Breast Cancer Classification using Computer Vision

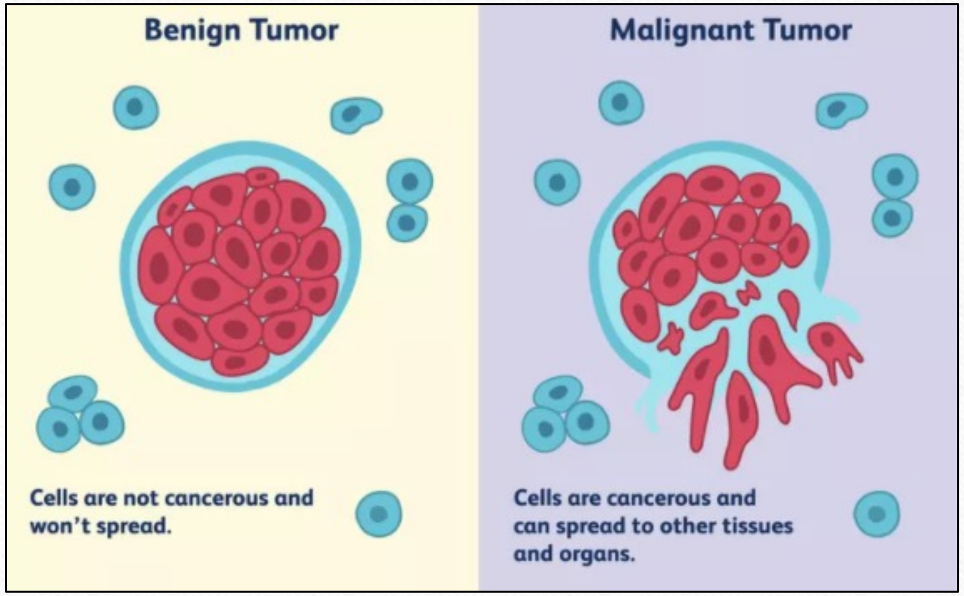
Image Classification Using CNN

- *Chandramouli Yalamanchili*

Abstract

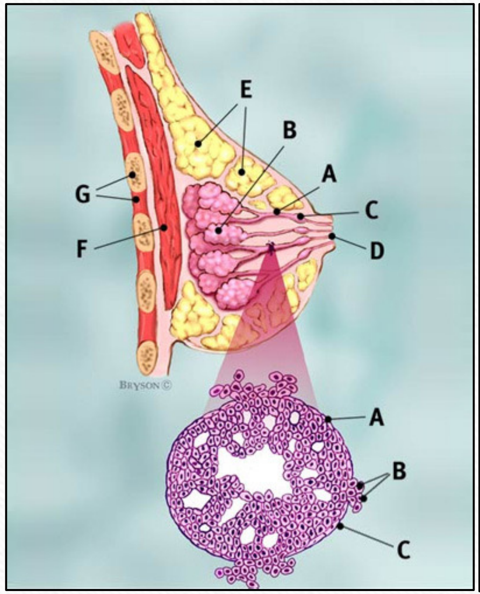
- Breast cancer is one among the foremost common forms of cancer in American women, with approximately 276K women were estimated to be diagnosed with breast cancer in 2020.
- Invasive Ductal Carcinoma (IDC) is the common type of breast cancer, contributing to 85% of total carcinoma cases.
- Considering the progress that's been made in last few decades in computer vision and image classification technology, any automation that these technologies can bring in in early detection of breast cancer would be of tremendous help.
- Goal for this project is to build a Convolutional Neural Network (CNN) and see how efficiently this model can classify the cancer cells.

Domain Background - IDC, Benign, Malignant?



Benign Tumor – Non-cancerous, won't spread to nearby cells.

Malignant Tumor – Cancerous, and it invades near by cells and other body parts.



