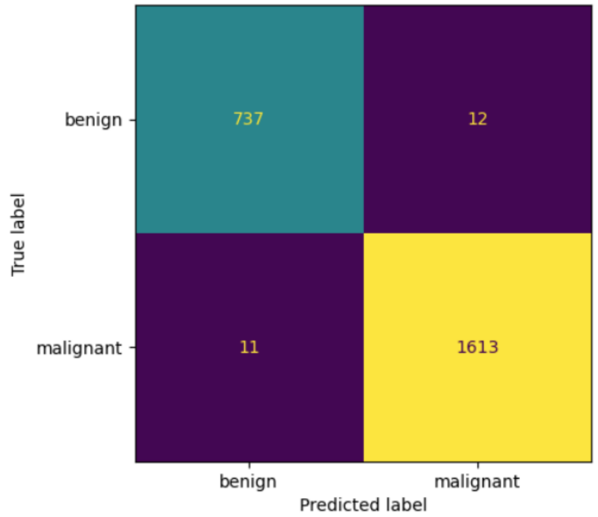
- Default weights from PyTorch pre-trained library
- Initial learning rate = 0.0003, decreases every 5 epochs by factor of 0.5
- Adam optimizer better performance than SGD
- Batch size = 32; Max epoch = 50
- Cross entropy loss function
- PyTorch for data loader and normalization
- Scikit-learn for splitting data
- Scikit-learn for performance matrices calculation
- Smote for augmentation
  - Training set before and after over-sampling: 5536 and 19232, respectively
  - Sample count in each class: 2404
- Trained on Google Colab

Charles Sturt University | MELBOURNE INSTITUTE OF TECHNOLOGY | University of Rajshahi | THE UNIVERSITY OF NOTRE DAME AUSTRALIA | WESTERN SYDNEY UNIVERSITY | GCSTMR

Binary Classification Results (no data augmentation)

