*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

# An Augmentation in the Diagnostic Potency of Breast Cancer through A Deep Learning Cloud-Based A.I. Framework to Compute Tumor Malignancy & Risk

**Om Agarwal**

Grade 10 Student @ Citadel High School, Halifax, Nova Scotia, Canada

***Abstract* - This research project focuses on developing a web-based multi-platform solution for augmenting prognostic strategies to diagnose breast cancer (BC), from a variety of different tests, including histology, mammography, cytopathology, and fine-needle aspiration cytology, all in an automated fashion. The respective application utilizes tensor-based data representations and deep learning architectural algorithms, to produce optimized models for the prediction of novel instances against each of these medical tests. This system has been designed in a way that all of its computation can be integrated seamlessly into a clinical setting, without posing any disruption to a clinician's productivity or workflow, but rather an enhancement of their capabilities. This software can make the diagnostic process automated, standardized, faster, and even more accurate than current benchmarks achieved by both pathologists, and radiologists, which makes it invaluable from a clinical standpoint to make well-informed diagnostic decisions with nominal resources.**

***Purpose:*** To develop an application for predicting breast cancer malignancy, along with its risk factors, from multiple diagnostic procedures through the integrated utilization of cutting-edge deep learning models.

***Output:*** The efficiency of the AI models embedded in the system proves that this technology can be used to predict breast cancer from a variety of sources, at the level of experienced individuals in the field in a clinically-supportive channel.

***Keywords:*** Breast Cancer, Early Detection, Improved Diagnostics, Deep Learning, AI, Histology, Cytopathology, Mammography, Fine-Needle Aspiration Cytology.

## I. BACKGROUND RESEARCH

BC is one of the most common cancers among women worldwide. It accounts for 25% of all cancer cases, making it a significant public health problem in today's society [1, 4, 5]. Complementarily, however, according to the World Health Organization, there is a global shortage of radiologists, with over 5 billion people in the world having little to no access to radiology services for proper diagnosis [2]. This alone motivates the need for software that can diagnose BC at the standard of these experts, to minimize consequences posed by such shortage. Additionally, depending on one's location, both pathologists, and radiologists take anywhere from 1-16 weeks to finalize their diagnostic conclusions, which is precious time of the patient which could be used to get started on therapeutic endeavors if BC is present. Countless researches have shown that the early diagnosis of BC can improve the prognosis and chances of survival significantly, as it can promote timely clinical treatment to patients [6, 9].To solve this problem, machine learning models have been applied, in hopes of trying to exploit patterns and relationships among a large number of cases and predict the outcome of BC using historical cases stored in datasets. But in these cases, the performance of most conventional classification systems is dependent on appropriate data representation and much of the efforts are dedicated to feature engineering, a difficult and time-consuming process that uses prior expert domain knowledge of the data to create useful features. However, deep learning, advancement in traditional machine learning, can extract and organize the discriminative information from the data, not requiring the design of feature extractors by a domain expert.

Furthermore, even disregarding the shortages of radiologists, and the time-resources constraints associated with BC diagnosis, there is still a major problem of the diagnostic predictions themselves being quite subjective.

Several studies have shown that pathologist-related diagnostic variabilistic error fares very high for BC, and the primary reason for this, in histology, is the subtle differences of professional opinion on whether the specific morphological features present, meet the diagnostic criteria; these are often related to an individual pathologist's threshold for a particular diagnosis in a specific case [3, 10, 14]. The average general agreement within breast pathologists, as derived by a prominent national study, shows an alarmingly low 68.39% (95% CI) interclass coefficient (ICC) agreement [7, 11].

ISSN (online): 2581-3048
Volume 3, Issue 6, pp 1-24, June-2019

In addition to this, in the case of fine-needle aspirations, countless studies show a pathologist's chance of accurate prediction to be in the ranges of 90-92%, with a specificity of 94-96%, and sensitivity of 93-95%, yielding relatively high false-negative rates of 10-15% [15]. This not only puts people at a huge risk, but also makes pointless a test that is designed to provide accurate results while decreasing patient trauma, expense, and be able to be performed on an outpatient basis.

Also, recently documented reports have deemed that mammographers are very likely to overestimate tumor extent when studying a mammogram and that in general, they have anywhere between 60-75% overall accuracy in detecting BC from a mammogram, and that if they are asked to review the same study as reviewed by them 5 years ago, they display 21-24.5% disagreement, which in itself is huge [12, 18]. This is also very much correlated with the mammographic density parameters, namely, percent-based density (PBD), and breast imaging-reporting and data system (BI-RADS5), which are computed from the mammogram and add crucial weight age to a diagnostic decision –a separate derivation from the report. As closely observed in one European study (which sought to compare individual density predictions with their respective consensus), predictions for both PBD & BI-RADS5 tasks come with a very high root mean squared error of around 17.548, and alarming accuracy of 57.777%, respectively [13].

Despite these unsatisfactory results, both pathologists, and radiologists continue to diagnose patients of malignancy, who, unfortunately, end up suffering the repercussions, namely late and misdiagnosis [16, 17, 24]. This is exactly what I would like to change, using my novel application.

## II. RELATED WORK

Recently, deep learning-based approaches using CNNs have begun to achieve impressive performance on medical tasks such as chest pathology identification in X-Ray and CT, thoraco-abdominal lymph node detection, and interstitial lung disease classification [19, 25]. Currently, in the context of breast masses, the detection of malignancy uses a combination of manual feature extraction and traditional machine learning algorithms [20, 21]. Multiple works tackle the problem of breast lesion classification, but typically adopt a multi-stage approach, by extracting hand-engineered semantic (such as calcification) and textual features, and classify a partial mammogram by extracting features from each view of the breast, and combining them to output a prediction [22]. Not only does this require extensive pre-processing but also heavy clinical domain knowledge, before training a CNN [23]. To the best of my knowledge, this paper presents the first work to directly classify pre-detected breast masses using CNN architectures achieving state-of-the-art results, from more than one diagnostic source in a single application.

## III. DEEP LEARNING MODULE 1: HISTOLOGY

### A. Dataset

In the context of Histopathologic means of diagnosis, I utilized the publicly available BACH dataset for detecting BC malignancy, particularly Invasive (Infiltrating) Ductal Carcinoma (IDC) – which accounts for 80%+ of all BC diagnosis. This dataset can be accessed from here: http://gleason.case.edu/webdata/jpi-dl-tutorial/IDC-regular-ps50-idx5.zip.

This dataset consists of 162 H&E-stained whole mount slide images (RGB-scaled .tiff) of BC specimens scanned at 40x.To assign an aggressiveness grade to a whole mount sample, pathologists typically focus on the regions which contain the IDC, and therefore each image has a corresponding list of labeled coordinates that enclose IDC and Non-IDC regions. In total, there are 277,524 records of x & y coordinates with an associated label, of which 198,738 are IDC negative, and 78,786 are IDC positive.

### B. Preprocessing

As a result of this area-focus labelling, the preprocessing steps for automatic aggressiveness grading first require the delineation of the exact regions of IDC inside of a whole mount slide, which I implemented by looping through each of the patch coordinate parameters, and storing each partition of the image into a separate Numpy array (.npy) file, and its label (after One Hot Encoding into a list) in a new CSV log explicitly linked to the array's ID. In this way, I was able to generate 277,524 patches and resized them all to a unified size of 50*50px. A sample of these preprocessed patches from each class is shown in Fig. 1. I then enforced balanced classes by randomly deleting the difference between the IDC negative and positive images, from the negative, as proved by Fig. 2. Then, after applying feature scaling, for the primary purpose of speeding up the training process, the data is split into training, and testing subsets, with a ratio of 80:20. After the data preprocessing, the training tensor has the shape, (126058, 50, 50, 3), and the testing tensor, (31514, 50, 50, 3), while the output tensor is expected to have the shape (2, 1).

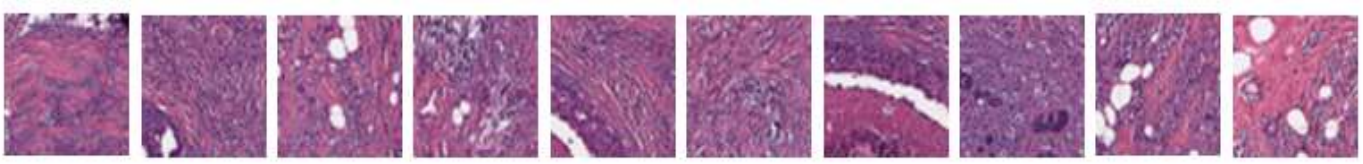
 

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

**IDC (+)**



**Non-IDC (-)**

**Figure 1: Displaying a sample of preprocessed histology image patches of both IDC positive and negative classes**

**Previously Imbalanced** **Now Balanced**

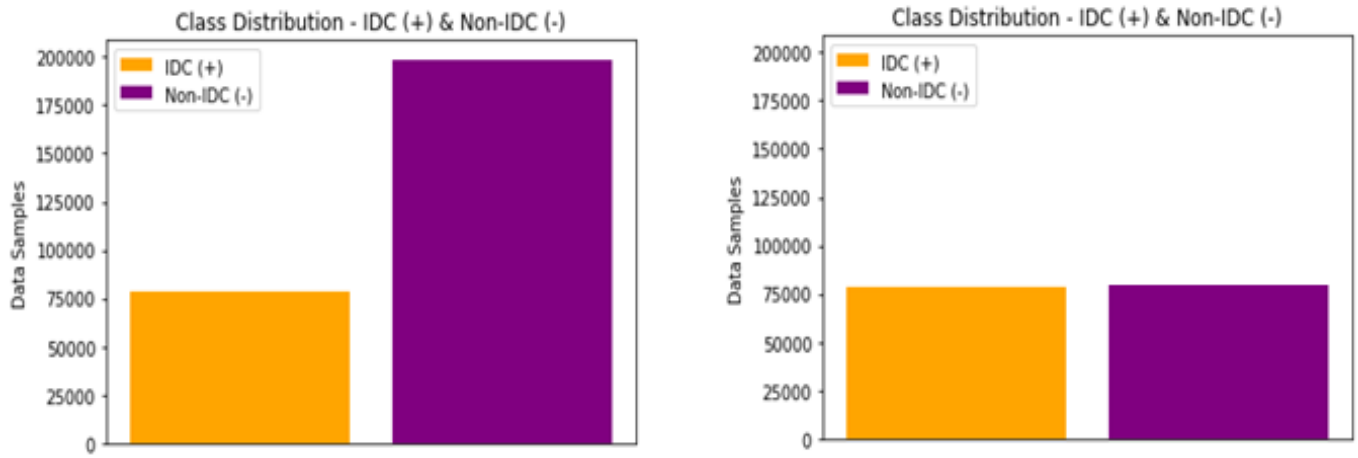



**Figure 2: Displaying the correction of class imbalancement**

### C. Model Structure

In terms of structure, I have defined a custom TensorFlow model which starts by accepting input tensors in batches of 128 images into an initial 2D Convolutional layer. It is then followed up with 3 sets of Convolutions, Max Poolings, Batch Normalizations, Dropouts, and subsequently trailed with several dense layers. All mentioned layers carry a Re Lu activation function, except the last dense layer, which is activated via Soft Max. The gradient descent optimization algorithm used for this is called AdaDelta. A fully graphed structure of the model is displayed in Fig. 3.