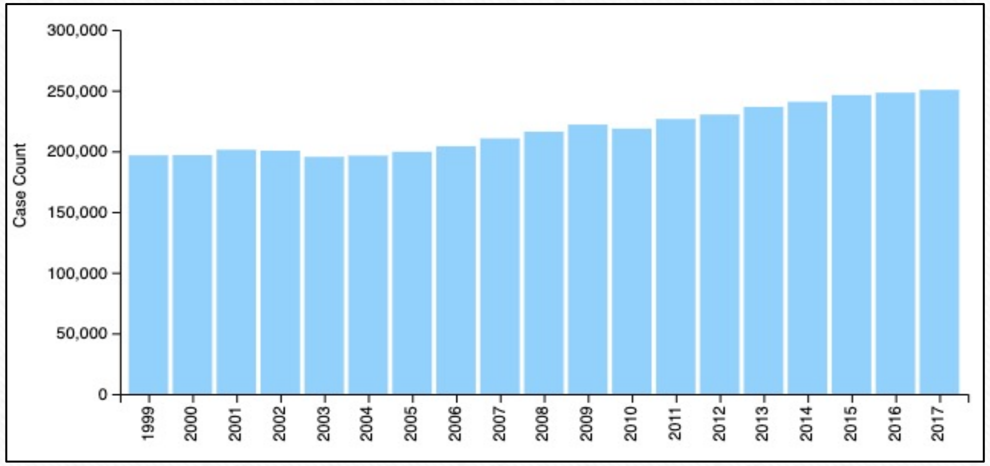
Breast profile:
A Ducts
B Lobules
C Dilated section of duct to hold milk
D Nipple
E Fat
F Pectoralis major muscle
G Chest wall/rib cage

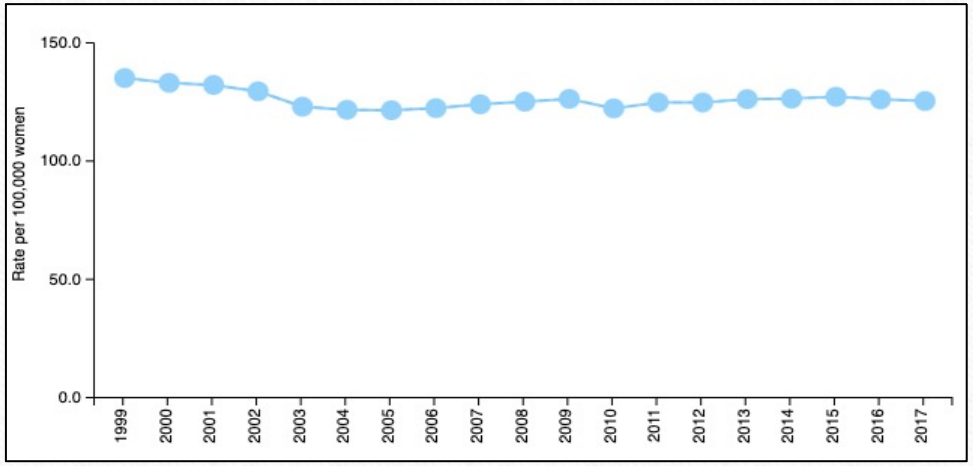
Enlargement
A Normal duct cell
B Ductal cancer cells breaking through the basement membrane.
C Basement membrane

Invasive Ductal Carcinoma (IDC).

Showing the abnormal growth in the enlarged cell within the duct.