ResNet-50: 99.03% accuracy

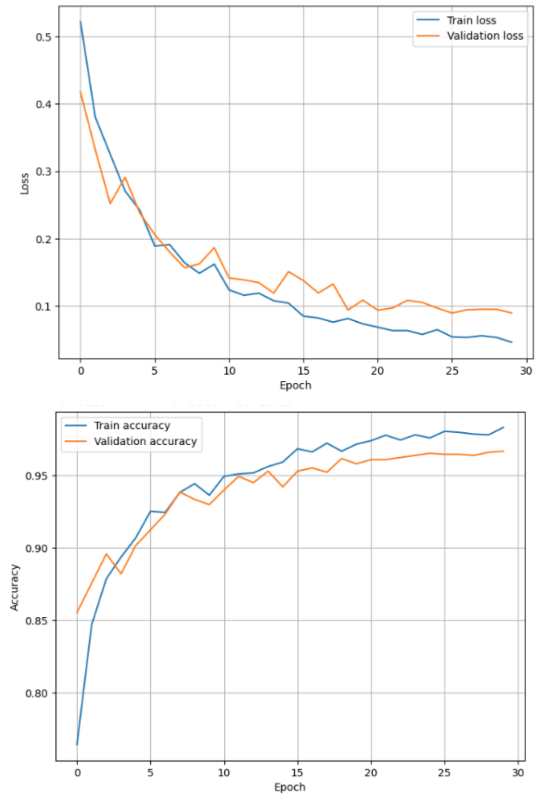


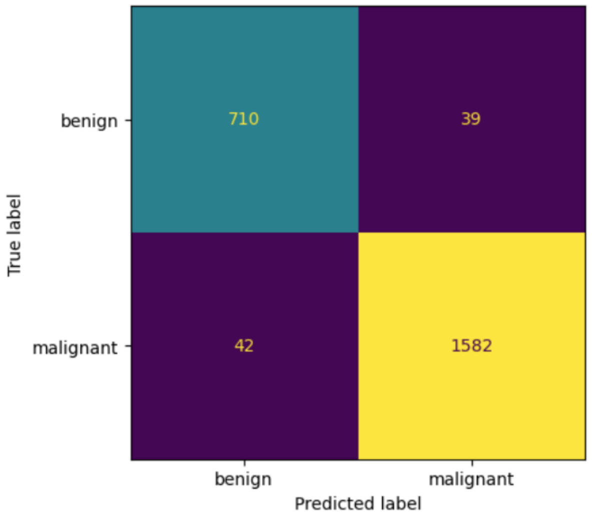


	precision	recall	f1-score	support
benign	0.9853	0.9840	0.9846	749
malignant	0.9926	0.9932	0.9929	1624
accuracy			0.9903	2373
macro avg	0.9890	0.9886	0.9888	2373
weighted avg	0.9903	0.9903	0.9903	2373

```
Accuracy for benign: 0.9903
Accuracy for malignant: 0.9903
```

VGG-19: 96.59% accuracy





	precision	recall	f1-score	support
benign	0.9441	0.9479	0.9460	749
malignant	0.9759	0.9741	0.9750	1624
accuracy			0.9659	2373
macro avg	0.9600	0.9610	0.9605	2373
weighted avg	0.9659	0.9659	0.9659	2373

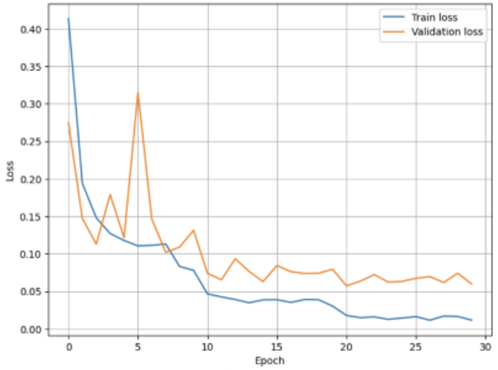
```
Accuracy for benign: 0.9659
Accuracy for malignant: 0.9659
```

FMLDS

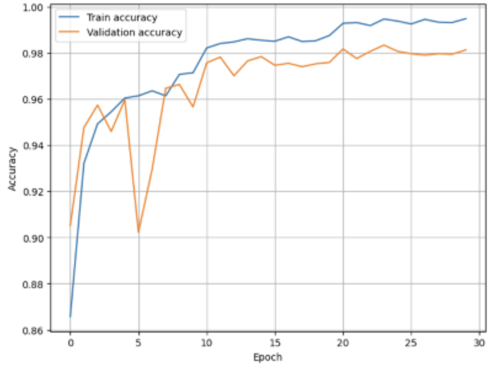
Multi-Class Classification Results (no data augmentation)

IEEE

ResNet-50: 93.97% accuracy



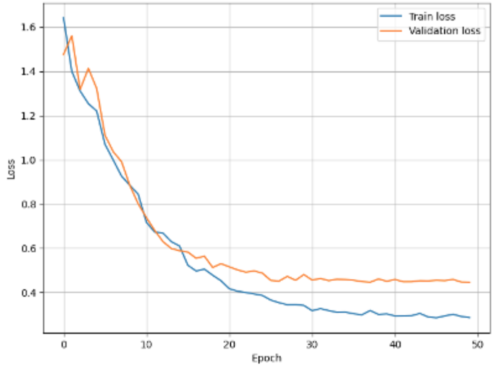
ResNet50 loss curves



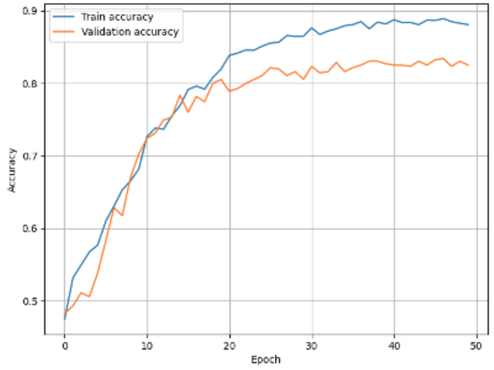
ResNet50 accuracy curves

True label \ Predicted label	A	DC	F	LC	MC	PC	PT	TA
A	123	0	0	0	1	0	0	0
DC	1	992	3	41	3	7	0	0
F	2	0	295	0	2	1	7	1
LC	2	34	0	144	1	0	0	0
MC	0	7	1	1	227	1	0	1
PC	1	4	1	0	1	151	0	0
PT	2	2	9	2	0	1	131	0
TA	0	0	3	0	0	0	0	167

