Breast Cancer Detection : CNN based approach

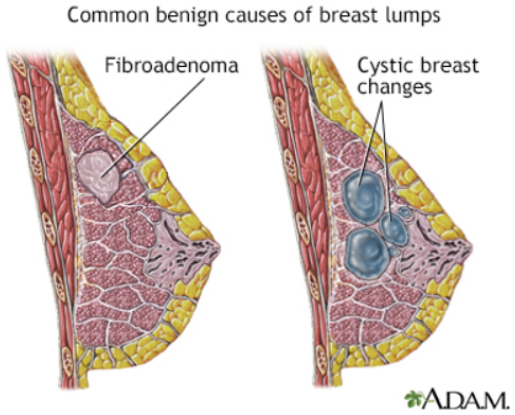
Ravina Lad
lad.ra@northeastern.edu

Abstract*— Breast cancer is the second leading cause of death in women.** ***Breast cancer mortality is reduced by early identification and treatment. Because it can detect early breast lumps or calcification regions, mammography plays an important role in breast cancer screening. Breast cancer masses are more difficult to detect in exceptionally dense breast tissue, which is one of the disadvantages of breast mammography. Manual detection of cancer cells involves human error, hence the deep learning modules are applied to obtain better results. Cancer detection can be done by extracting features through convolution neural network. In this paper, I have trained a CNN model and obtained prediction accuracy of up to 96.04%.

Keywords—CNN, mammography, benign, malignant, maxpooling, batch normalization

I. Introduction

Breast cancer affects one out of eight women worldwide. It is a type of cancer that starts when cells begin to grow out of control. Most breast cancers begin in the ducts that carry milk to the nipple (ductal cancers). Breast cancer can spread when the cancer cells get into the blood or lymph system and are carried to other parts of the body. Cancerous breast tumors are detected by a special type of examination, which is screening mammograms. Mammography is the gold standard for detecting early signs of breast cancer, which can help cure the disease during its early stages. Breast cancer tumors occupy only a small portion of the image of the entire breast. For example, FFDM stands for full-field digital mammography image is usually 227 x 227 and the potential cancerous region can occupy 50 X 50 pixels or even less. This explains why we have a large number of data in our dataset that is 24,518 images (227 x 227 pixels). Our goal for the project is to detect if a given patient and its breast patch is cancerous or not. However, incorrect mammography diagnosis is very common and can harm the patient and there is a probability that the patient can go through unnecessary treatments and operations or sometimes, lack of treatment. Hence, I have built a model that can learn to detect cancerous breast tissues on their own and could potentially help to reduce the number of incorrect interpretations and missed cases.



Implementation of this project can help in the following ways:

- Help physicians for early detection to maximize patients' survival rate.
- Minimize the number of "untrained eyes" that are wrong interpretations and increase the accuracy of screening.
- Prevent late treatments as well as unnecessary treatments in case of false positives.

A) Motivation

My friend's mother was recently diagnosed with early stage breast cancer. She had to travel places for the treatment and move to city areas from rural areas. She is completely recovered now and going through radiation therapies every week to make sure the residues of the operated tumor gets burnt out and in future, there remains no scope for the cancer to spread again. Being a woman myself, I think one should take mammogram screen tests once in five years to make sure the breast ducts are normal. Despite decades of research and effort, the accuracy of risk models used in clinical practice remains limited. Pathologists have to go through Mammography test results and to decide whether a patient has IDC (Invase Ductual Carcinoma – most common type of breast cancer). This is a manual and time consuming process. Moreover, the decision depends on the expertise and their equipment. This motivates us to build an efficient model which will detect the probability of masses and chances of breast cancer in patients. This way one would be able to overcome the dependency of pathologists in the places where no experts are available.

B) Approach

This paper introduces the Convolution neural network approach. The CNN [5] architecture has two main types of transformation. The first is convolution, in which pixels are convolved with a filter or kernel. A second important transformation is sub-sampling, which can be of many types (max_pooling, min_pooling and average_pooling) and used as per requirement. The outline of the approach contains image processing, feature extraction and classification.