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

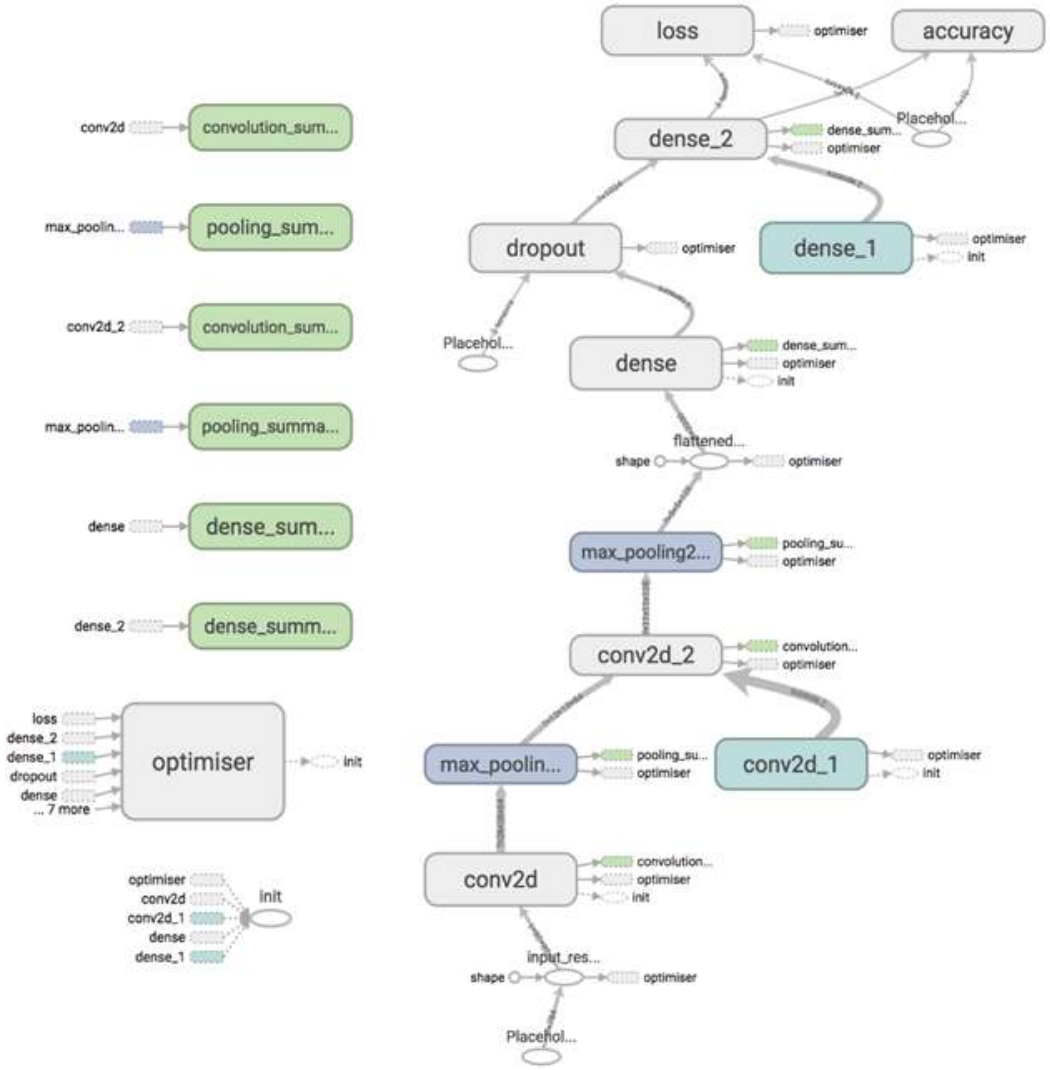



**Figure 3: Displaying model structure**

### D. Model Results

See Fig. 4 for the validation results computed on the test set.

| | |
|---|---|
| **Overall Accuracy** | **84.45458000760168%** |
| **Area Under the Curve** | 84.23874203482222% |
| **Brier Score/Hamming Loss/Zero-One Classification Loss** | 0.1554541999239833 |
| **Average Hinge Loss (Non-Regularized)** | 0.6554541999239833 |
| **Logistic/Cross entropy Loss** | 0.3631732057210411 |
| **Cohen's Kappa (Inter-Annotator Agreement)** | 0.9091600152033333 |
| **Matthews Correlation Coefficient** | 0.6891128602975093 |

**Figure 4: Listing validation data; revealing the results**

Average time to predict BC presence from browser: 18.2 seconds.

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

As can be seen by considering accuracy alone, this model already performs far better than traditional pathologists, who, as stated earlier, diagnose BC from histopathological means with an accuracy of just 68.39%. This is an unprecedented 16.06458% increase. Though there is limited data to compare other metrics, it can be inferred that this positive spike is also prevalent in the other stated losses and agreements.

Additionally, see Fig. 5 for a detailed Classification Report (gauging values of the precision, recall, and f1 score for each class) & Fig. 6 for a Confusion Matrix (computing class-based accuracies), both derived from the validation set. Furthermore, Fig. 7 shows epoch-wise trends for training & testing accuracies & losses.

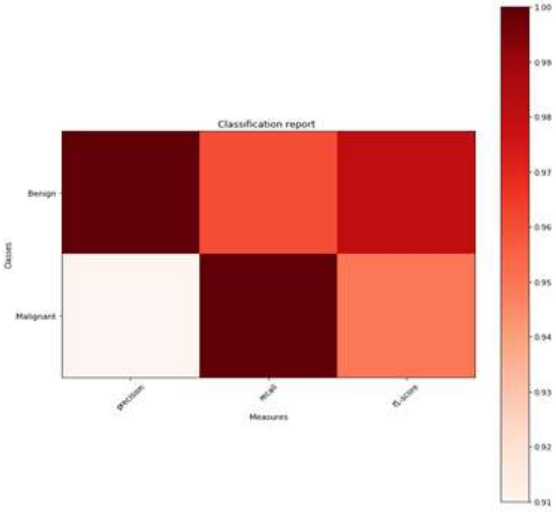



**Figure 5: Graphing classification report**

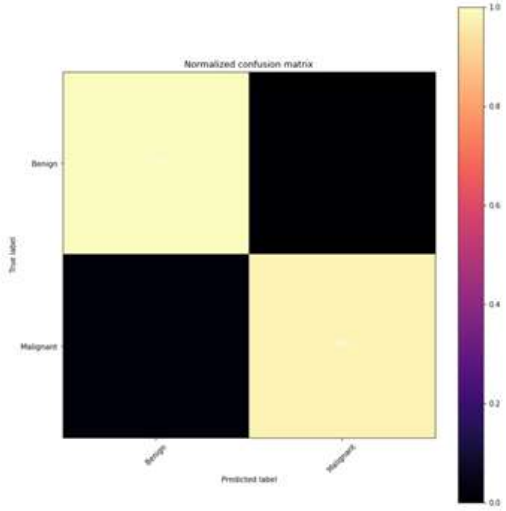



**Figure 6: Graphing confusion matrix**

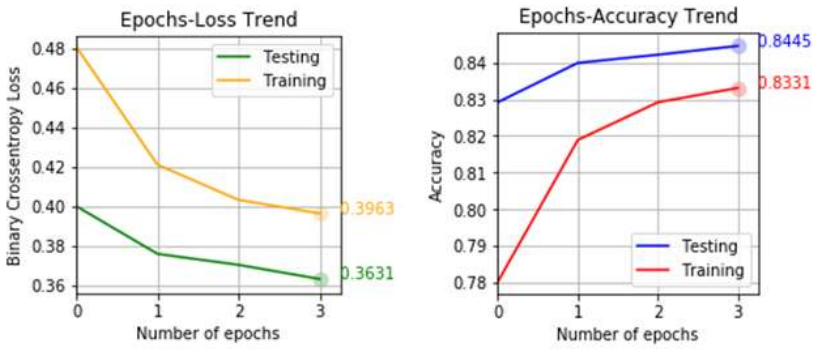



**Figure 7: Plotting epoch-wise trends for training & testing accuracies & losses**

## IV. DEEP LEARNING MODULE 2: MAMMOGRAPHY

### A. Dataset

In the context of Mammographical means of diagnosis, I trained 3 separate models, 2 utilizing the publicly available CBIS1-DDSM dataset for detecting breast density (both percent-based density [PBD], and breast imaging-reporting and data system[BIRADS5] categorical density) from a mammogram, and the other using the publicly available MIAS-M dataset for detecting malignancy from these 2 density parameters. The CBIS1-DDSM dataset can be accessed from here: https://wiki.cancer-imaging-archive.net/display/Public/CBIS-DDSM, and the MIAS-M dataset from here:

https://www.repository.cam.ac.uk/handle/1810/250394.

The CBIS1-DDSM dataset contains 1182 (3328px x 4084px) gray-scaled DICOM (.dcm) mammogram images, with associated PBD & BIRADS5 markings stored in the image format, while the MIAS-M dataset consists of a single CSV file with 936 records, and 2 columns for the 2 density parameters, with the prognostic truth label

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

(MALIGNANT/BENIGN) for each. There are 452 MALIGNANT records, and 484 BENIGN.

### B. Preprocessing

Preprocessing for CBIS1-DDSM dataset is quite extensive. Firstly, each DICOM file must be appropriately read and converted into a pixel array. Then the density parameter labels which are embedded in the image format need to be extracted and stored into a Data Frame separately after One Hot Encoding the BIRADS5 labels and accounting for outliers in the PBD. Further, after implementing feature scaling, the images are resized into 256*256px and saved as Numpy array (.npy) files. A sample of these preprocessed images is shown in Fig. 8.These Numpy arrays are then run through a series of various augmentations, namely, random noise, log/gamma/sigmoid adjustments, rotations/flips, contrasting, exposure, etc., in order to increase the magnitude of the data by more than 3x. The preprocessing required for the MIAS-M dataset is nominal, apart from loading the data into a Pandas Data Frame. All the data is finally split into training, and testing subsets, with a ratio of 80:20. After the data preprocessing, the training tensor for the CBIS1-DDSM dataset has the shape, (2564, 256, 256, 1), and the testing tensor, (641, 256, 256, 1), while the output tensor is expected to have the shape (2, 1) for both models collectively. In the same sense, the training tensor for the MIAS-M dataset is of shape (791, 2), while the rest (145, 2) is for testing.