VGG-19: 88.33% accuracy



VGG19 loss curves



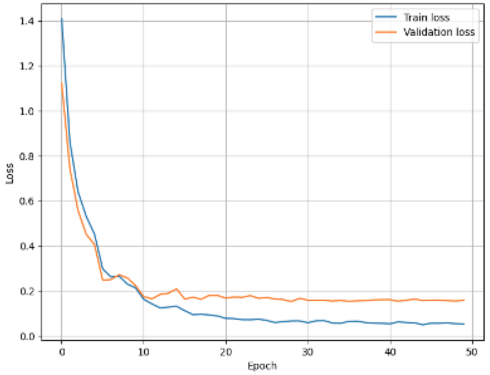
VGG19 accuracy curves

True label \ Predicted label	A	DC	F	LC	MC	PC	PT	TA
A	117	0	0	1	2	0	4	0
DC	1	975	14	29	9	15	1	3
F	1	6	282	1	2	0	10	6
LC	1	56	0	116	7	0	1	0
MC	0	16	6	8	202	4	1	1
PC	0	9	6	0	6	137	0	0
PT	9	7	18	0	1	0	111	1
TA	0	2	10	0	1	0	1	156

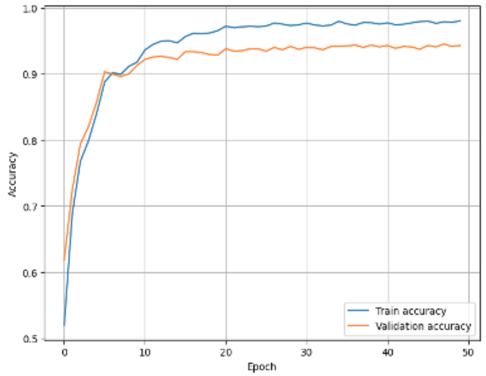
Charles Sturt University | Melbourne Institute of Technology | University of Rajshahi | The University of Notre Dame Australia | Western Sydney University | GCSTMR

Multi-Class Classification Results after data augmentation

ResNet-50: 94.18% accuracy



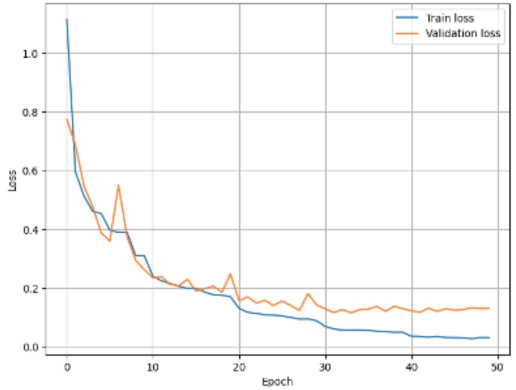
ResNet50 loss curves



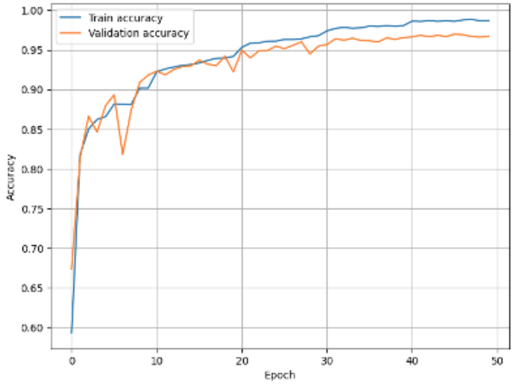
ResNet50 accuracy curves

True label \ Predicted label	A	DC	F	LC	MC	PC	PT	TA
A	123	1	0	0	0	0	0	0
DC	0	981	3	45	9	7	2	0
F	0	2	298	0	3	0	3	2
LC	0	19	0	158	3	0	1	0
MC	0	0	0	1	234	1	0	2
PC	1	5	1	0	0	149	1	1
PT	2	3	13	2	0	0	126	1
TA	0	0	4	0	0	0	0	166