Domain Background – Breast Cancer Stats

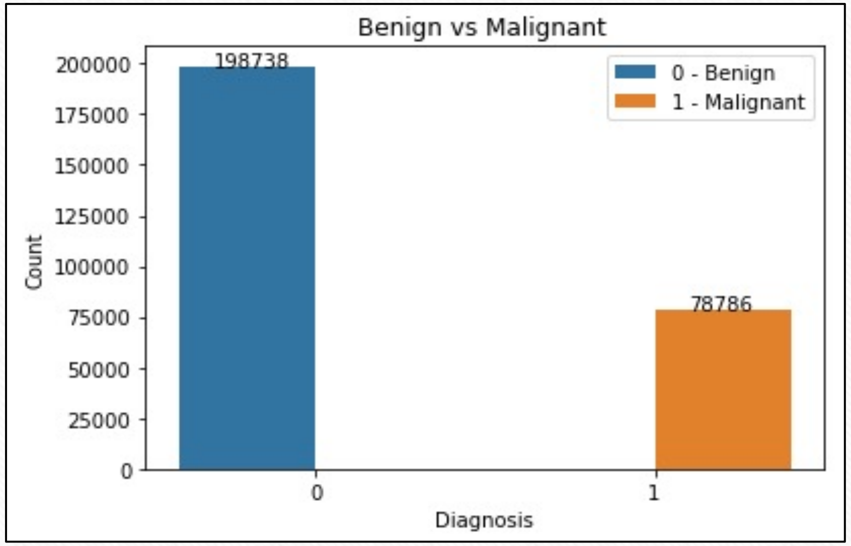




- 276,480 women are estimated to be diagnosed with breast cancer in 2020.
- 42,170 women are estimated to die from breast cancer in the year 2020.
- 81% of breast cancer diagnosis are invasive.

Exploratory Data Analysis - Input dataset

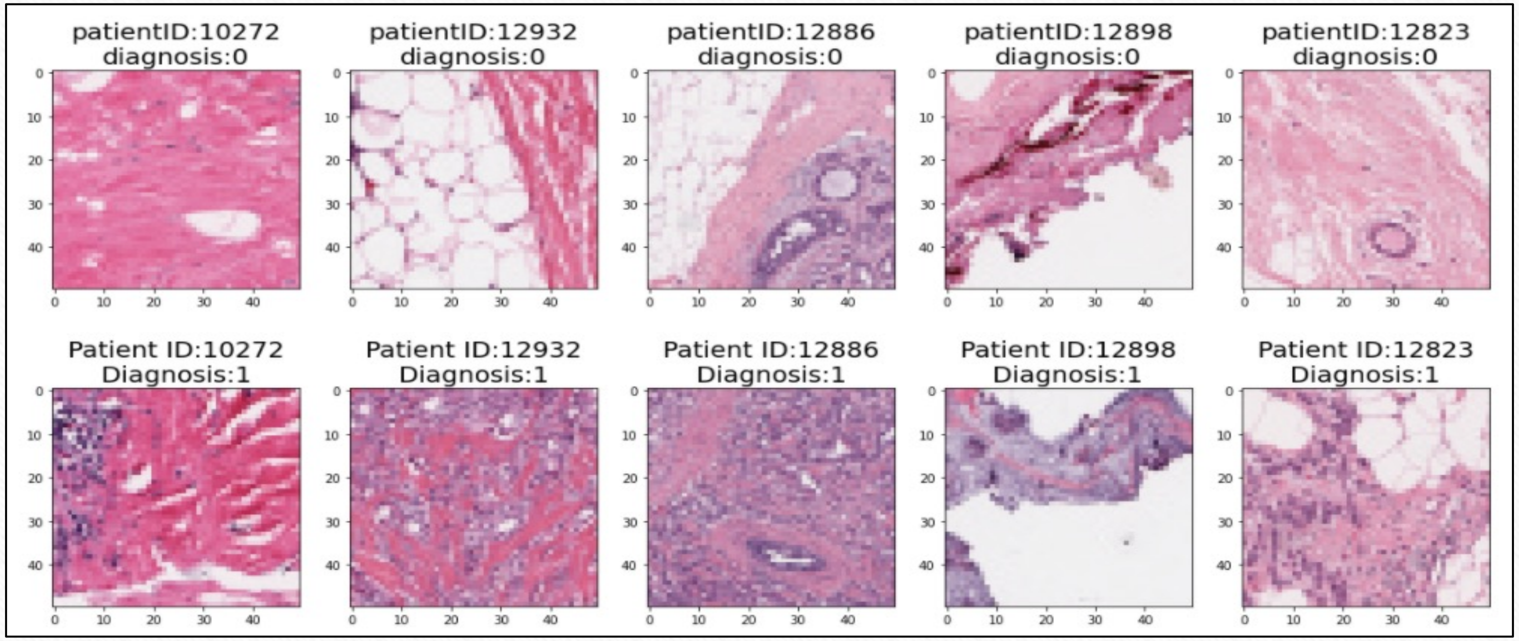
Parameter	Value
Number of patients	279
Total number of images	277,524
Number of benign cells	198,738
Number of malignant cells	78,786



