Classifying breast tumours with convolutional networks

Team: Anyone Welcome!

Z5259601, z5408907, z5255835, z5361944, z5308765

Matthew Leo, Yiheng Wei, Angelo Zhou, Keerthan Gopal, Ethan Reinhard

### Introduction

Breast cancer is one of the most common causes of death among women worldwide. In Australia, it is the second most common cancer in people and the most common form of cancer found in women. With 1 in 7 women diagnosed in their lifetime and the diagnosis rate increasing by 21% over the last 10 years [2], it is pivotal that we can improve our techniques to screen and diagnose women early. Why it is important is that early detection increases patients’ survival rates greatly, where stage 1 has on average a 100% survival rate however this drops as the cancer advances, with stage 2 cancers having an average survival rate of 92%, stage 3 at 82% and stage 4 with a 32% survival rate [1]. Thus it is pivotal that medical professionals are able to screen patients and detect any tumours of concern early to maximise patient survival rate. It is shown that in many countries around the world, due to cost and portability, it can be hard for many to access a mammogram in comparison to an ultrasound machine, which is much cheaper for the patient and able to be deployed in many more locations [3]. In addition to above, there is also a growing shortage of radiologists and other medical professionals, in Australia, where there is a growing disparity in the number of radiologists per capita in rural areas compared to metropolitan [5]. The US is expected to have a shortage of 17000 to 42000 radiologists, pathologists and psychiatrists [7] whereas in the rest of the world, two-thirds of the population is currently not able to access medical imaging technology [6]. There is potential in convolutional neural networks improving medical professional diagnostic accuracy, efficiency and time of diagnosis [8]. Our data is from a Kaggle dataset [4] of breast ultrasound images, which we trained various convolutional networks upon to help both identify and classify whether or not the patient has a tumour of concern as we aim to demonstrate and confirm that neural networks indeed have the potential to aid in medical screening.

### Literature Review

As we can see in existing literature, especially Zhang’s survey of existing works on breast cancer identification tasks with convolutional networks[9], there is overwhelming evidence in support of the convolutional neural networks potential to improve radiologists accuracy when screening patients. Many of the models reviewed by Zhang achieving high 80-90% accuracy, specificity and sensitivity [9], which shows the architecture has a good potential in not only correctly identifying the presence of a tumour from a various scans, depending on what dataset the model was trained on, but in correctly identifying true positives and true negatives. This lends itself for us to further explore the use of convolutional networks to help us try to demonstrate the effectiveness and flexibility of a convolutional network when we train these models against contexts such as the kaggle dataset.

### Methods

We will be training our own implementations of the following well known convolutional network architectures: AlexNet, VGG16, Inception, ResNet152v2. Our method was to train-test split the ultrasound data and then simply replicate the model's architecture as stated in their original papers with the same hyperparameters chosen. We would then train the model ourselves on the training split of data, which we then evaluated on the evaluation split of ultrasound data.

### Experimental Setup

Our data is sourced from [here](https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset). The input data was 780 images that were labelled as malignant or non-malignant. The data was split into a training and test set with a 70-30 ratio. Four models were trained to maximise the ratio of AUC on the training set and validated by being run on the test set. Model learning rate and momentum variables were controlled by using the ADAM optimiser upon each one using a binary cross entropy as our loss function.

**Model parameters:**

**Resnet158v2:** This model has 60.4 million parameters and has 158 convolutional layers all using 3x3 filters with 4 max pooling layers. The first convolution layer has a kernel of 2 and padding of 3 before it is batch normalised and ReLu is applied. Layer 0 has a stride of 1 but all other layers have a stride of 2 with padding at 1.

**VGG16:** This model had 138 million parameters, with 13 convolutional layers. 2 of which have 64 filters, 2 with 128, 3 with 256 and 5 with 512 filters with 5 max pooling layers. Linear layers have a dropout rate of 0.5. All kernels have a stride of 1 with padding of 1. Followed by a batch normalisation before ReLu is applied to each layer's output.

**Inception V3:** This model has 25.8 million parameters. It utilises different sized kernels in the same layer to determine the most important features. We look at kernels of size 1, 3 and 7.

**Alex net:** This model has 21.5 million parameters. It utilises 5 convolution layers from 96, 256, 384, 384 and finally 256 filters. The first layer has a stride of 4 and padding at 0, the rest of the layers have stride at 1 and padding at 1 except layer 2 which has padding at 2. Finally, all linear layers have a dropout rate of 0.5. Each convolutional layer has its output batch normalised and all layers have ReLu applied to them except the final classification layer.