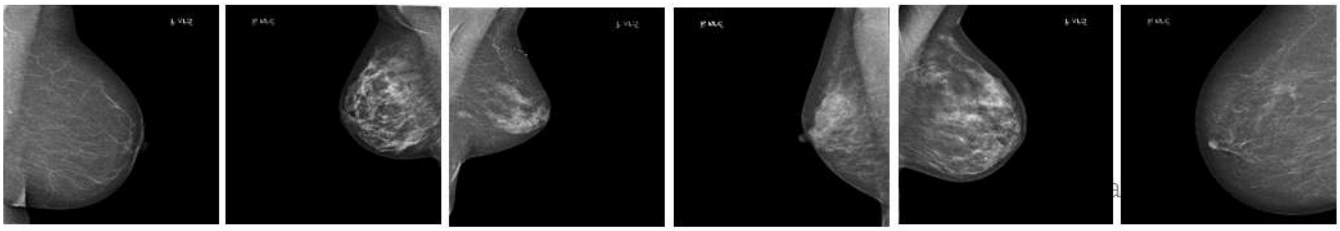
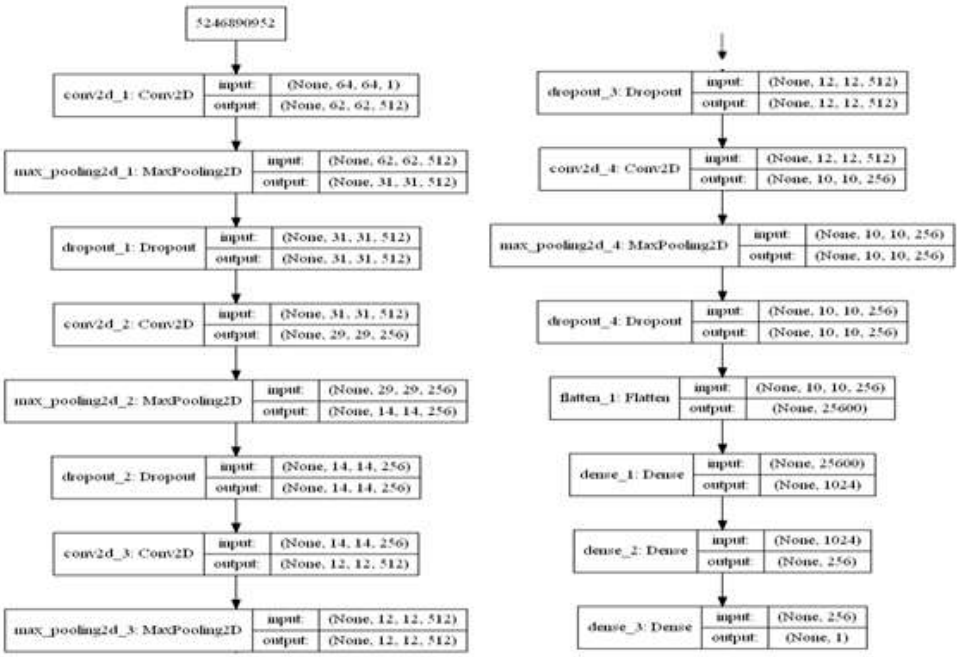


**Figure 8: Displaying a sample of preprocessed mammogram images**

### C. Model Structure

The custom TensorFlow model structure/summary of the PBD& BIRADS5 prediction models can be seen in Fig. 9.Both of these models utilize a TanH activation function for every layer, except the last, where the PBD model uses no activation, and the BIRADS5 model uses Soft Max.

The third model, which is fed these numerical density outputs, actually performs better as a traditional machine learning Random Forest classifier, as opposed to a deep learning classifier, presumably due to limited training data. Fig. 10 graphical displays the central node of the RF Decision Tree.



*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

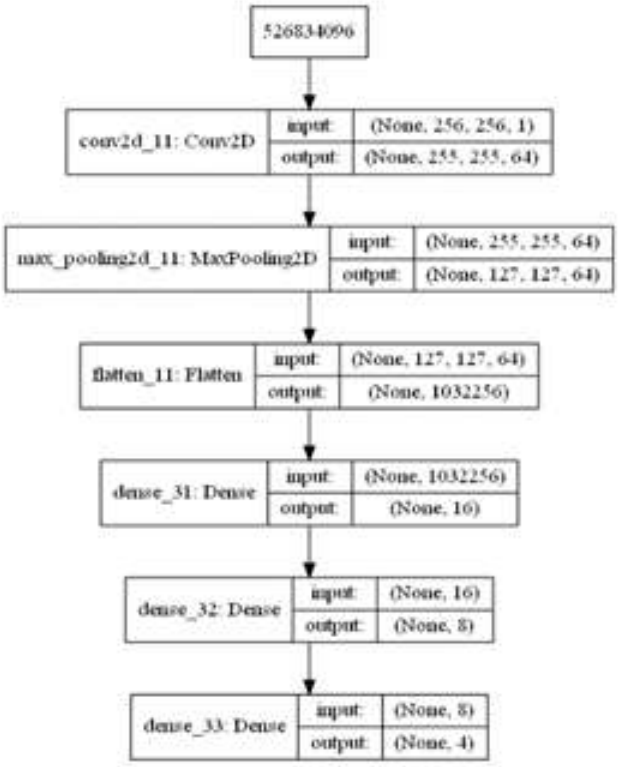



**Figure 9: Displaying PBD & BIRADS5models'structures**

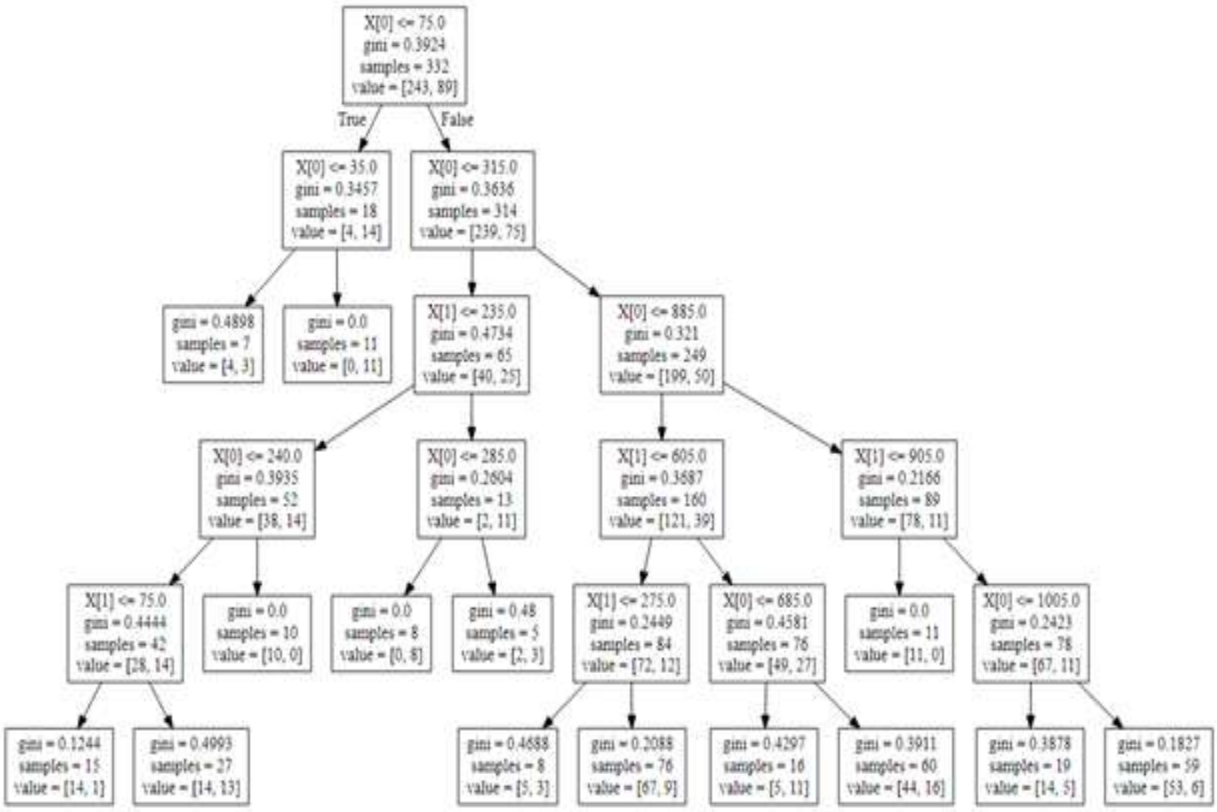



**Figure 10: graphical displays the central node of the RF Decision Tree**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

### D. Models' Results

**** PBD Model***

See Fig. 11 for the validation results computed on the test set.

| | |
|---|---|
| ***Root Mean Squared Error (RMSE)*** | ***5.2321181999999999*** |
| ***Mean Squared Error (MSE) Loss*** | 27.375060712888888 |
| ***Mean Absolute Error (MAE)*** | 3.9111124037821011 |
| ***Median Absolute Error (Med-AE)*** | 4.7013566669285345 |
| ***R2 Score*** | 0.9874539608491023 |
| ***Explained Variance Regression Score*** | 0.9378082300000000 |
| ***Pearson Correlation Coefficient (Pearson's R)*** | 0.9666666666666666 |
| ***Kendall Rank Correlation Coefficient (Kendall's Tau)*** | 0.8277829344387999 |
| ***Spearman's Rank Correlation Coefficient (Spearman's Rho)*** | 0.8591247189612479 |

**Figure 11: Listing validation data; revealing the results**

Average time to predict PBD from browser: 3.4 seconds.

Judging by the RMSE, this model shows an increase of nearly 12.32, which is more than competitive for what's currently available.

Also, see Fig. 12 for a linear comparison of actual values, and predicted values, and see Fig. 13 for a detailed comparison of the residuals and fits. Furthermore, Fig. 14 outlines the training & testing, RMSE & MAE loss trends, as the epochs progress.

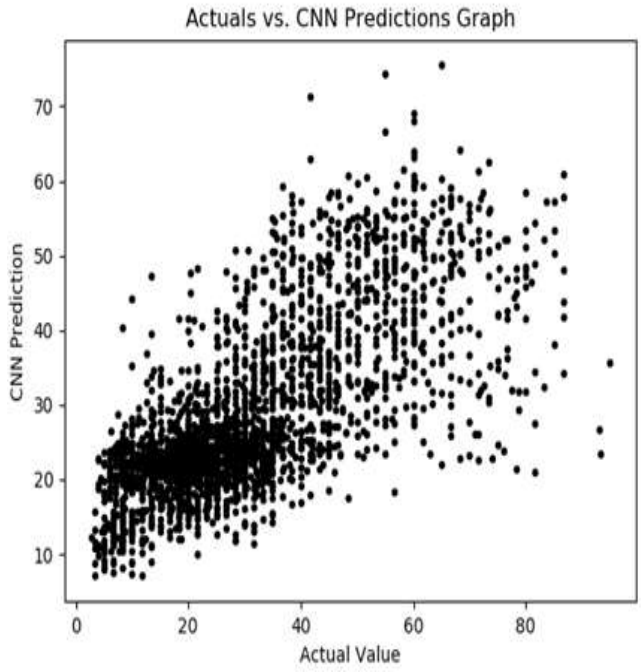



**Figure 12: Graphing actuals vs. CNN predictions**

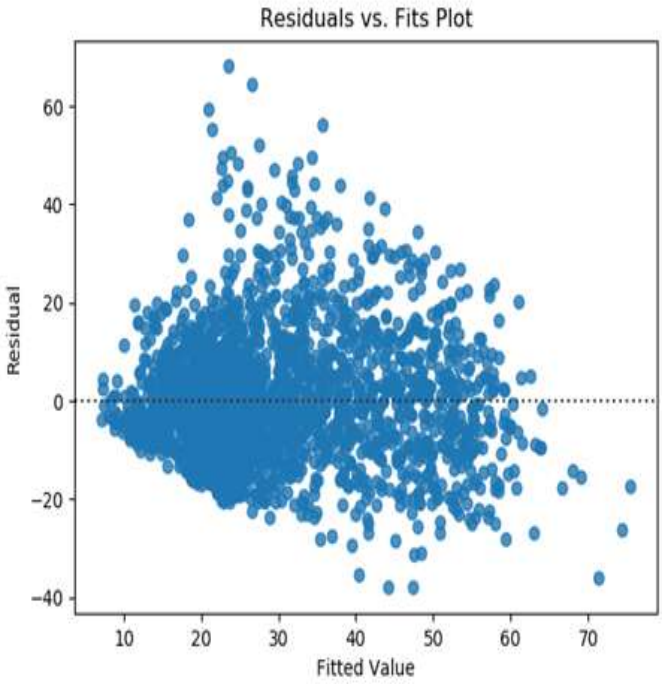



**Figure 13: Graphing residuals vs. fits**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

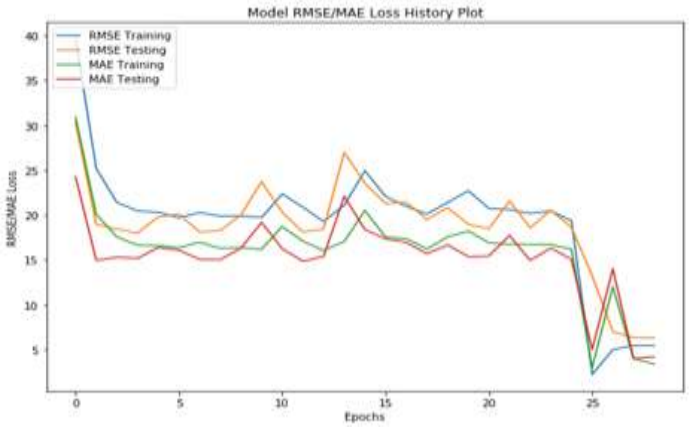



**Figure 14: Plotting RMSE/MAE loss progression history**

**** BIRADS5 Model***

See Fig. 15 for the validation results computed on the test set.

| | |
|---|---|
| ***Overall Accuracy*** | 88.29837599999999% |

**Figure 15: Listing validation data; revealing the results**

Average time to predict BIRADS5 from browser: 1.8 seconds.