Dataset Location:

https://www.kaggle.com/paultimothymooney/breast-histopathology-images

Exploratory Data Analysis – Benign vs. Malignant cells



Benign vs. Malignant histology tissue images for 5 random patients.

Modeling - Convolutional Neural Network

- CNN with 31 layers, that accepts 48X48 images and outputs 2 classes.
- Used Adagrad optimizer with initial learning rate of 0.01.
- Keras ImageDataGenerator for training, and test data.
- Class Weights to handle imbalanced dataset.
- Trained for 20 Epochs, with selected sample of patient's data.

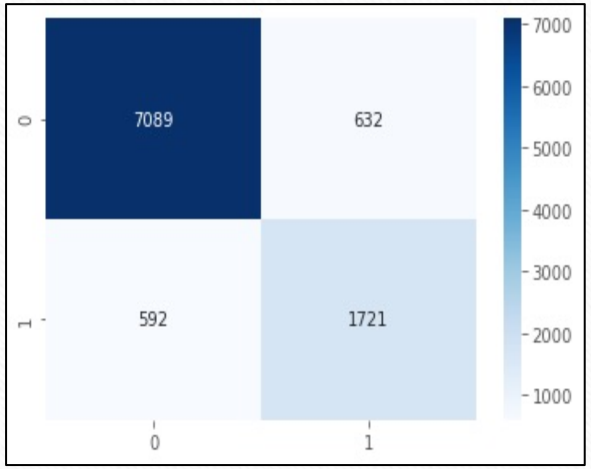
Model: "sequential_1"

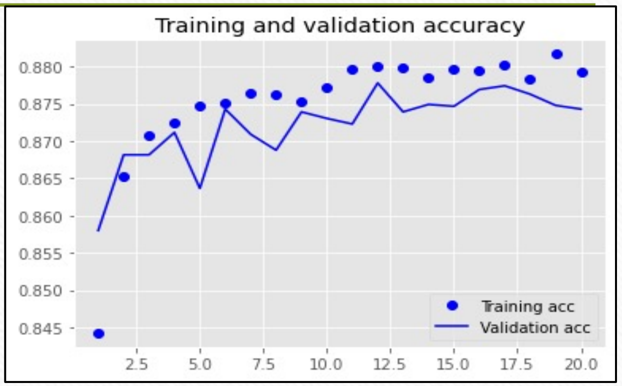
Layer (type)	Output Shape	Param #
separable_conv2d_1 (Separabl	(None, 48, 48, 32)	155
activation_1 (Activation)	(None, 48, 48, 32)	0
batch_normalization_1 (Batch	(None, 48, 48, 32)	128
max_pooling2d_1 (MaxPooling2	(None, 24, 24, 32)	0
dropout_1 (Dropout)	(None, 24, 24, 32)	0
separable_conv2d_2 (Separabl	(None, 24, 24, 64)	2400
activation_2 (Activation)	(None, 24, 24, 64)	0
batch_normalization_2 (Batch	(None, 24, 24, 64)	256
separable_conv2d_3 (Separabl	(None, 24, 24, 64)	4736
activation_3 (Activation)	(None, 24, 24, 64)	0
batch_normalization_3 (Batch	(None, 24, 24, 64)	256
max_pooling2d_2 (MaxPooling2	(None, 12, 12, 64)	0
dropout_2 (Dropout)	(None, 12, 12, 64)	0
separable_conv2d_4 (Separabl	(None, 12, 12, 128)	8896
activation_4 (Activation)	(None, 12, 12, 128)	0
batch_normalization_4 (Batch	(None, 12, 12, 128)	512
separable_conv2d_5 (Separabl	(None, 12, 12, 128)	17664
activation_5 (Activation)	(None, 12, 12, 128)	0
batch_normalization_5 (Batch	(None, 12, 12, 128)	512
separable_conv2d_6 (Separabl	(None, 12, 12, 128)	17664
activation_6 (Activation)	(None, 12, 12, 128)	0
batch_normalization_6 (Batch	(None, 12, 12, 128)	512
max_pooling2d_3 (MaxPooling2	(None, 6, 6, 128)	0
dropout_3 (Dropout)	(None, 6, 6, 128)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_1 (Dense)	(None, 256)	1179904
activation_7 (Activation)	(None, 256)	0
batch_normalization_7 (Batch	(None, 256)	1024
dropout_4 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 2)	514
activation_8 (Activation)	(None, 2)	0

