RSNA Intracranial Hemorrhage Detection

jfooyh@stanford.edu, https://drive.google.com/open?id=1P98uk8qTqm-_-eskHzfQm2Ke9qOSLlK0

Problem

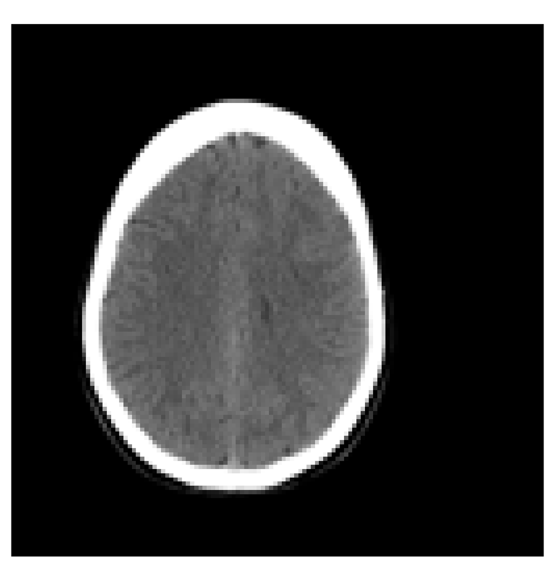
Intracranial haemorrhage, bleeding in the brain, requires urgent medical treatment. Deep learning can be used to detect hemorrhages from medical images, to screen and filter likely cases that can then be referred to medical personnel for confirmation.

Data

674, 257 raw DICOM images were provided by Kaggle in stage 1 of the competition, each image labelled by a binary class of whether a haemorrhage exists or not. The images were then pre-processed accordingly, to reduce the size of the images for computation[1]:

(1) Transform DICOM values to Hounsfield units
(2) Apply windowing to rescale pixel values
(3) Resize images to 128x128x3
(4) Load and train images in minibatches

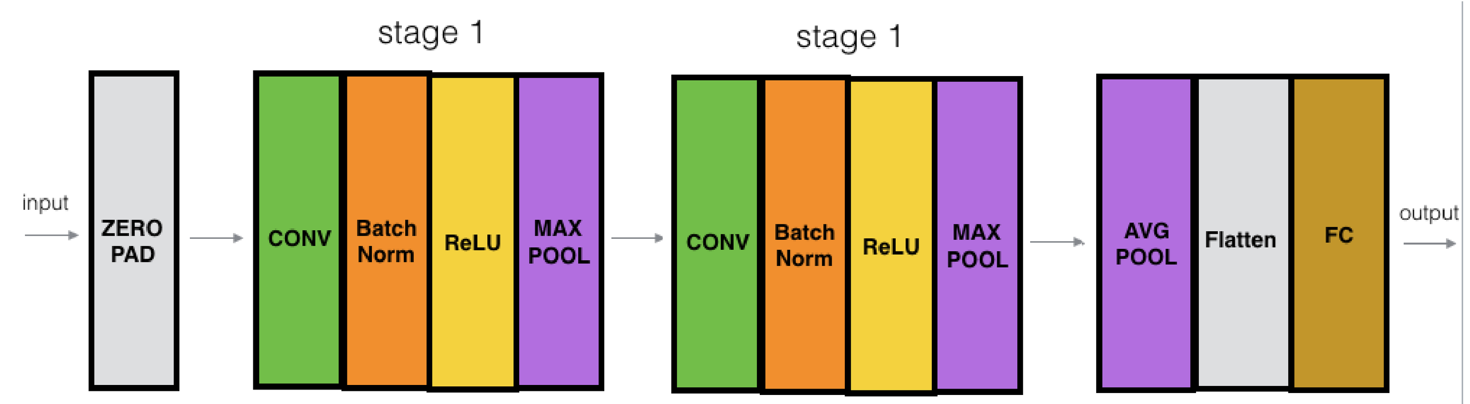
(i) Sample labelled 1



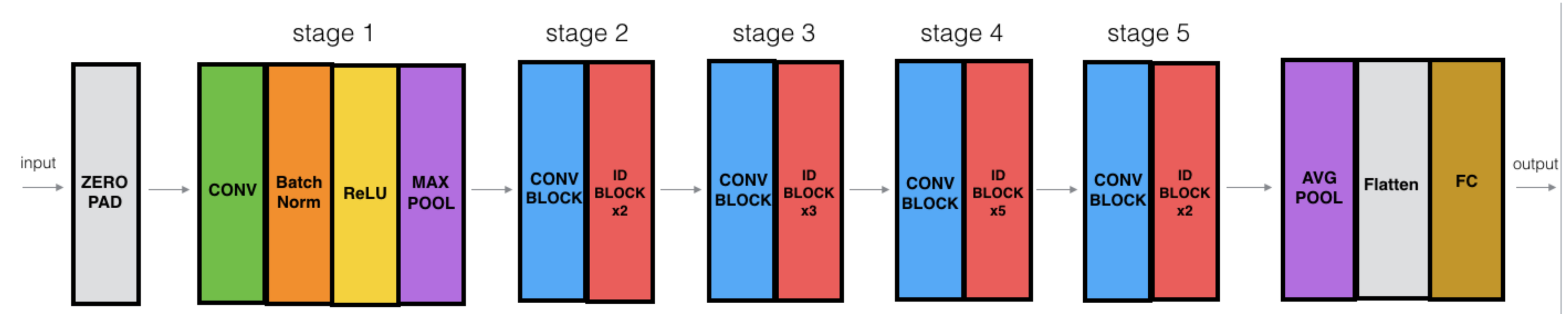
(ii) Sample labelled 0

Model

1) **2-layer Convolutional Neural Network**



2) **ResNet50** – Transfer Learning from ImageNet



- Allowed weights of Stage 5 and last block (i.e., FC layer) to be trained

For each model, the minibatch size was tuned (64, 128, 256), and the drop out probability in the last block