Judging by the accuracy, this model shows an increase of nearly 30.52, which is more than competitive for what's currently available.

Additionally, see Fig. 16 for a Confusion Matrix (computing class-based accuracies) derived from the validation set.

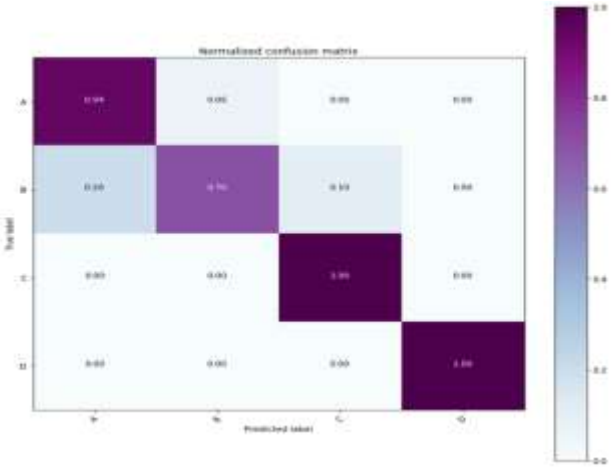


**Figure 16: Graphing confusion matrix**

**** Overall Density -> Malignancy Model***

See Fig. 17 for the validation results computed on the test set.

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

| | |
|---|---|
| ***Overall Accuracy*** | **92.19287411188888%** |
| ***Area Under the Curve*** | 81.22222222222222% |
| ***Brier Score/Hamming Loss/Zero-One Classification Loss*** | 0.1398597151111111 |
| ***Average Hinge Loss (Non-Regularized)*** | 0.5555555555555555 |
| ***Logistic/Crossentropy Loss*** | 0.3279283202222222 |
| ***Cohen's Kappa (Inter-Annotator Agreement)*** | 0.9339725875029388 |
| ***Matthews Correlation Coefficient*** | 0.5347275390870328 |

**Figure 17: Listing validation data; revealing the results**

Average time to predict BC presence from browser: 12.1 seconds.

As can be seen by considering accuracy alone, this model already performs far better than traditional radiologists, who, as stated earlier, diagnose BC from mammographic means with an accuracy of just 60-75%. This is an unprecedented 17.193-32.193% increase. Though there is limited data to compare other metrics, it can be inferred that this positive spike is also prevalent in the other stated losses and agreements.

Additionally, see Fig. 18 for a detailed Classification Report (gauging values of the precision, recall, and f1 score for each class) & Fig. 19 for a Confusion Matrix (computing class-based accuracies), both derived from the validation set. Furthermore, Fig. 20 shows epoch-wise trends for training & testing accuracies & losses.

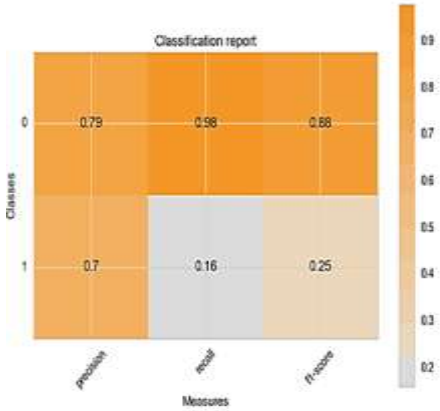



**Figure 18: Graphing classification report**

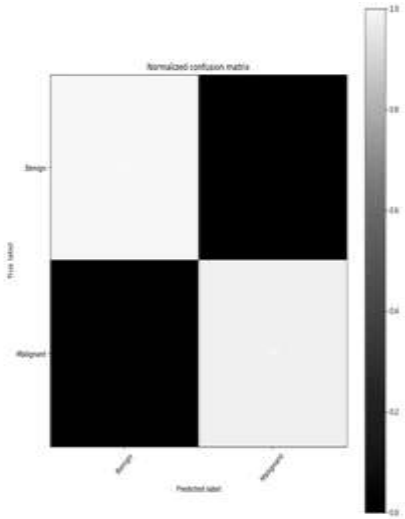


**Figure 19: Graphing confusion matrix**

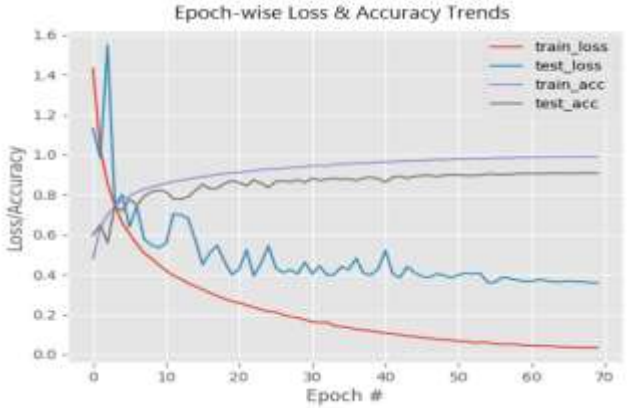



**Figure 20: Plotting epoch-wise trends for training & testing accuracies & losses**

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

## V. DEEP LEARNING MODULE 3: CYTOPATHOLOGY

### A. Dataset

In the context of Cytopathological means of diagnosis, I utilized the publicly available Breast Cancer Wisconsin (Diagnostic)dataset, for detecting BC malignancy, and can be accessed from here:

http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic).

This dataset consists of features which are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the characteristics of cell nuclei present in the image. Aside from the ID, and diagnosis (m = malignant, B = benign), there are 10 real-valued features computed for each cell nucleus, and the mean, standard error, and worst/largest (mean of the 3 largest values) of these features are recorded resulting in a total of 30, as follows,

a) mean, SE, & worst - radius (mean of distances from the center to points on the perimeter)
b) mean, SE, & worst - texture (standard deviation of gray-scale values)
c) mean, SE, & worst - perimeter
d) mean, SE, & worst - area
e) mean, SE, & worst - smoothness (local variation in radius lengths)
f) mean, SE, & worst - compactness (perimeter^2 / area - 1.0)
g) mean, SE, & worst - concavity (severity of concave portions of the contour)
h) mean, SE, & worst - concave points (number of concave portions of the contour)
i) mean, SE, & worst - symmetry
j) mean, SE, & worst -fractal dimension ("coastline approximation" - 1)

There are 569 records in total, of which 357 are benign, and 212 are malignant. Fig. 21 shows the correlations among and between these dimensions and Fig. 22 shows their distribution.