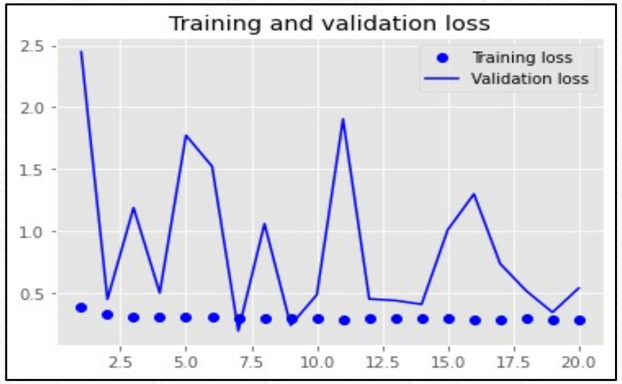
Total params: 1,235,133
Trainable params: 1,233,533
Non-trainable params: 1,600

Modeling – CNN Model performance

Performance Metric	Value
Accuracy Score	**88%**
Precision Score	**73%**
Recall/Sensitivity Score	**74%**
F1 Score	**74%**
Specificity	**74%**







- Accuracy of 88% reflects that the model has done a decent job with image classification.
- The 74% of recall and specificity values indicates the model's moderate performance with true positives and true negatives.
- Based on the accuracy and loss plots, we can see that the model is overfitting as the validation loss is not quite following the training loss.
- Based on confusion matrix as well, model has 74% of accurate cancer prediction results (TN).

Conclusion

- Breast cancer is an increasing concern in today's world, it is one among the foremost common forms of cancer in American women. Within breast cancer, Invasive Ductal Carcinoma (IDC) is the most common category, accounting for 85% of all carcinoma cases.
- There has been substantial progress made in the computer vision and image classification technology over last few decades, making it possible to use these techniques in healthcare domain for medical imaging and diagnosis. Any contributions the computer vision technology can add in early detection of breast cancer would help in reducing the mortality rate of cancer patients every year.
- In this project we have taken the histology images of the IDC patient tissues and used them to train Keras based Convolutional Neural Network (CNN) to be able to predict if a particular histology image is benign or malignant. We have established that the CNN model has done a good job in predicting the classification of histology images with 88% accuracy score. We believe using the additional training data and allowing more time for training the model would yield even better prediction accuracy.
- Computer vision and CNN models have already made their way into the healthcare domain to support the medical imaging, it is interesting to learn that CNN models are also being used for working with 3D images as well as AR (Augmented Reality), these studies would open much more opportunities to improve healthcare industry in the future.