Optimization

1) **Weighted cross-entropy binary loss**

$$-(w_1 y \log(p) + w_0(1-y)\log(1-p))$$

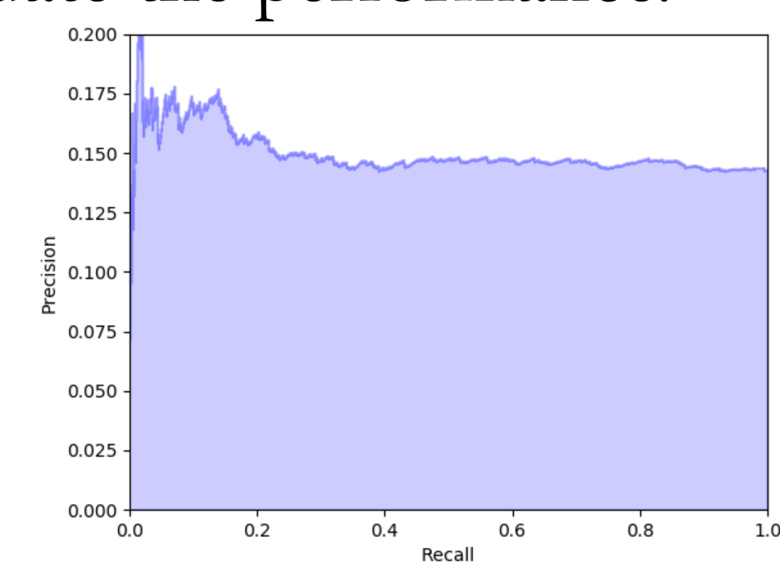
where p is the predicted probability of class 1, y is the true label, $w_1 = \frac{1}{\#\ samples\ in\ class\ 1}$ and $w_0 = \frac{1}{\#\ samples\ in\ class\ 0}$

2) **Learning rate (LR)** – overfitted on 10 minibatches to find best LR (0.1, 0.01, 0.05, 0.001, 0.0001), decay rate (0.1, 0.01)

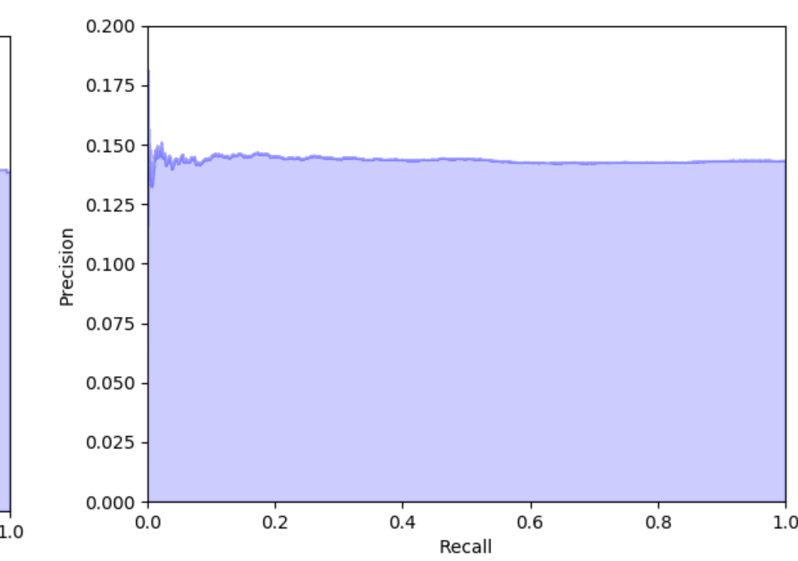
3) **Optimizer** – SGD (Nesterov momentum = 0.9), Adam

Results and Discussion

Both models achieve a validation accuracy of around 86%. However, accuracy is not a good metric since majority of the observations in the dataset are negative (85%). Instead, we use the precision-recall (PR) curve to evaluate the performance.



(i) PR curve (2-layer CNN)



(ii) PR curve (ResNet50)

- 2-layer CNN performs slightly better than ResNet50, since the area under the PR curve for the simpler model is larger. This could be because we used pretrained weights from ImageNet, which may not transfer well to medical images.
- In general, precision seems to stagnate at 0.15, even as recall increases. This means that the model is not able to correctly identify that a hemorrhage exists but is good at identifying that a hemorrhage does not exist. While this may be acceptable in the context of medical screening, there's scope to improve the model

Future Work

- Use raw DICOM images to capture higher resolution[2]
- Segment patches of the scan to generate local, patch-level classifications

[1] Guillermo Fernandez, *Prepare dataset (resizing and saving as png)*, Kaggle kernel, https://www.kaggle.com/guiferviz/prepare-dataset-resizing-and-saving-as-png

[2] Krzysztof J. Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S. Gene Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy, Kyunghyun Cho, *High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks* (2017), arXiv:1703.07047