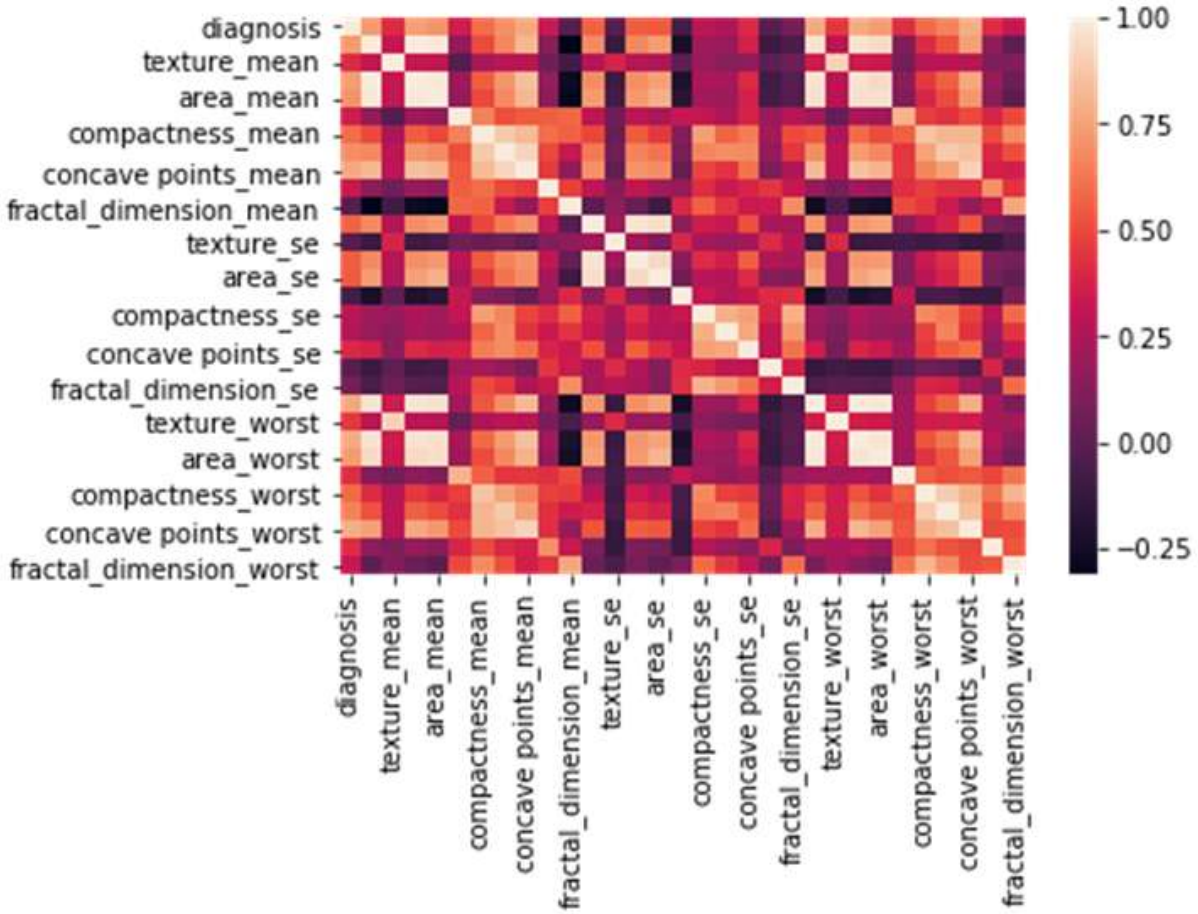



**Figure 21: Showing correlation heat map for the dataset**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

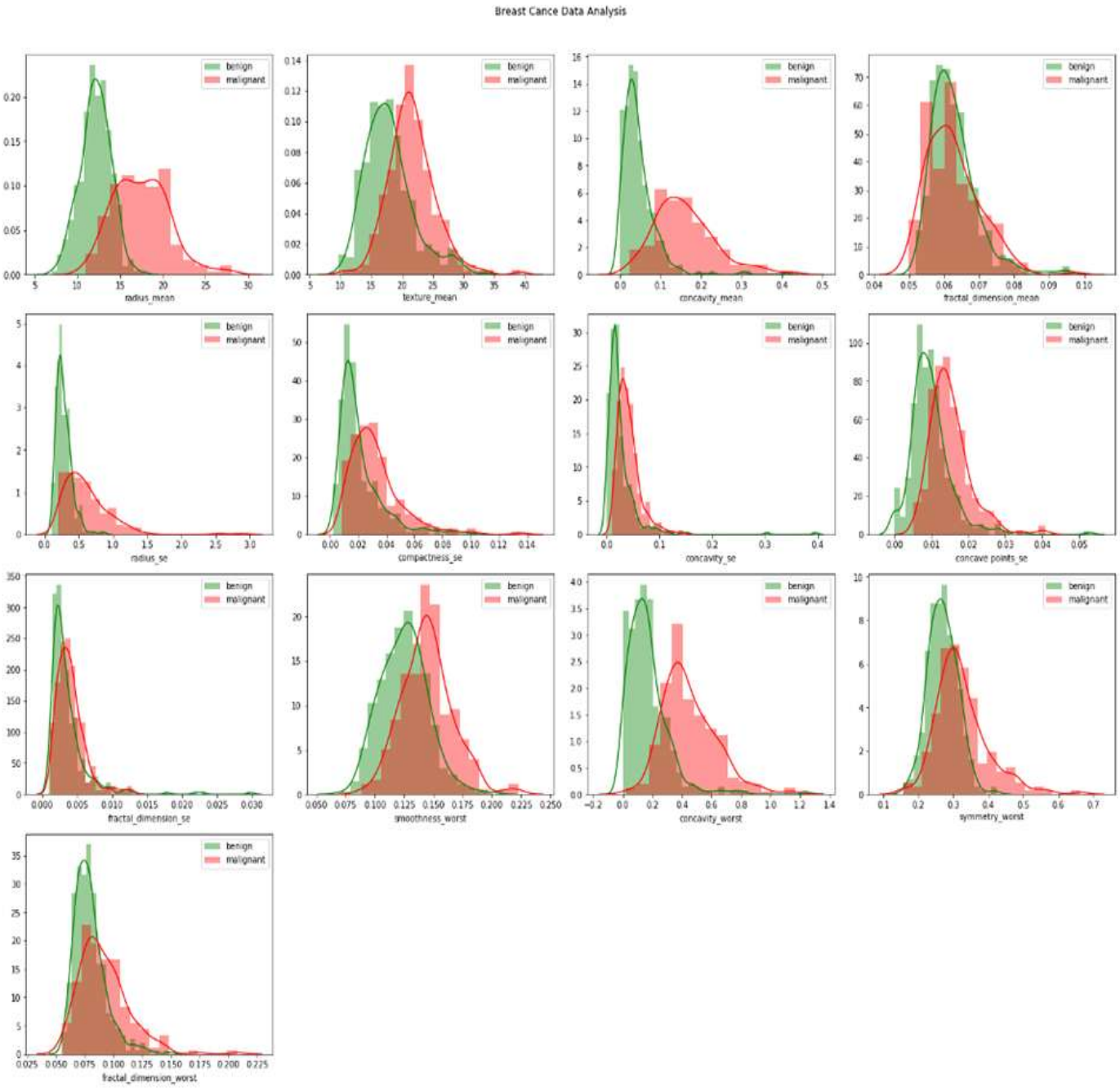



**Figure 22: Showing distribution plots for the selected features**

### B. Preprocessing

The preprocessing required for this dataset is minima land requires loading the data into a Pandas Data Frame, applying feature scaling + normalization, and splitting into training, and testing subsets, with a ratio of 95:5. After the data preprocessing, the training tensor has the shape, (541, 31), and the testing tensor, (28, 31), while the output tensor is expected to have the shape (2, 1).

### C. Model Structure

In terms of structure, I have defined a custom TensorFlow model which starts by accepting input tensors in batches of 32records into an initial Dense layer. It is followed up with 3 sets of Dense layers, all of which carry a ReLu activation function, except the last, which is activated via Sigmoid (causing probability-based predictions/outputs). The gradient descent optimization algorithm used for this is called RMS Prop. A fully graphed structure of the model is displayed in Fig. 23.

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

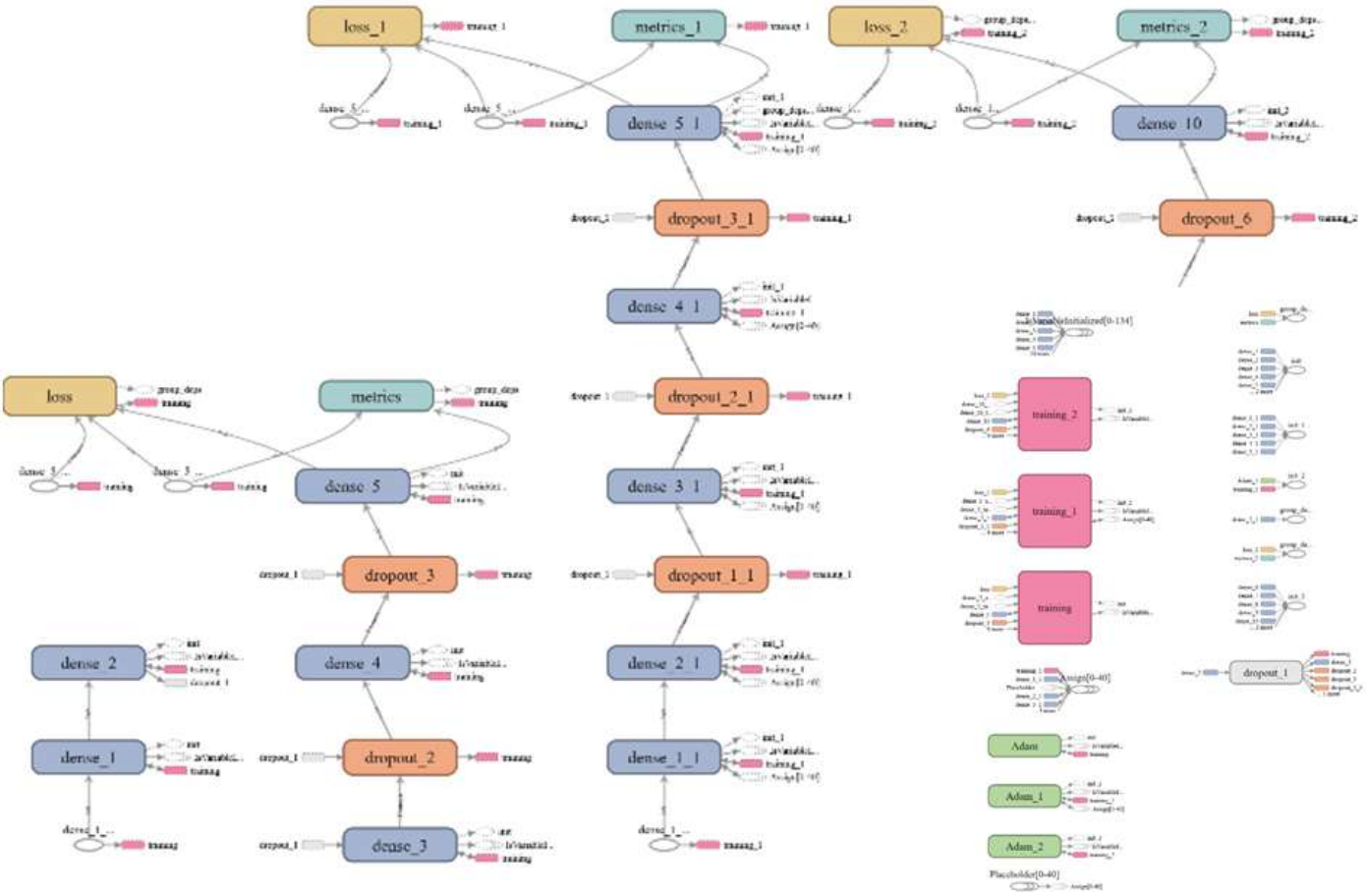



**Figure 23: Displaying model structure**

### D. Model Results

See Fig. 24 for the validation results computed on the test set.

| | |
|---|---|
| ***Overall Accuracy*** | **96.4949494949%** |
| ***Area Under the Curve*** | 92.9361702128% |
| ***Brier Score/Hamming Loss/Zero-One Classification Loss*** | 0.008771929821 |
| ***Average Hinge Loss (Non-Regularized)*** | 0.596491228077 |
| ***Logistic/Crossentropy Loss*** | 0.191881722762 |
| ***Cohen's Kappa (Inter-Annotator Agreement)*** | 0.981841350749 |
| ***Matthews Correlation Coefficient*** | 0.902003265278 |

**Figure 24: Listing validation data, revealing the results**

Average time to predict BC presence from browser: 11.7 seconds.

As can be seen by considering accuracy alone, this model already performs far better than traditional pathologists, who, as stated earlier, diagnose BC from cytopathological means with an accuracy of just 90-92%. This is an unprecedented 4.494-6.494% increase. Though there is limited data to compare other metrics, it can be inferred that this positive spike is also prevalent in the other stated losses and agreements.

Additionally, see Fig. 25 for a detailed Classification Report (gauging values of the precision, recall, and f1 score

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

for each class) & Fig. 26 for a Confusion Matrix (computing class-based accuracies), both derived from the validation set. Since the outputs of this model are probabilistic rather than discrete, a Receiver Operating Characteristic Graph has been plotted, allowing for greater transparency into the model's capability of distinguishing between classes (see Fig. 27). Also view Fig. 28 for an illustrative measure of the valued distribution of the validation predictions. Furthermore, Fig. 29 shows epoch-wise trends for training & testing accuracies & losses.