C) Dataset

The dataset contains mammography with benign and malignant masses. Images in this dataset were first extracted 106 masses images from INbreast dataset, 53 masses images from MIAS dataset, and 2188 masses images DDSM dataset. Then we use data augmentation and contrast-limited adaptive histogram equalization to preprocess our images. After data augmentation, Inbreast dataset has 7632 images, MIAS dataset has 3816 images, DDSM dataset has 13128 images.

II. Ease of Use

A. Abbreviations and Acronyms

- CNN : Convolution Neural Networks
- IDC : Invase Ductual Carcinoma
- MIAS : Mammographic Imaging Analysis Society
- DDSM : Digital dataset for screening Mammography
- MLE : Maximum Likelihood Estimation.

III. Background

A. Current state-of-art

A group of researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) have built a predictive analytics model that can forecast a patient's breast cancer risk over multiple time periods. However, this model is currently being developed. Clinicians from Novant Health in North Carolina, Emory in Georgia, Maccabi in Israel, TecSalud in Mexico, Apollo in India, and Barretos in Brazil invested in this technology for detection. This model takes into account a variety of parameters such as age, hormones, genetics, and breast density to evaluate the need for further testing and how frequently women should be screened. Despite decades of research and effort, the accuracy of risk models used in clinical practice remains limited, according to relevant published data. While predictive analytics and other AI methods have shown promise in predicting cancer risk, they frequently perform poorly in new patient populations, finding masses in dense breast tissue can be difficult.

B. Literature survey

There are various methods and manual machine learning networks proposed by various scholars to classify breast tissues. Machine learning networks are dependent on MLE [3] (Maximum likelihood Estimation) and GRU-SVM [4] (gated recurrent unit)(State vector machine) model. The above network setup has obtained an accuracy ranging from 80 to 85%. Many used pre-trained models using VGG16 and RESNET50 weights and resultant accuracy was above 95%.

IV. Approach

This paper assesses convolution neural network architecture for detecting breast cancer. I have used labeled data benign and malignant input images. The CNN [6] was trained using 2480 benign and 5429 malignant samples belonging to the RGB color model. CNN was trained separately for MIAS and INBreast data sets.

A. Image Processing

Most of the pixels in the image are redundant and do not contribute substantially to the information of an image [7]. It is required to eliminate them to avoid unnecessary computational overhead. This can be achieved by compression techniques. This is necessary to remove redundancy from the input data which only contributes to the computational complexity of the network.

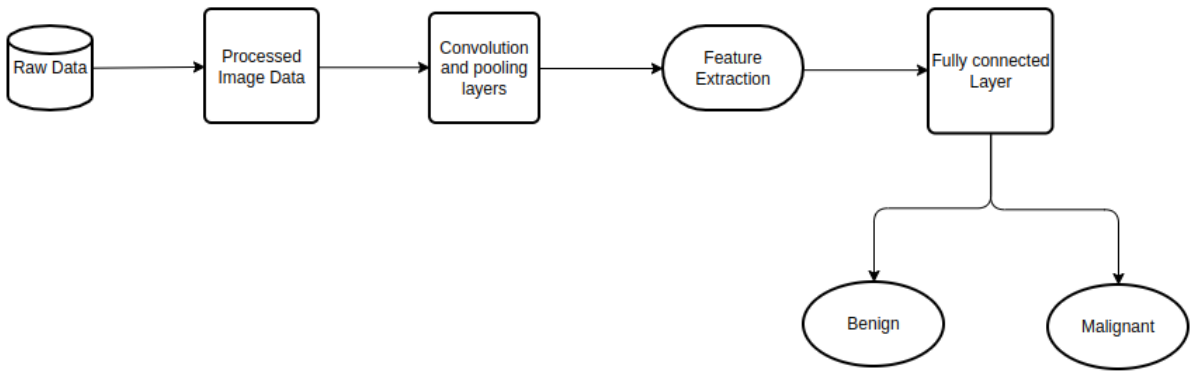
In addition, image augmentation is performed on image datasets in order to avoid overfitting of the CNN model. Data augmentation is performed so that the model will have more images to learn. Different types of transformations done on the image datasets like rotation, reshaping, shifting, shearing zooming. This will generate images with various perturbations.