References

1. Dataset - https://www.kaggle.com/paultimothymooney/breast-histopathology-images
2. Code - https://github.com/chandu85/data-science/blob/main/Project%202%20-%20Breast%20Cancer%20Classification%20using%20Computer%20Vision/Code/Breast_Cancer_Classification.ipynb
3. Doru Paul; Differences Between a Malignant and Benign Tumor. Retrieved May 9, 2021, from https://www.verywellhealth.com/what-does-malignant-and-benign-mean-514240
4. F. Milletari, N. Navab and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79. Retrieved May 9, 2021 from https://arxiv.org/pdf/1606.04797.pdf
5. Junfeng Gao, Yong Yang, Pan Lin, Dong Sun Park, "Computer Vision in Healthcare Applications", Journal of Healthcare Engineering, vol. 2018, Article ID 5157020, 4 pages, 2018. https://doi.org/10.1155/2018/5157020. Retrieved May 9, 2021 from https://www.hindawi.com/journals/jhe/2018/5157020/
6. Nadim Mahmud, Jonah Cohen, Kleovoulos Tsourides, Tyler M. Berzin, Computer vision and augmented reality in gastrointestinal endoscopy, Gastroenterology Report, Volume 3, Issue 3, August 2015, Pages 179–184, https://doi.org/10.1093/gastro/gov027. Retrieved May 9, 2021 from https://academic.oup.com/gastro/article/3/3/179/613495
7. Esteva, A., Chou, K., Yeung, S. et al. Deep learning-enabled medical computer vision. npj Digit. Med. 4, 5 (2021). https://doi.org/10.1038/s41746-020-00376-2. Retrieved May 9, 2021 from https://www.nature.com/articles/s41746-020-00376-2
8. J. Thevenot, M. B. López and A. Hadid, "A Survey on Computer Vision for Assistive Medical Diagnosis From Faces," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1497-1511, Sept. 2018, doi: 10.1109/JBHI.2017.2754861. Retrieved May 9, 2021 from https://www.researchgate.net/publication/320250581_A_Survey_on_Computer_Vision_for_Assistive_Medical_Diagnosis_From_Faces
9. Chaohui Wang, Nikos Komodakis, Nikos Paragios. Markov Random Field Modeling, Inference & Learning in Computer Vision & Image Understanding: A Survey. Computer Vision and Image Understanding, Elsevier, 2013, 117 (11), pp.1610-1627. ff10.1016/j.cviu.2013.07.004ff. ffhal-00858390v2f. Retrieved May 9, 2021 from https://hal.archives-ouvertes.fr/hal-00858390/document
10. Tim F. Cootes and Christopher J. Taylor "Statistical models of appearance for medical image analysis and computer vision", Proc. SPIE 4322, Medical Imaging 2001: Image Processing, (3 July 2001); https://doi.org/10.1117/12.431093. Retrieved May 9, 2021 from https://www.spiedigitallibrary.org/conference-proceedings-of-spie/4322/0000/Statistical-models-of-appearance-for-medical-image-analysis-and-computer/10.1117/12.431093.pdf
11. Sultana, F., Sufian, A., & Dutta, P. (2018, November). Advancements in image classification using convolutional neural network. In 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) (pp. 122-129). IEEE. Retrieved May 9, 2021 from https://arxiv.org/pdf/1905.03288.pdf
12. O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2019, April). Deep learning vs. traditional computer vision. In Science and Information Conference (pp. 128-144). Springer, Cham. Retrieved May 9, 2021 from https://arxiv.org/pdf/1910.13796.pdf
13. Data-flair-training python projects. Project in Python – Breast Cancer Classification with Deep Learning. Retrieved May 9, 2021 from https://data-flair.training/blogs/project-in-python-breast-cancer-classification/
14. Breast cancer statistics 2020. Retrieved May 9, 2021 from https://www.nationalbreastcancer.org/wp-content/uploads/2020-Breast-Cancer-Stats.pdf
15. What is Breast Cancer? Retrieved May 9, 2021 from https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html
16. United States Cancer Statistics. Retrieved May 9, 2021 from https://gis.cdc.gov/Cancer/USCS/DataViz.html
17. Invasive Ductal Carcinoma (IDC). Retrieved May 9, 2021 from https://www.breastcancer.org/symptoms/types/idc