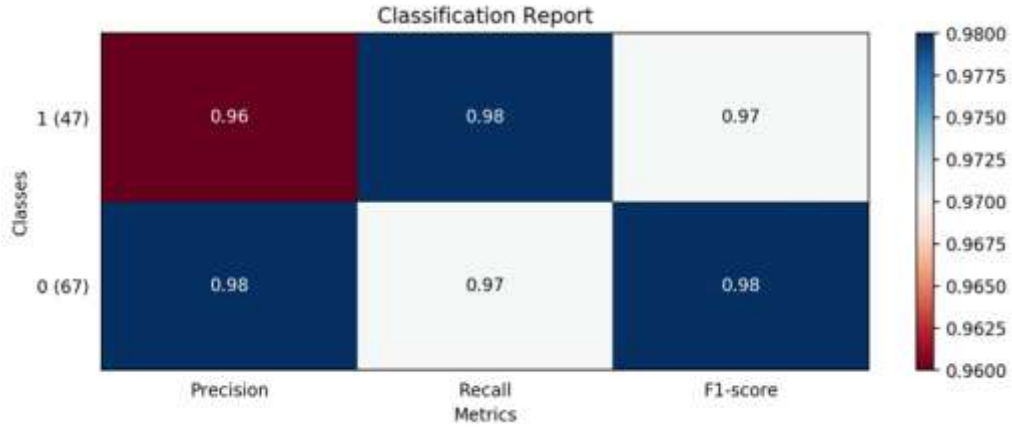



**Figure 25: Graphing classification report**

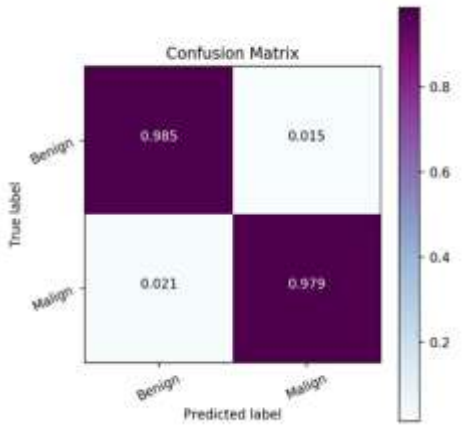



**Figure 26: Graphing confusion matrix**

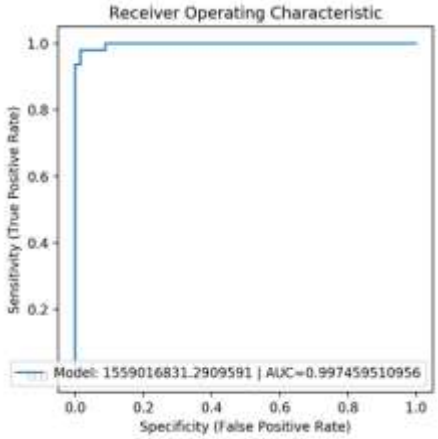



**Figure 27: Graphing ROC AUC Curve**

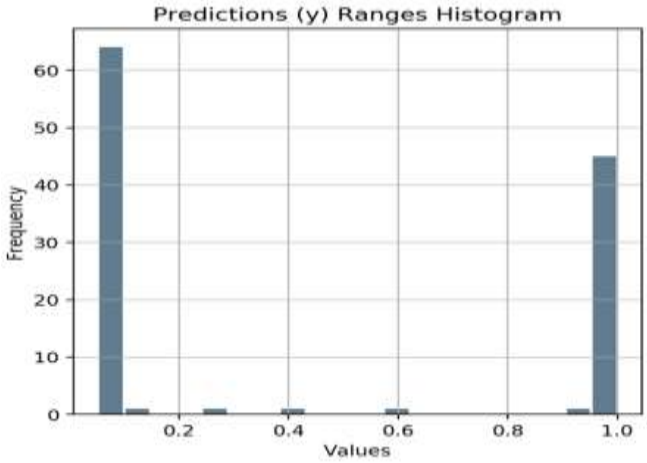



**Figure 28: Graphing predictions probability distribution**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

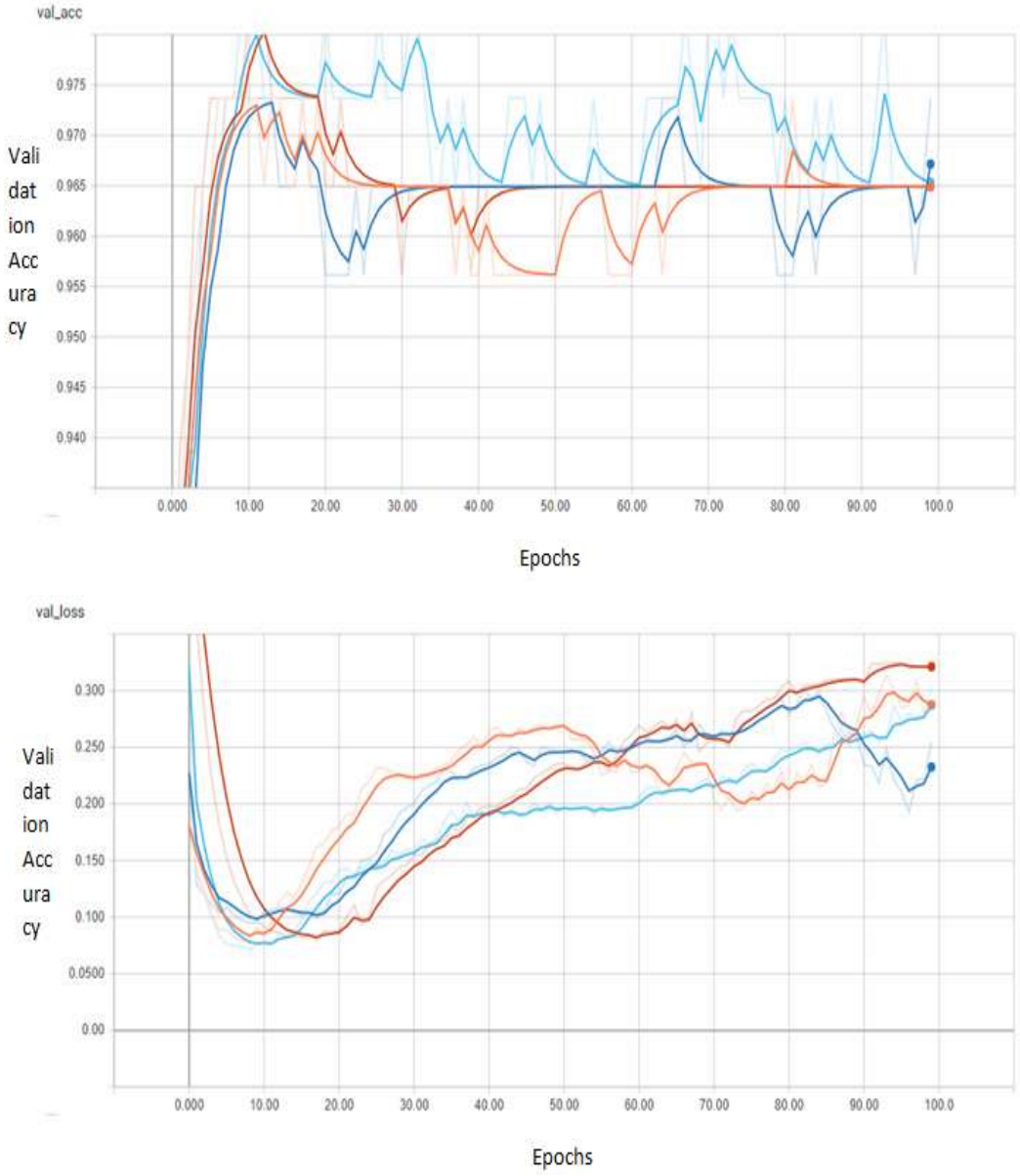



**Figure 29: Plotting epoch-wise trends for training & testing accuracies & losses for several models**

## VI. DEEP LEARNING MODULE 4: FINE-NEEDLE ASPIRATION CYTOLOGY

### A. Dataset

In the context of Fine-Needle Aspiration Cytological means of diagnosis, I utilized the publicly available Breast Cancer Wisconsin (Original) dataset, for detecting BC malignancy, and can be accessed from here: http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original).This dataset consists of pathological parameters which have been recorded of fine needle aspirations of breast masses. Aside from the ID, and diagnosis (m = malignant, B = benign), there are 10 categorical features (encoded 1-10), as follows,

a) clump thickness
b) uniformity of cell size
c) uniformity of cell shape
d) marginal adhesion
e) single epithelial cell size
f) bare nuclei
g) bland chromatin
h) normal nucleoli
i) mitoses

There are 699 records in total, of which 458 are benign, and 241 are malignant. Fig. 30 shows the correlation within these dimensions.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 3, Issue 6, pp 1-24, June-2019

**Figure 30: Showing correlation heatmap for the dataset**

### B. Preprocessing

The preprocessing required for this dataset is minimal, and requires loading the data into a Pandas Data Frame, applying feature scaling + normalization, performing Principal Component Analysis (see Fig. 31) to reduce the dimensions to only the most significant (hence removing single epithelial cell size), and splitting into training, and testing subsets, with a ratio of 95:5. After the data preprocessing, the training tensor has the shape, (664,10), and the testing tensor, (35, 10), while the output tensor is expected to have the shape (2, 1).

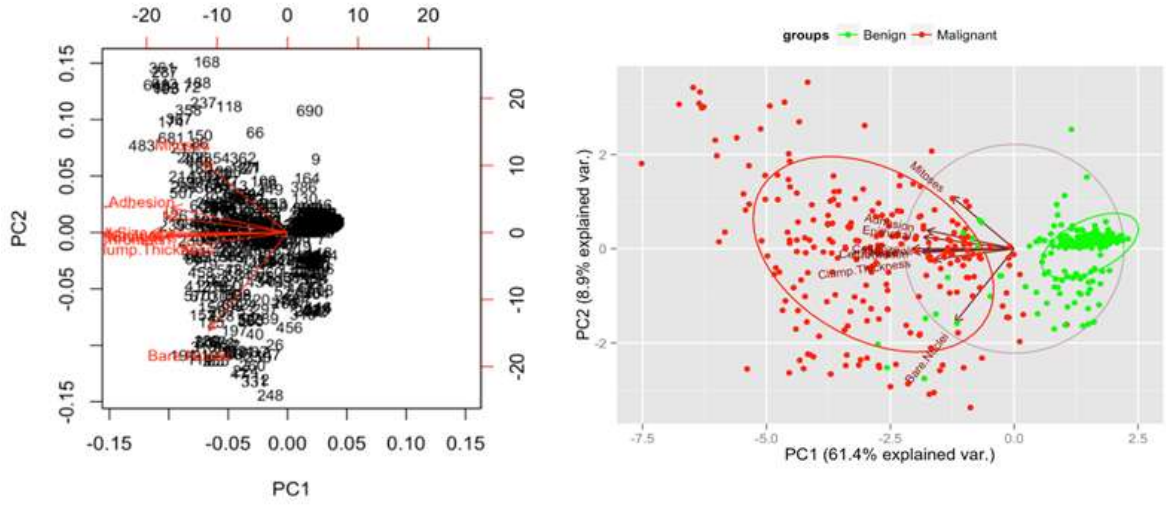



**Figure 31: Showing removal of "single epithelial cell size" through principal component analysis**

### C. Model Structure

In terms of structure, I have defined a custom TensorFlow model which starts by accepting input tensors in batches of 32 records into an initial Dense layer. It is followed up with several Dropouts, and 4 sets of Dense layers, all of which carry a ReLu activation function, except the last, which is activated via Sigmoid(causing probability-based predictions/outputs). The gradient descent optimization algorithm used for this is called Stochastic Gradient Descent (SGD). A fully graphed structure of the model is displayed in Fig. 32.