B. Feature extraction

Feature extraction is a crucial part in neural networks as we have to process the patterns and textures [8] of the images. Malignant tumors tend to have large and irregular nuclei or multiple nuclear structures. In the case of supervised learning, images are fed to an architecture such as CNN, along with its class as a label (Benign or Malignant). From filter values in the training process, CNN is able to extract the computational features. In short, for a given architecture of CNN filters and their weights, are features that are used at the time of testing for model evaluation. In this approach, CNN takes raw pixels of an image and gives output as learned filter weights.

For the final prediction, these weights serve input to the dense architecture for the deep neural network. This paper contains the CNN model which has 2 types of hidden layers combination of convolution layer and pooling layer.
The details of the network are illustrated in Table 1 .

When the convolution process between images and kernels is done, the depth of input is increased by the number of filters used, and when the pooling layer is applied, depth remains the same and size is reduced. The architecture of my model is initiated by a sequential model. The first two layers are defined as : Convolution - Max pooling – ReLu activation function. Then the output layer uses softmax activation function. The last step is classification, which places an image into the respective class (benign or malignant) and can be done using (support vector machine) or with a fully connected layer using an activation function such as Softmax.



C. *Classification*

The process of classification is done by taking the flattened weighted feature map obtained from the final pooling layer, and is used as input to the fully connected network. The fully connected network will calculate the loss and modify the weights of the internal hidden nodes. I have defined the layer attributes in tabular form that can be found in Table 2.

In the classification layer, pre-processed data is given to layers mentioned in Table 2 and the output of the last layer is taken as the final output. Last layer will have 2 numbers of nodes as outputs that are benign and malignant classification. Softmax will normalize the CNN output to fit between 0 and 1. Softmax is used in case of multi-classification problems. The model is compiled with adam optimizer and loss as categorical cross entropy. As we have two classes to categorize the image data, I have also tried to use other optimizers like SGD (Stochastic gradient descent) and RMSprop, but there was no improvement in the performance of the model.

Table : 1 - Parameters of CNN architecture

Layer 1	Layer 1	Layer 1	Layer 1	Layer 4	Layer 5	Layer 6
Type	Conv	pool	conv	pool	conv	pool
Channel	32	-	64	-	128	-
Filter	5 x 5	-	5 x 5	-	5 x 5	-
Conv. stride	1 x 1	-	1 x 1	-	1 x 1	-
Pooling size	-	3 x 3	-	3 x 3	-	3 x 3
Pool stride	-	1 x 1	-	1 x 1	-	1 x 1
Activation	ReLu	-	ReLu	-	ReLu	-

Table : 2 – Details of Fully Connected Network

Layer Attribute	FC-1	FC-2	FC-3
Number of Nodes	64	64	2
Activation Function	ReLu	Relu	Softmax

Outline of Approach -

1. Load the datasets Inbreast and MIAS.
2. Visualize the data set and preprocess the raw images using transformation and data augmentation techniques.
3. Design and train the breast cancer classification problem. Save the model.
4. Test on never seen before data and make the predictions using a loaded saved model.

V. Experimental Results

A. *Datasets*

The dataset to be used in this project is put together by Mendeley Data. It is free to use and open source. The dataset contains mammography with benign and malignant masses. These include below datasets:

1. INbreast dataset [1]
Dataset Images : 7632
Benign : 2520 Malignant: 5112
2. MIAS (Mammographic Imaging Analysis Society) [2]
Dataset Images : 3816
Benign : 2376 Malignant: 1140
3. DDSM (Digital Database for Screening Mammography)
Dataset Images : 13128
Benign : 5742 Malignant: 6576

All the images were resized to 227*227 pixels.