### Evaluation Metric

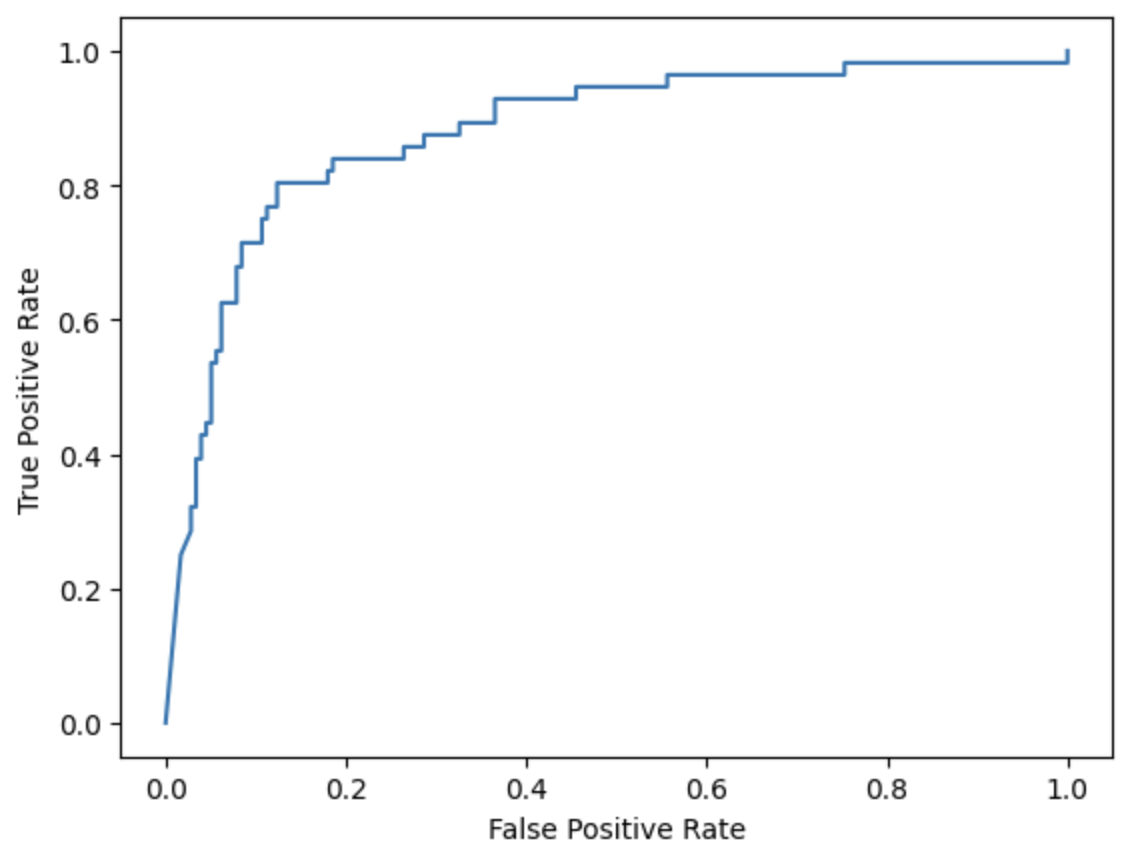
The ratio of AUC was used as the evaluation metric for the models. The purpose of the model is screening for cancer, so we want a model with a high TPR paired with a reasonable FPR.

**Validation**

The model was trained on training data. Its performance was then assessed by the AUC score it achieved on the test data set.