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

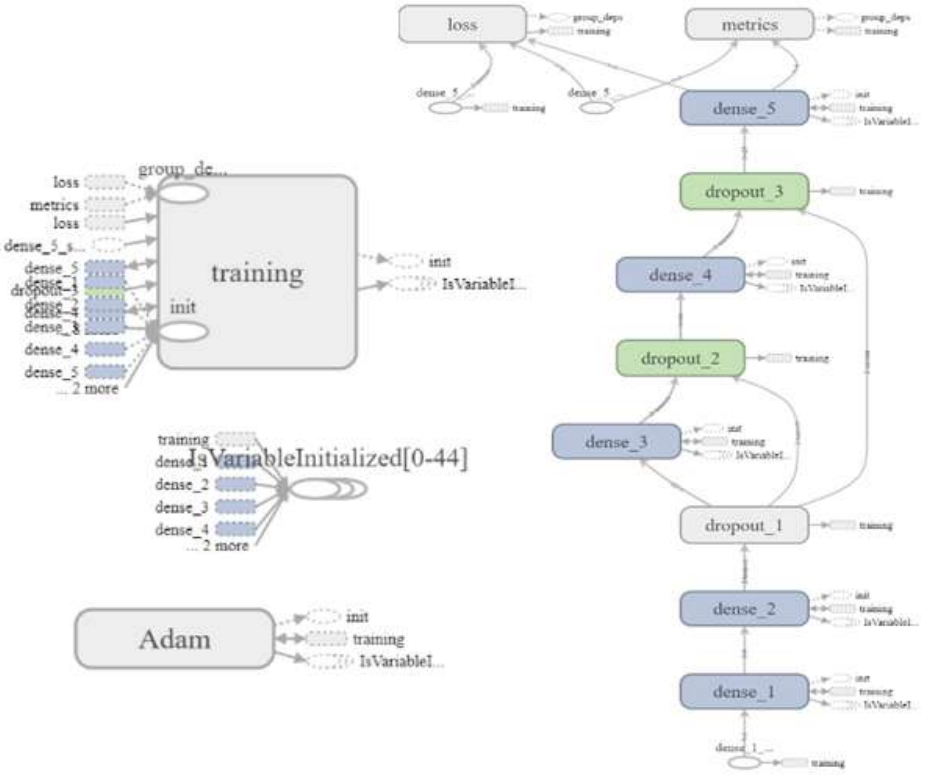



**Figure 32: Displaying model structure**

### D. Model Results

See Fig. 33 for the validation results computed on the test set.

| | |
|---|---|
| ***Overall Accuracy*** | **99.1212121212%** |
| ***Area Under the Curve*** | 98.0934752222% |
| ***Brier Score/Hamming Loss/Zero-One Classification Loss*** | 0.005238083222 |
| ***Average Hinge Loss (Non-Regularized)*** | 0.333829787888 |
| ***Logistic/Crossentropy Loss*** | 0.098982975222 |
| ***Cohen's Kappa (Inter-Annotator Agreement)*** | 0.999325798232 |
| ***Matthews Correlation Coefficient*** | 0.969827530980 |

**Figure 33: Listing validation data, revealing the results**

Average time to predict BC presence from browser: 4.0 seconds.

As can be seen by considering accuracy alone, this model already performs far better than traditional pathologists, who, as stated earlier, diagnose BC from fine-needle aspiration cytological means with an accuracy of just 90-92%. This is an unprecedented 7.121-9.121% increase. Though there is limited data to compare other metrics, it can be inferred that this positive spike is also prevalent in the other stated losses and agreements.

Additionally, see Fig. 34 for a detailed Classification Report (gauging values of the precision, recall, and f1 score

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

for each class) & Fig. 35 for a Confusion Matrix (computing class-based accuracies), both derived from the validation set. Since the outputs of this model are probabilistic rather than discrete, a Receiver Operating Characteristic Graph has been plotted, allowing for greater transparency into the model's capability of distinguishing between classes (see Fig. 36). Also view Fig. 37 for an illustrative measure of the valued distribution of the validation predictions. Furthermore, Fig. 38 shows epoch-wise trends for training & testing accuracies & losses.

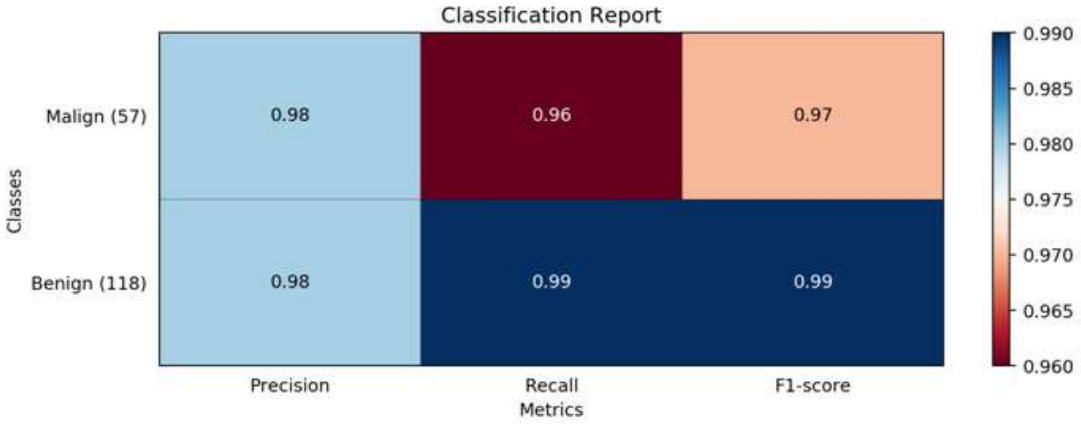



**Figure 34: Graphing classification report**

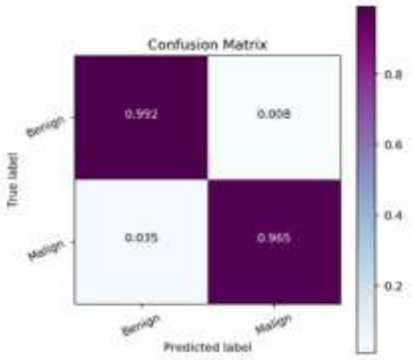



**Figure 35: Graphing confusion matrix**

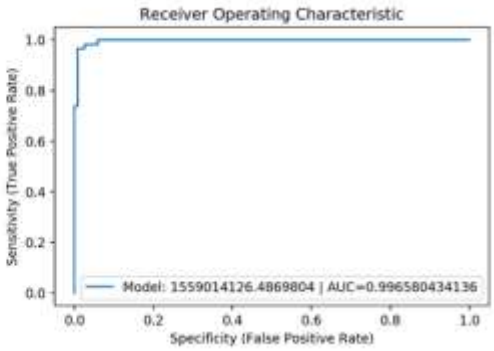



**Figure 36: Graphing ROC AUC Curve**

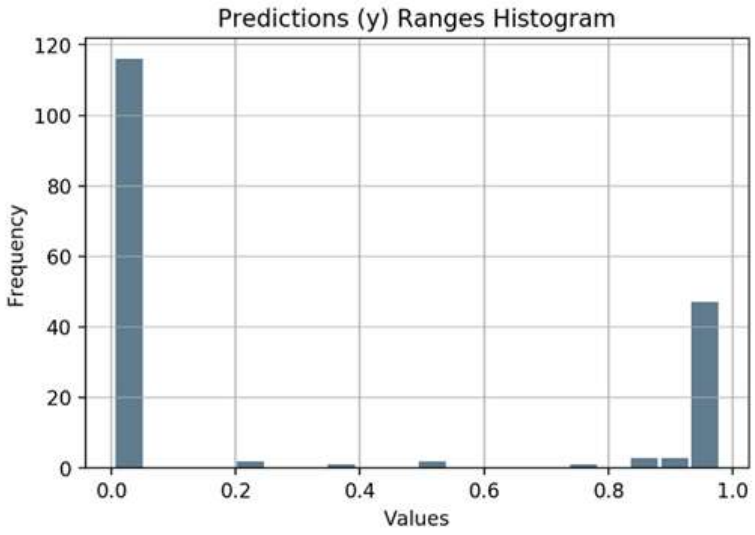



**Figure 37: Graphing predictions probability distribution**

*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**

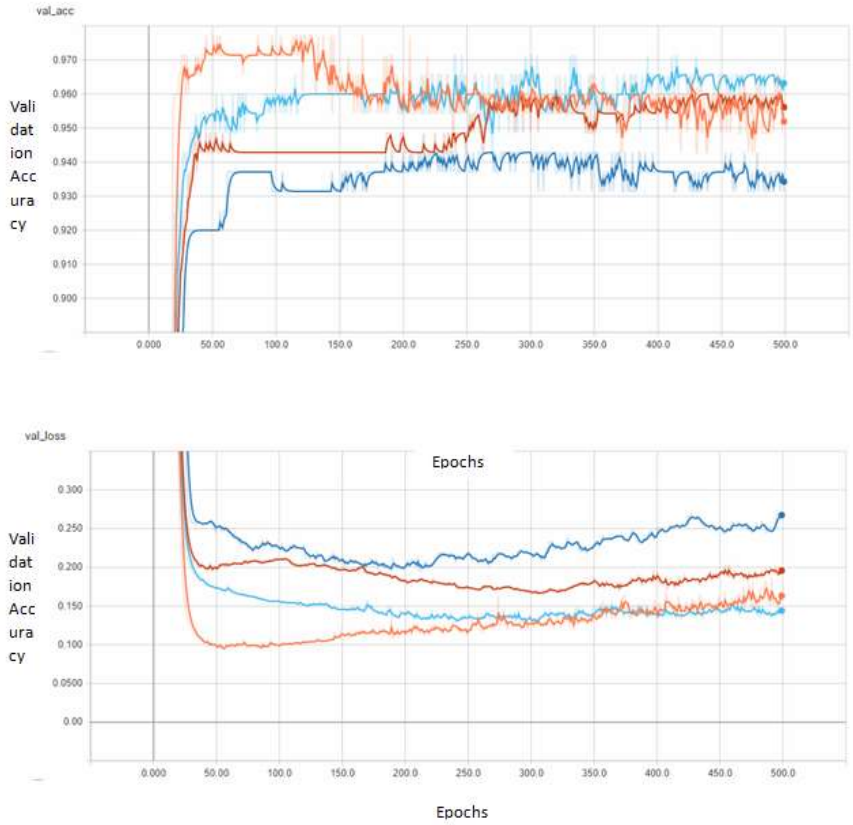



**Figure 38: Plotting epoch-wise trends for training & testing accuracies & losses for several models**

## VII. CLOUD-BASED PLATFORM

One of the main aspects of this system is its access and availability. This is primarily due to the software's web-based component. This application has user-friendly forms (see Fig. 39) in the cloud that accept inputs, predict against each of the trained models, and return outputs (see Fig. 40) instantaneously.