B. *Experiments and Performance Evaluation*

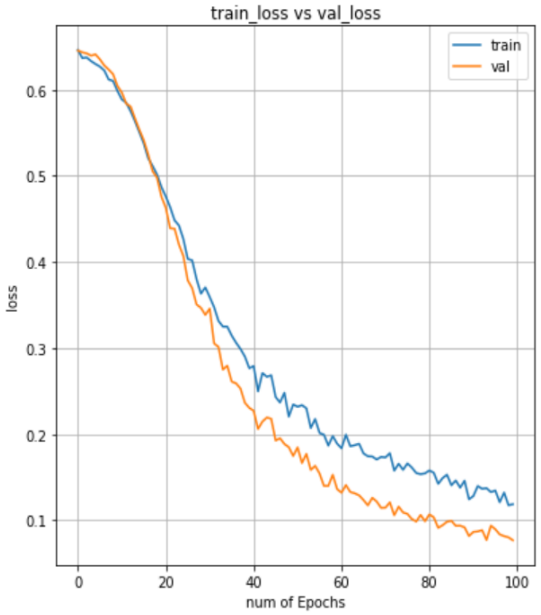
By considering the previously described setup we have obtained a training accuracy of 92.00% with a test train split of 0.2 for MIAS dataset.

As expected for a network, both losses shown Figure 2 and Figure 3 start with a high value and decrease while training proceeds. This behavior is similar to the standard training procedure for deep learning. The difference in saturation levels of Training Loss and Validation Loss is 0.2, which is within the permissible range for a network to avoid underfitting or overfitting.

A. INBreast Dataset

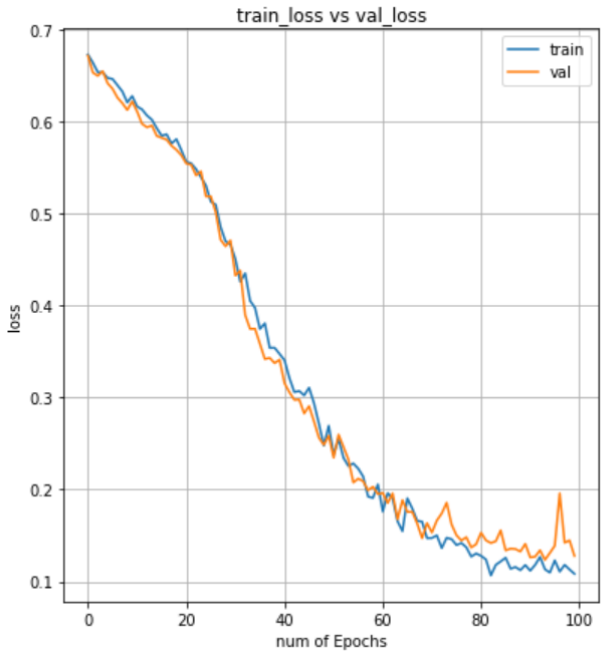
Figure: 2 - Training loss and validation loss.

B.



MIAS Dataset

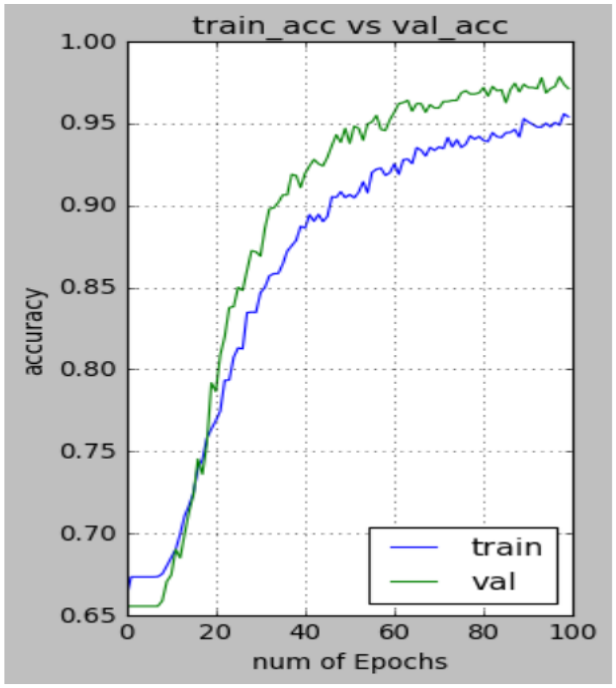
Figure: 3 - Training loss and validation loss.



