Breast Cancer Tumour Tissue Classification: A Deep Learning CNN Approach

3244-1910-0026-6

Chen Huijia, Thangavel Sharan, Khaw Yew Onn, Huang Yida,
Jeffrey Tan Jian Sheng, Low Si Jia Jessica

NUS National University of Singapore

School of Computing

Overall Objective: **To classify breast cancer tumour tissue stains as benign or malignant using a deep learning CNN approach.**

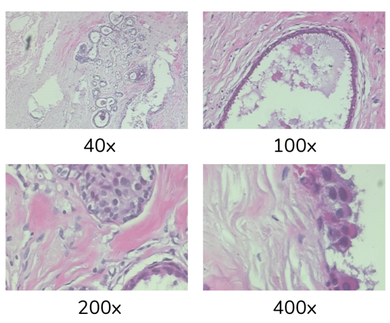
Abstract

Breast cancer is the most common disease among women, with 2 million new cases in 2018 alone (Bray, 2018). Diagnosing breast cancer based on biopsy images, however, can be challenging as visually similar tumours can be either malignant (cancerous) or benign (harmless) in nature. Classification into malignant or benign affects the treatment received by the patient, and incorrect classification could lead to undesirable health consequences, including death.

Thus, by harnessing the potential of available data **(>7000 tagged images)** and CNN, we aim to enhance existing capabilities in breast cancer tumour classification in the areas of efficiency and accuracy.

Our Dataset

We requested and obtained **7,909 microscopic images** of breast tumour tissue that were collected from **82 patients** by the Laboratory of Vision, Robotics and Imaging at the Federal University of Parana. This breast cancer histopathological database will be used for training & testing purposes. The data is divided into **4 magnifications** as shown on the right:



Pre-processing

Different methods of pre-processing were utilised, including data augmentation, to prevent overfitting and allow more images for training.

Data augmentation

We attempted to flip the dataset horizontally/ vertically, rotate the dataset and zoom randomly into the dataset using the Keras Preprocessing library ImageDataGenerator.

Accounting for the underlying class imbalance between benign and malignant training images

Our initial dataset had considerably more malignant cell images (**5429** vs **2480**). To counter the resultant 'Accuracy Paradox' effect, we applied data augmentation on both the benign and malignant data, but subsequently randomly removed part of the malignant dataset for training. This allowed us to expose the model to malignant and benign cell data, and thus features, more equally. Eventually, we used **2480 images** for malignant and benign each (before augmentation), and they were split into **training : validation : test** as **0.5 : 0.25 : 0.25.**

Methodology

		To:
1	Perform Exploratory Data Analysis (EDA) on dataset, involving data augmentation & manipulation	Understand dataset better and minimise overfitting
2	Test out available pre-trained neural networks on our whole dataset for binary classification	Analyse performance and assess shortcomings
3	Divide the dataset and analysis further, based on available patient metadata	Minimise contamination of test set and understand effect of certain patients on result
4	Account for underlying class imbalance and tweak hyperparameters further	Overcome the accuracy paradox and improve model relevance & performance

Noteworthy Addition: We parallelized the loading of data from Google Drive to the Google Colab Server, which reduced the time taken from >60mins to <10mins!

Results

We were able to achieve the best result using a combination of the following:

- **Model:** VGG16
- **Weights:** 'ImageNet'-trained, unfrozen
- **Image Size:** 300 by 300
- **Optimiser:** RMSprop, lr=1e-5
- **Loss:** Binary Cross Entropy
- **Epochs:** 30

	Precision	Recall	f1-score
0 (Benign)	0.85	0.88	0.86
1 (Malignant)	0.88	0.85	0.86

Table 1: Performance of best-performing model

	Predicted: Benign	Predicted: Malignant
Actual: Benign	591	81
Actual: Malignant	107	600

Table 2: Confusion Matrix

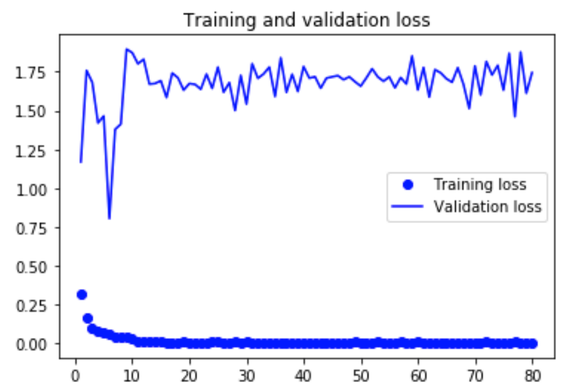


Figure 3: Training and Validation Loss

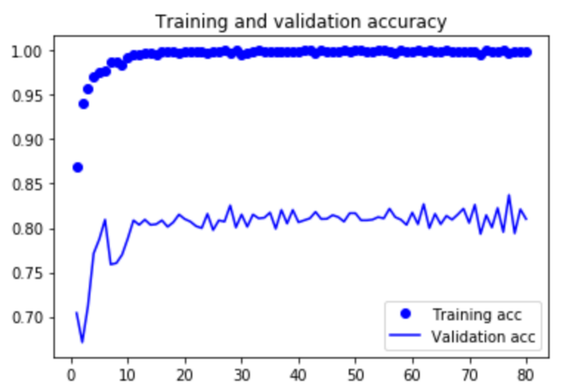


Figure 4: Training and Validation Accuracy

Evaluation / Discussion

Why did we unfreeze the 'ImageNet' weights?

	Precision	Recall	f1-score
0 (Benign)	0.79	0.79	0.79
1 (Malignant)	0.80	0.80	0.80

Table 5: Performance with frozen 'ImageNet' weights

The custom added layer's weights are far from the true value because they are randomly initialised. Subsequently, this would propagate very large weight updates through the network. We avoid this situation by freezing the 'ImageNet'-trained weights while training only the custom added layers in our first run. Retraining the model after unfreezing the 'ImageNet' weights in the bottom layers gave us better performance as the 'ImageNet' dataset has limited cellular-level images for relevant training in our case.

How did data augmentation affect the model's performance?

Making minor alterations to the images (flipping, rotating, zooming), allowed us to expose our model to more variations, thus reducing overfitting. Training on the **400x** magnification images with the most detail produced the best performance.

	Precision	Recall
0 (Benign)	0.77	0.66
1 (Malignant)	0.72	0.82

Table 6: No data augmentation

How did adjusting the imbalance between benign and malignant cell data affect the model's performance?

Our initial model was trained on **5429** malignant cell images, and only **2480** benign cell images. This split gave good prediction for malignant cases, but performed no better than coin flip for benign cases as a result of the 'Accuracy Paradox', where the accuracy merely reflected the underlying distribution across classes. Cutting down malignant cell images to **2480** then improved performance as shown in Table 1.

Why did we split the dataset by individuals?

Ensuring that each person's data does not exist in more than one subset (training, validation, test) was crucial to preventing cross-contamination that initially gave us an overly optimistic accuracy score of **>0.90** for both benign and malignant classification. We improved on this by studying the image metadata, and separating the **~100** cell images/person to evaluate the model on an individual-independent test set.

Next Steps:

- **Training the model with data of higher quantity and diversity so as to reduce any potential overfitting, and improve ability to generalise outside of our given dataset.**
- **Testing out the model with fine-needle aspirate cell tissue images from other types of cancer, such as lung cancer.**
- **Upon procuring more data for each of the 8 subclasses within malignant and benign, attempting multi-class classfication.**