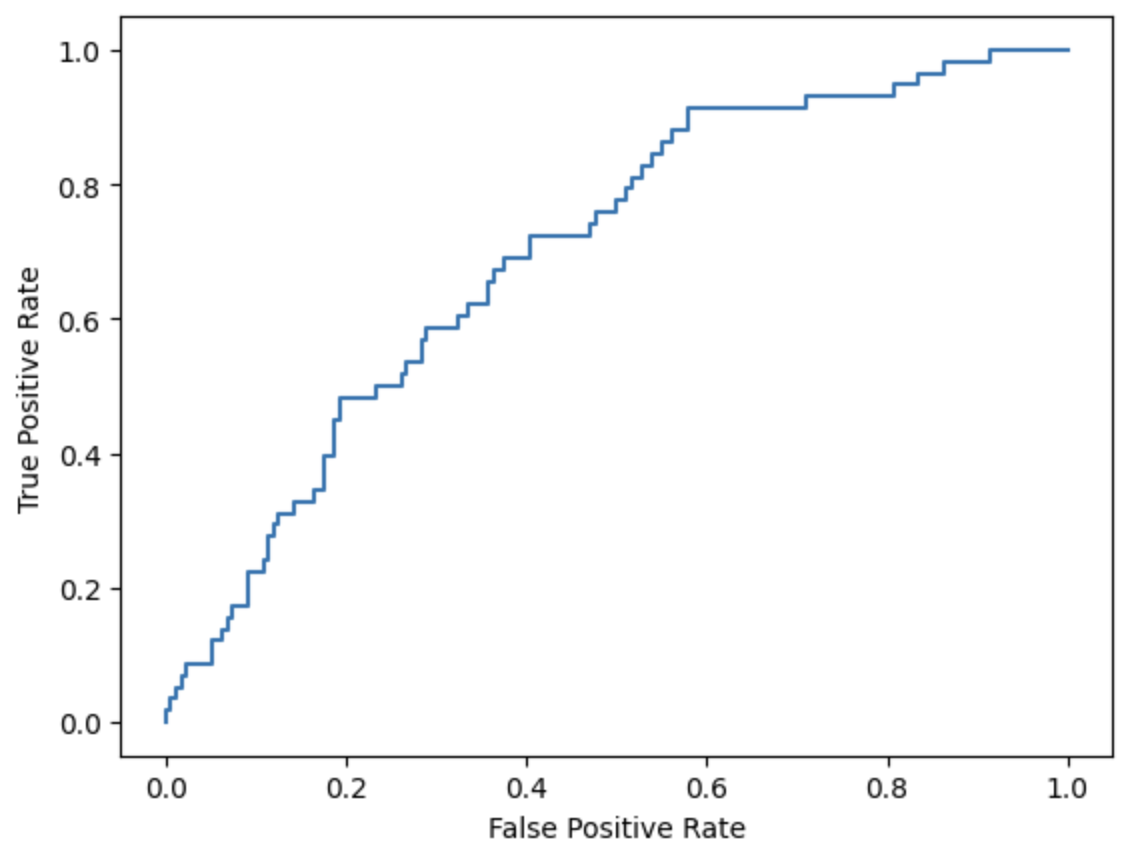
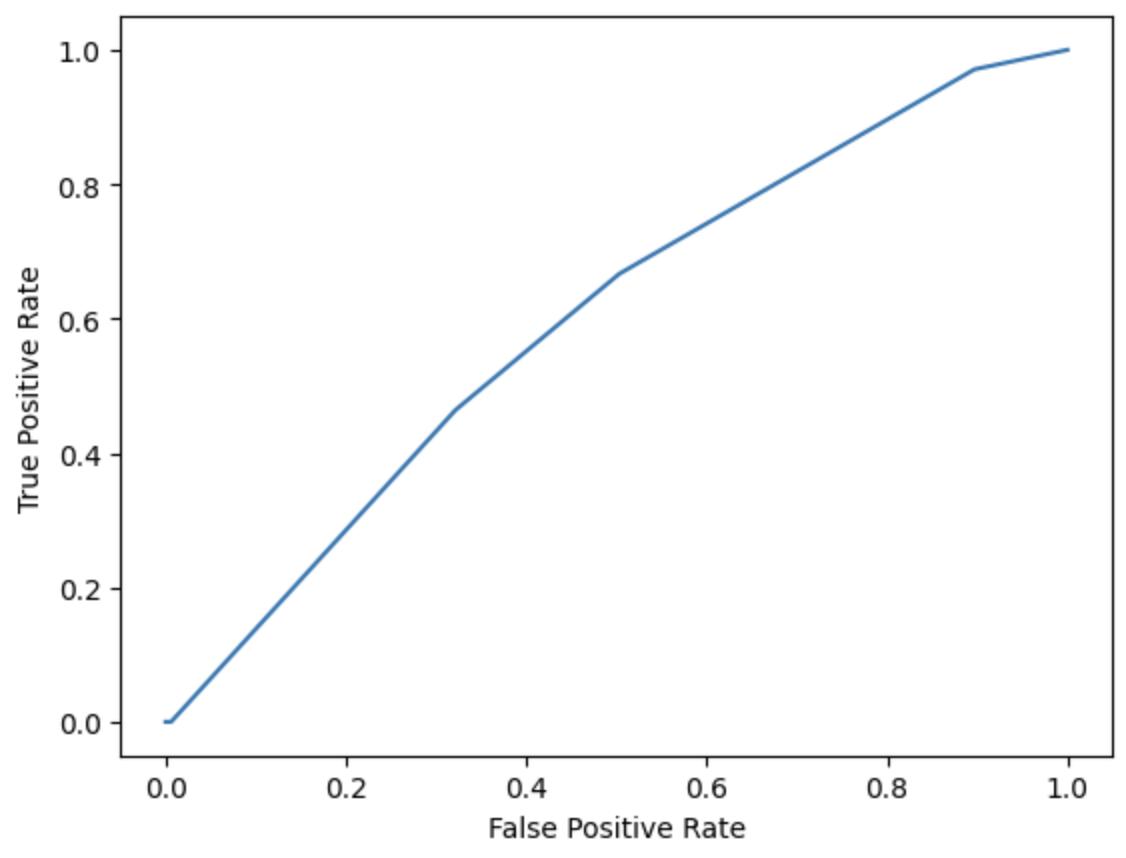
### Results