The graphical plot of accuracy distribution is presented in figure 4 and figure 5. Accuracy starts to increase with the number of epochs, and ultimately saturates, which shows that the training on the dataset is completed for the designed network. Moreover, an important conclusion from this graph is that the network is trained without having characteristics of underfitting and overfitting, as validation accuracy and training accuracy curves are similar in distribution.

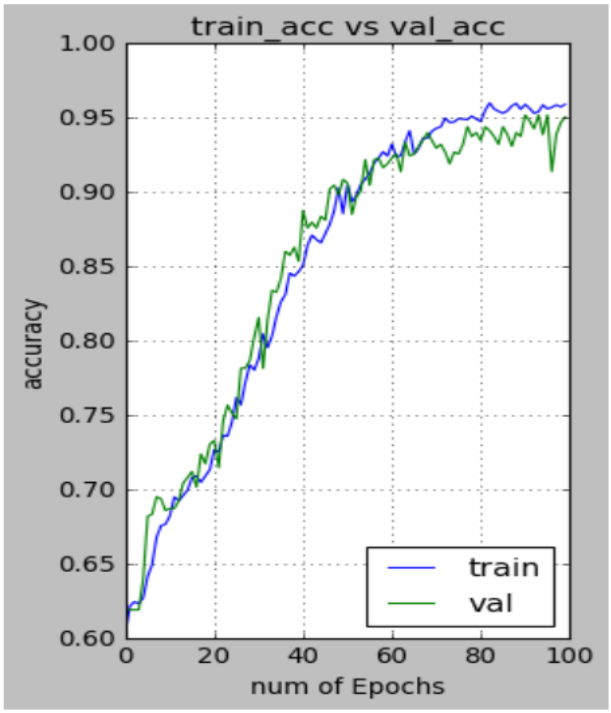
A. INBreast Dataset

Figure: 4 - Training accuracy and validation accuracy.



B. MIAS Dataset

Figure: 5 - Training accuracy and validation accuracy.



Prediction results on test data:

A. Inbreast Dataset

```
# Final evaluation of the model
scores = model.evaluate(X_test, y_test, verbose=1,batch_size=batchsize)
```

```
12/12 [==============================] - 3s 271ms/step - loss: 0.0767 - accuracy: 0.9712
```

B. MIAS Dataset

```
# Final evaluation of the model
scores = model.evaluate(X_test, y_test, verbose=1,batch_size=batchsize)
```

```
6/6 [==============================] - 2s 401ms/step - loss: 0.2450 - accuracy: 0.9202
```

Testing on any random image :

```
***** TESTING ANY RANDOM IMAGE *****

Shape of test image 2:
(1, 92, 140, 3)
Predicted accuracies:
[[0.01409874 0.9859013 ]]
Predicted class:
[1]
```

As you can see, I have provided an unknown image to the loaded trained model and it predicted the results correctly. The probability of the image being malignant is 98.59 %. Hence, the predicted class is malignant, that is 1.

Classification Report :

1. Precision is a probabilistic measure to determine whether a positive case, defined on our terms, actually belongs to the positive class.

2. A recall is a probabilistic measure to determine if an actual positive case is correctly classified with the positive class.

3. The F1 score is calculated as the geometric mean between precision and recall.

Formula :

*F1=2(precision * recall) / (precision+recall)*

1. Classification Report : MIAS Dataset

	precision	recall	f1-score	support
0	0.92	0.95	0.94	467
1	0.92	0.87	0.89	285
accuracy			0.92	752
macro avg	0.92	0.91	0.91	752
weighted avg	0.92	0.92	0.92	752

2. Classification Report : INbreast Dataset

	precision	recall	f1-score	support
0	0.95	0.98	0.96	526
1	0.99	0.97	0.98	1001
accuracy			0.97	1527
macro avg	0.97	0.98	0.97	1527
weighted avg	0.97	0.97	0.97	1527

C. *Discussion*

The accuracy obtained above was improved over many state-of-art experiments. I have tested our model with various resolutions of images and the results are relatively insensitive to resolution. By using this automated process there is a possibility of detection of cancer in the early stages, which can ultimately increase survival rate among breast cancer patients. For comparison, we have compared our result (92.02% validation accuracy from test set)with several published studies shown in Table 3