VGG-19: 92.50% accuracy



VGG19 loss curves



VGG19 accuracy curves

True label \ Predicted label	A	DC	F	LC	MC	PC	PT	TA
A	123	0	0	0	1	0	0	0
DC	1	979	4	39	17	6	1	0
F	3	4	296	0	0	1	3	1
LC	0	34	0	141	6	0	0	0
MC	1	3	1	2	230	1	0	0
PC	1	2	1	0	0	151	2	1
PT	3	5	21	4	0	1	113	0
TA	1	1	4	1	0	1	0	162

FMLDS · IEEE · Charles Sturt University · Melbourne Institute of Technology · University of Rajshahi · The University of Notre Dame Australia · Western Sydney University · GCSTMR

Performance Metrics Results Summary

Classification	Model	Accuracy	Precision	Recall	F1-score	Augmentation
Binary	ResNet50	99.03%	98.53%	98.40%	98.46%	No
	VGG19	96.59%	96.00%	96.10%	96.05%	No
8-Class	ResNet50	93.97%	93.08%	93.45%	93.24%	No
	ResNet50	94.18%	93.11%	94.12%	93.53%	Yes
	VGG19	88.33%	86.95%	85.25%	85.97%	No
	VGG19	92.50%	91.46%	91.38%	91.26%	Yes

Comparison with Previous Research

Ref	Year	Model	Magnification	Accuracy	Precision	Recall	F1-Score
Binary Classification							
[6]	2023	DenseNet169	Dependent (200x and 400x)	87.14%	91.21%	90.54%	78.93%
		Bit-S-R101x1		86.88%	90.67%	90.79%	77.45%
[5]	2023	VGG19	Independent	96.46%	93.14%	95.77%	94.43%
This Work	2024	ResNet50	Independent	99.03%	98.90%	98.86%	98.88%
		VGG19	Independent	96.59%	96.00%	96.10%	96.05%
Mult-class Classification							
[7]	2020	Handcrafted Features VGG16+ SVM	Dependent (40x)	93.97%	94.00%	93.00%	94.00%
			Dependent (200x)	91.23%	92.00%	92.00%	92.00%
[8]	2022	Handcrafted Features + DNN	Dependent (40x)	97.89%	97.00%	98.00%	97.00%
[9]	2022	DenseNet121	Dependent (40x)	89.00%	90.00%	85.00%	87.00%
		VGG19	Independent	68.00%	-	-	-
		ResNet50	Independent	80.00%	-	-	-
		DenseNet121	Independent	87.20%	85.00%	83.00%	87.00%
This Work	2024	ResNet50	Independent	94.73%	94.30%	94.23%	94.25%
		VGG19	Independent	92.50%	91.46%	91.38%	91.26%

Conclusion

- The established transfer learning-based deep learning classifiers are **not susceptible to variations in magnification factors**
- **The Z-score image normalization** helps reducing color and brightness variation and improving deep learning model performance
- ResNet50 is a well-fitted model, with training loss and validation loss in proximity throughout the training process, even without data augmentation.
  - **99.03%** and **94.18%** acc. for binary and multi-classification, respectively
- VGG19: overfitting issue was addressed by data augmentation
  - **96.59%** and **92.50%** acc. for binary and multi-classification, respectively
- Our deep learning framework achieved better accuracy for the challenging task of classifying histopathological images of breast tumors in the BreakHis dataset

Limitations and Future work

- Limitations
  - The BreakHis dataset was extracted from only 82 patients with imbalanced number of images
  - Several tumor categories have only a few hundreds of images which are insufficient to build a robust classifier
  - The current framework is limited to pre-trained Resnet50 and VGG19 models.
  - The current model performance is based on BreakHis dataset.
- Future Work
  - Increasing the training set size by merging with another dataset or finding a larger & balanced dataset
  - Exploring other pre-trained models, such as ResNet-RS, EfficientNetV2, and Vision Transformers
  - Extending the established framework to other biomedical imaging tasks beyond breast cancer (we have successfully validated them on retinal fundus images for diabetic retinopathy classification)
  - Investigating other image enhancement methods to further improve the performance
  - Further hyperparameter tuning

Thanks, and Q&A