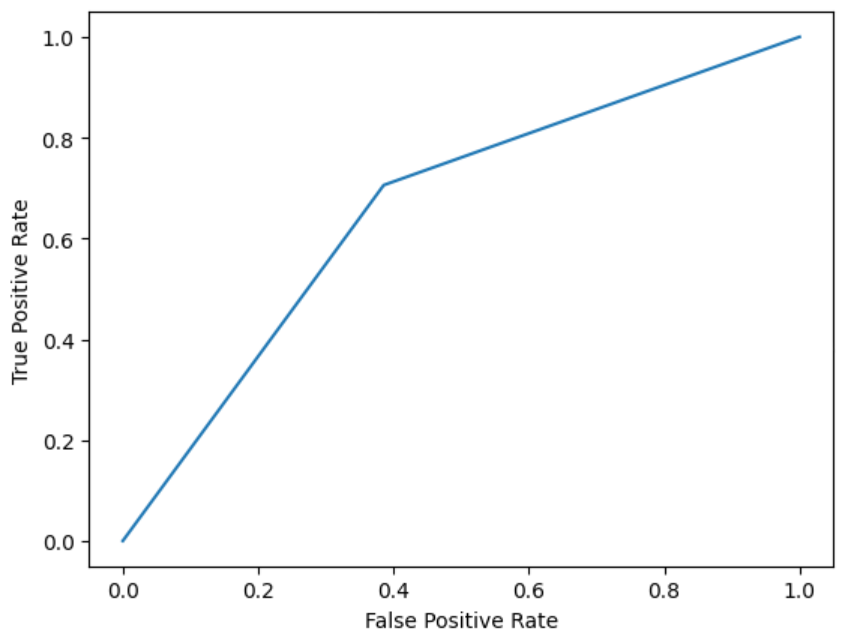
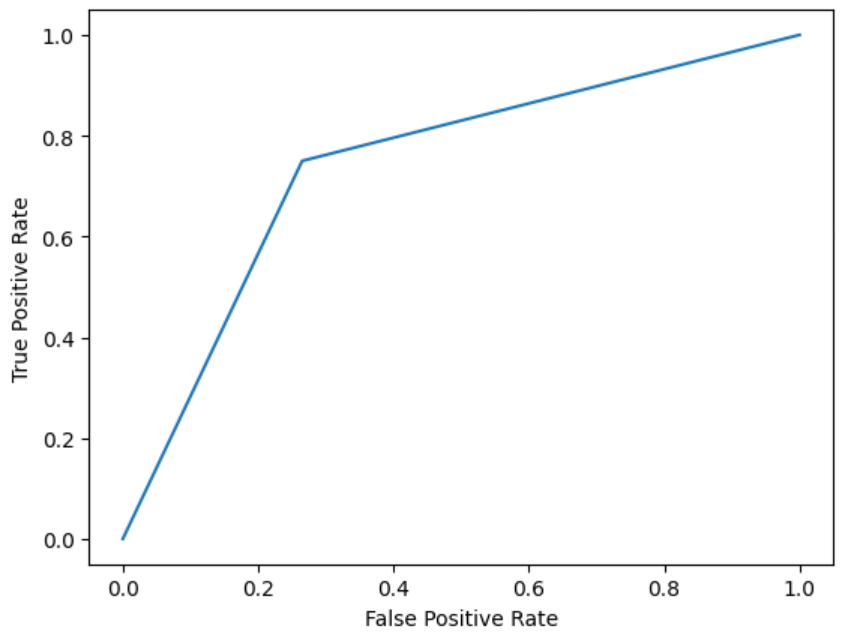
**ResNet152V2:**



**VGG16:**

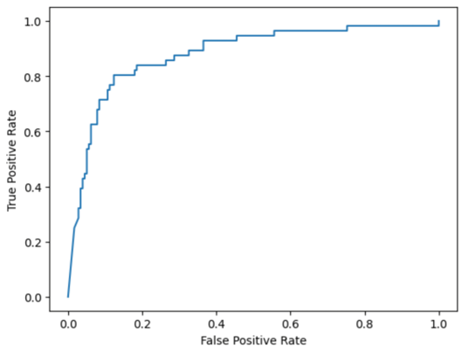
Without Batch Normalisation: With Batch Normalisation:



**Inception V3:**  **AlexNet:**

**AUC**

|  | Training AUC | Test AUC |
| --- | --- | --- |
| ResNet152V2 | 0.977 | 0.860 |
| VGG16 | 0.5 | 0.5 |
| VGG16(Batch Normalisation) | 0.718 | 0.728 |
| InceptionV3 | 0.713 | 0.736 |
| AlexNet | 0.874 | 0.817 |

A well-performing model will produce a roc curve which tends towards the top left-hand side of the graph and produces a high AUC (area under the curve). A significant element that can be interpreted from a ROC curve is the rate of false and true positives as a certain probability threshold. The further an ROC deviates towards the top left corner signals the existence of certain probability thresholds where the rate for false positives is low, meanwhile, the rate for true positives is high. Now applying this to our own results graphs such as ResNet152V2 and AlexNet, it can be seen that at the probability threshold of 0.8 the model performs extremely well. Furthermore, the AUC is also high as it deviates towards the top left corner. On the other hand VGG16 without batch normalisation and InceptionV3 performed quite poorly. The default shape for an ROC curve is a straight diagonal line, symbolising the chance for a false positive is just as likely as a true positive, making no real significance when predicting. The graph for VGG16 without batch normalisation deviates from this default shape by a very small amount, signifying its lack of ability to decrease the chance of false positives while increasing the chance of true positives. This shape also produces a very average AUC of 0.5. 

ResNet152v2 achieved an AUC score of 0.86 on the test set. This was the best performing model out of the 4 models selected. It could detect over 80% of breast cancer cases with a false positive rate of less than 20%. Although the model is useful, state of the art models in this area have achieved AUC scores of 0.945, significantly outperforming this model.

### Conclusion

A model that can detect over 80% of positive cases with a false positive rate of 20% is a useful model. However, obviously an even higher TPR and a lower FPR would be ideal. Two key limitations faced in this project included limited time and computing power. Having more time to customise the model parameters, add layers and more computing power will likely lead to better results.

Reference

[1] “Breast cancer,” Breast cancer in Australia statistics | Cancer Australia, https://www.canceraustralia.gov.au/cancer-types/breast-cancer/statistics (accessed Nov. 11, 2023).

[2] “Breast cancer stats,” National Breast Cancer Foundation (NBCF) | Donate Online, https://nbcf.org.au/about-breast-cancer/breast-cancer-stats/ (accessed Nov. 11, 2023).

[3] E. Devolli-Disha, S. Manxhuka-Kërliu, H. Ymeri, and A. Kutllovci, “Comparative Accuracy of Mammography and Ultrasound in Women with Breast Symptoms According to Age and Breast Density,” *Bosnian Journal of Basic Medical Sciences*, vol. 9, no. 2, pp. 131–136, May 2009, doi: <https://doi.org/10.17305/bjbms.2009.2832>.

[4] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi: <https://doi.org/10.1016/j.dib.2019.104863>.

[5] Sanjay Jeganathan, “The Growing Problem of Radiologist Shortages: Australia and New Zealand’s Perspective,” *Korean Journal of Radiology*, vol. 24, no. 11, pp. 1043–1043, Jan. 2023, doi: <https://doi.org/10.3348/kjr.2023.0831>.

[6] “Wayback Machine,” web.archive.org, Oct. 23, 2012. <https://web.archive.org/web/20121023154427/http://www.who.int/eht/en/DiagnosticImaging.pdf>

[7] Dall T, Reynolds D, Chakabarti R, et al. The complexities of physician supply and demand: projections from 2018 to 2033. Washington, D.C. Associaion of American Medical Colleges, June 2020. <https://www.aamc.org/system/files/2020-06/stratcomm-aamc-physician-workforce-projections-june-2020.pdf>

[8] K. G. van Leeuwen, M. de Rooij, S. Schalekamp, B. van Ginneken, and M. J. C. M. Rutten, “How does artificial intelligence in radiology improve efficiency and health outcomes?,” *Pediatric Radiology*, vol. 52, Jun. 2021, doi: <https://doi.org/10.1007/s00247-021-05114-8>.

[9] Z. Zhu, S. Wang, and Y. Zhang, “A Survey of Convolutional Neural Network in Breast Cancer,” Cmes-computer Modeling in Engineering & Sciences, vol. 136, no. 3, pp. 2127–2172, Jan. 2023, doi: https://doi.org/10.32604/cmes.2023.025484.