Table : 3 - Existing methods and respective Accuracy

	A	B	C	D
1	Year	Method Used	Accuracy(%)	Error Rate
2	2017	K-Nearest Neighbor [12]	83 to 86	19.28
3	2019	Pre-Trained Networks [10]	90 to 97	4.74
4	2017	Feature Extracted Using CNN	83 to 90	4.28
5	2018	Deep Convolution Neural Network [11]	91.54	8.54

VI. Conclusion and future scope

In the field of medical pathology, detection of breast cancer using digital images was a milestone. This is the reason it has opened doors to opportunities for research. Hence, now we can use tools like deep learning to solve such problems. In the future scope, one can implement an auto encoder instead of manually reducing image size. It can compress data without losing the prominent features, because autoencoders can regenerate up to 90% of the original image. We can also obtain higher accuracy by changing the model architecture and hyperparameters. to parentheses. This paper was able to classify breast cancer tissues into two classes benign and malignant with accuracy of 92.02 % for MIAS and 97.45% accuracy for INBreast datasets.

VII. Acknowledgment

I am highly grateful to Professor Jerome Braun for the guidance and support throughout the project. The course material was very useful and came absolutely handy while working on the breast cancer detection project. I would also like to thank my parents and roommates for their constant support.

VIII. References

[1] http://peipa.essex.ac.uk/info/mias.html

[2] A. Alias, B. Paulchamy, Detection of breast cancer using artificial neural network, International Journal of Innovative Research in Science 3 (3).

[3] A.F. Agarap On breast cancer detection: an application of machine learning algorithms on the Wisconsin diagnostic dataset, CoRR abs/1711.07831

[4] Y. LeCun, Y. Bengio, G. Hinton Deep learning, nature, 521 (7553) (2015), p. 436, 10.1038/nature14539 https://doi.org/10.1038/nature14539

[5] F. Xing, Y. Xie, L. Yang An automatic learning-based framework for robust nucleus segmentation IEEE Trans Med Imaging, 35 (2) (2016), pp. 550-566, 10.1109/TMI.2015.2481436 https://doi.org/10.1109/TMI.2015.2481436

[6] R.C. González, R.E. Woods, S.L. Eddins Digital image processing using MATLAB Pearson (2004)

[7] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F.A. Wichmann, W. Brendel Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness International conference on learning representations (2019) https://openreview.net/forum?id=Bygh9j09KX

[8] M. Sokolova, N. Japkowicz, S. Szpakowicz Beyond accuracy, f-score and ROC: a family of discriminant measures for performance evaluation AI 2006: advances in artificial intelligence, 19th Australian joint conference on artificial intelligence, hobart, Australia, december 4-8, 2006, proceedings (2006), pp. 1015-1021, 10.1007/11941439_114 https://doi.org/10.1007/11941439%5c_114

[9] J. Sun, A. Binder Comparison of deep learning architectures for h&e histopathology images 2017 IEEE conference on big data and analytics (ICBDA), IEEE (2017), pp. 43-48

[10] S.A. Adeshina, A.P. Adedigba, A.A. Adeniyi, A.M. Aibinu Breast cancer histopathology image classification with deep convolutional neural networks 2018 14th international conference on electronics computer and computation (ICECCO), IEEE (2018), pp. 206-212

[11] A.A. Samah, M.F.A. Fauzi, S. Mansor Classification of benign and malignant tumors in histopathology images 2017 IEEE international conference on signal and image processing applications, ICSIPA 2017, kuching, Malaysia, september 12-14, 2017 (2017), pp. 102-106, 10.1109/ICSIPA.2017.8120587 https://doi.org/10.1109/ICSIPA.2017.8120587

[12] M. Sokolova, N. Japkowicz, S. Szpakowicz Beyond accuracy, f-score and ROC: a family of discriminant measures for performance evaluation AI 2006: advances in artificial intelligence, 19th Australian joint conference on artificial intelligence, hobart, Australia, december 4-8, 2006, proceedings (2006), pp. 1015-1021, 10.1007/11941439_114 https://doi.org/10.1007/11941439%5c_114