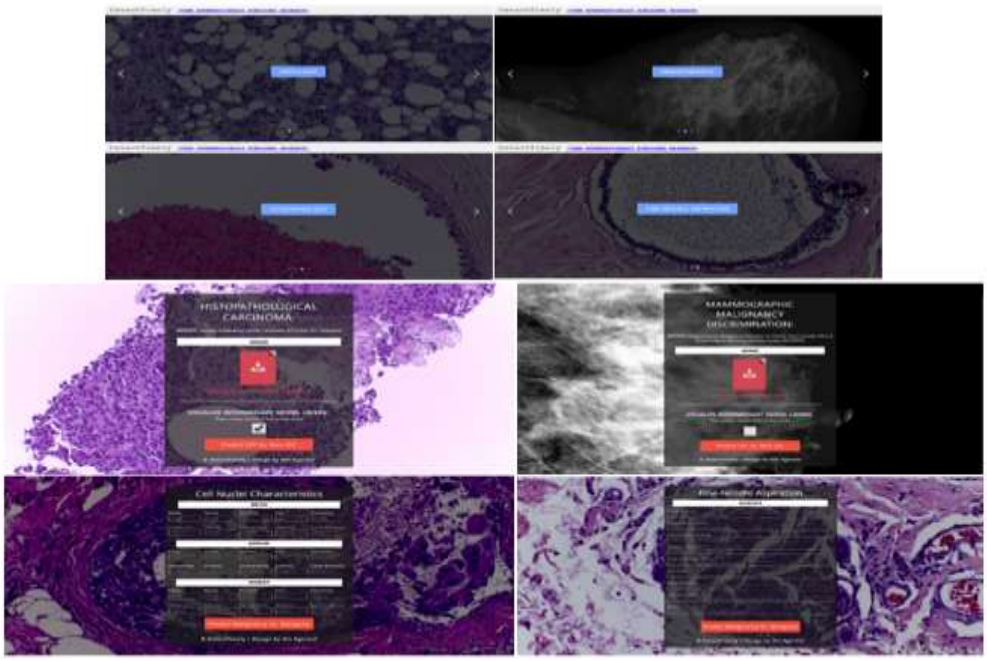


**Figure 39: Displaying web forms for inputting data, and interacting with the deep learning models**

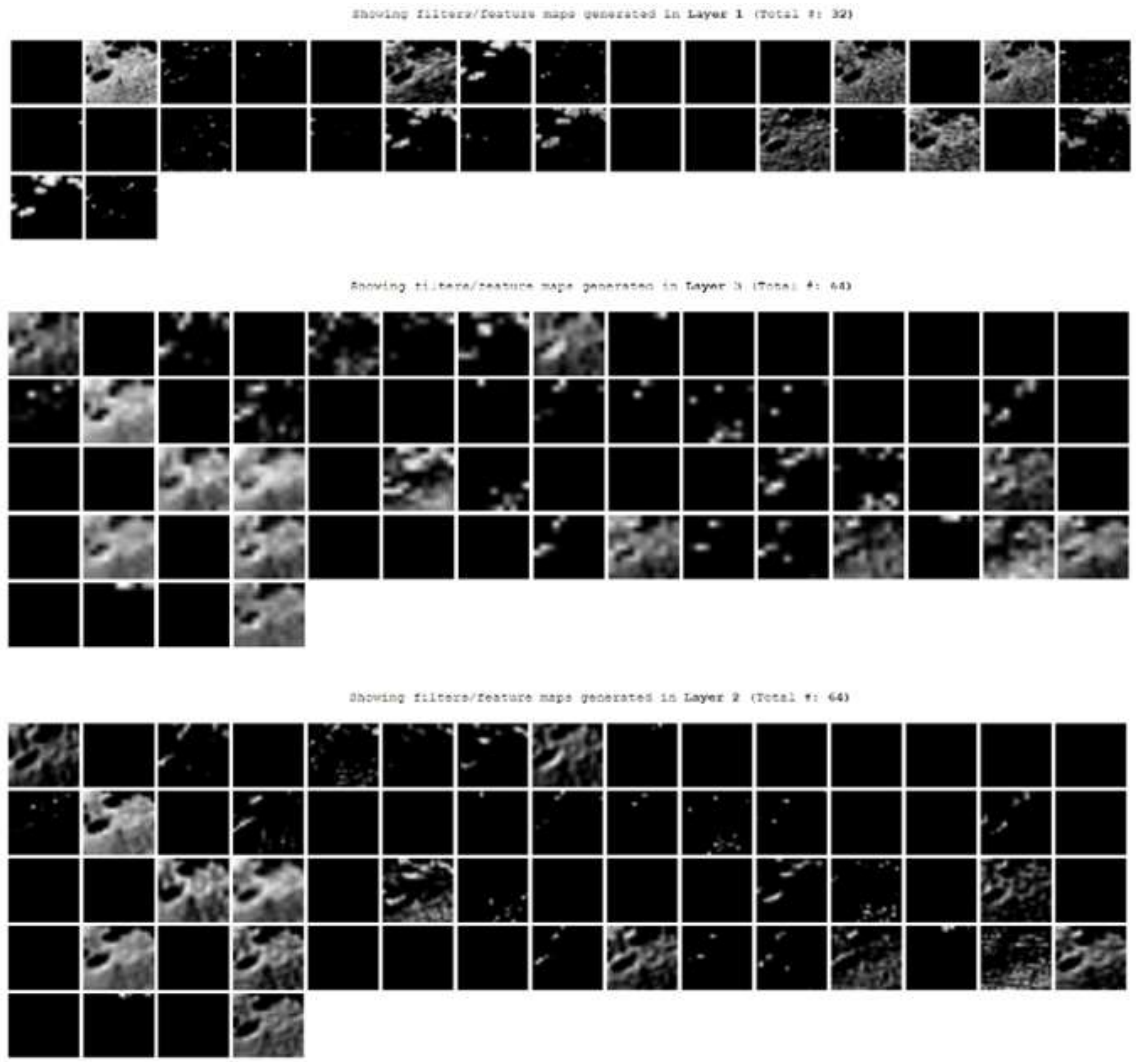
*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

**ISSN (online): 2581-3048**

**Volume 3, Issue 6, pp 1-24, June-2019**


**Figure 40: Displaying sample alertive results**

I also utilized ImgCNNViSual, my own fully-documented library for visualized an image input passing through each of the layers of the models, along with its respective filters/feature maps generated. As part of this software, this functionality induced additional transparency into the exact intermediary processes of the networks by bringing out distinct statures of inputs and outputs of the convolutional layers present in the contextual histograms, and mammograms (see Fig. 41).



*International Research Journal of Innovations in Engineering and Technology (IRJIET)*

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

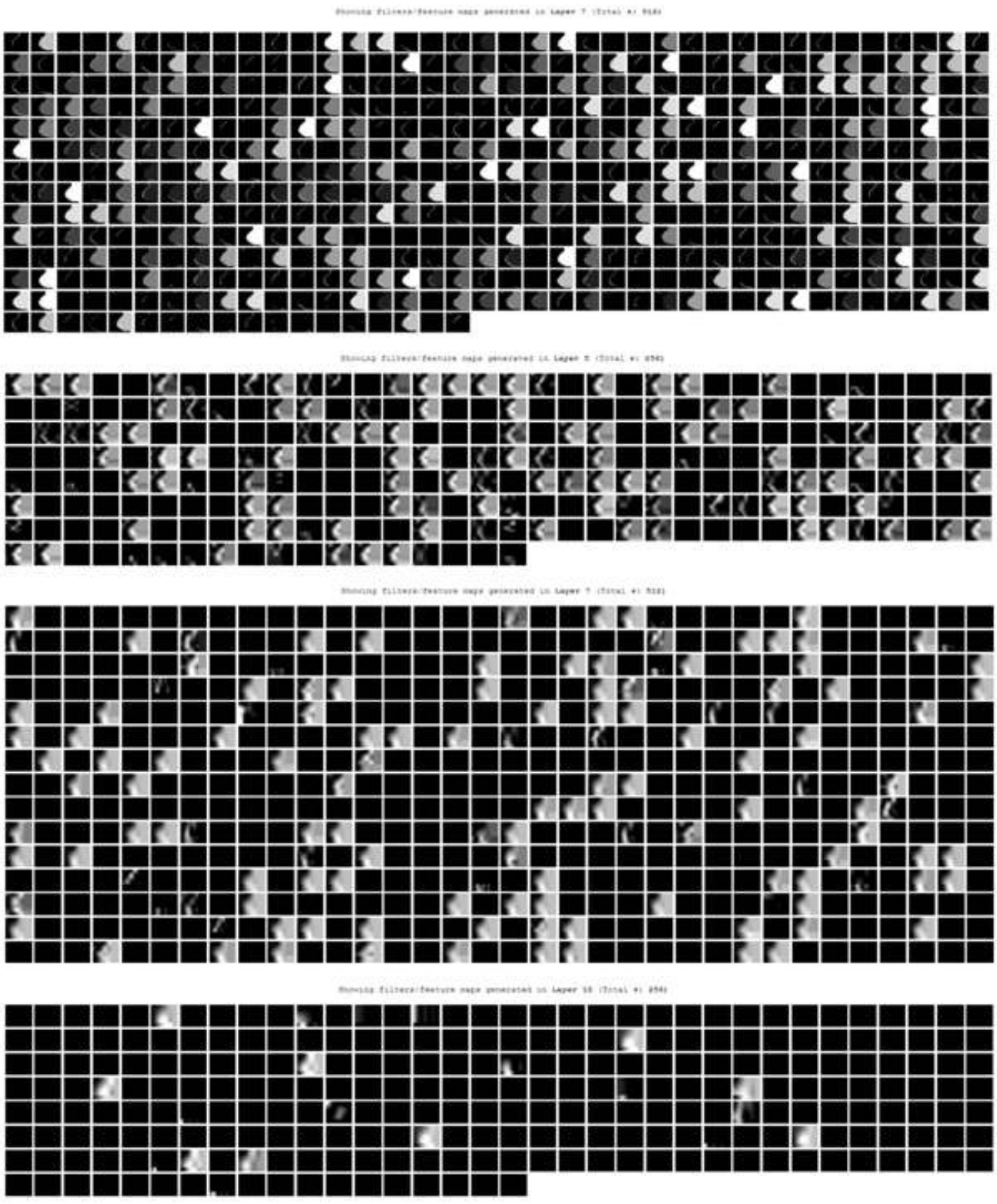


**Figure 41: Displaying sample visualizations of the uploaded image going through several models**

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 6, pp 1-24, June-2019

## VIII. TECHNOLOGIES / FRAMEWORKS

See Fig. 42 for a list of technologies utilized to make this project happen.


**Figure 42: List of technologies and frameworks used**

## IX. CONCLUSION

By leveraging the computational intuition of AI algorithms, this technology is indeed, successfully, able to perform a standardized and optimal medical diagnosis for BC, as proved by the accuracies, losses, and correlations computed on the validation matrices for each model. The model scan detects patterns which are far too complex to be recognized by even expert radiologists and pathologists and can efficiently apply their parameterized resolutions to the task of malignancy prediction.

DetectTimely v3.7 is an open-source, standardized, and comprehensive BC analytics platform, which delivers A.I. powered clinical optimization to cultivate a new standard of pathological/radiological interpretation through an automated environment with prognostic-based outcomes at its fore front. While it would be of substantial value to perform investigation-oriented clinical trials, this system is fully capable and ready to diagnose actual patients, as proved by the several stages of model validations. Currently the software's malignancy prediction task for the histology module has an accuracy of 84.5%, while the mammography module has an accuracy of 92.2%, the cytopathology module has an accuracy of 96.4%, and the fine-needle aspiration cytology module has an accuracy of 99.1%, making the overall technology significantly better than what typical pathologists and radiologists can achieve, which as mention previously are, 68.39%, 60-75%, 90-92%, and 90-92%, for the tasks respectively. More importantly though, this system is able to compute output predictions in a matter of seconds, as opposed to the prolonged weeks taken by pathologists, and radiologists (due to their world-wide shortage).Still, more global participation is required to adopt the intellectual logistics of these scientific findings and increase the predictive success on blind samples, where a traditional interpretation would deem it inconclusive. If more data were available to counterbalance rare outliers, it would be very valuable to further help this research project reach its full potential in terms of the enhanced development of its predictive capabilities, hopefully, one day meaning that breast cancer may no longer be a lethal disease.

The hope of this system is that it would make better the late/misdiagnosis issues commonly associated with breast cancer, which may appreciably aid in lowering chances of adversely evitable, and unnecessary delay in a patient's treatment. Furthermore, by adding functionality to visualize the image-related models, this framework provides precise insights into the intuitive inconsistencies of the patient's pathological report, which in turn helps provide full-scale intuition into the technicalities of the malignancy. Lastly, by optimizing the required deep learning models' predictions to the maximum, a radical alteration in the negative repercussions posed by human-error is ensured, inducing a viable causality for a lower rate of unhandled complexity within the data, and increasing the long-term success rate for the prognostic orchestration led by a positive decrease in the optimal trend for inconclusive results. This result will not only increase executional standards of incontrovertible productivity of the software's intended functionalities, but also potentially characterize an unseen paradigm-shift in the internal logistical processes of more general forms of such acute medical diagnosis.

## X. NEXT STEPS

Even though this project is beyond anything currently in the marketplace for early cancer detection systems, there is always scope for advancement. In the short-term, I plan to

ISSN (online): 2581-3048
Volume 3, Issue 6, pp 1-24, June-2019

acquire more datasets in order for this platform's features to accommodate other types of tests for breast cancer (MRIs, CTs, Ultrasounds,& PET Scans). In the long-term, I would like to get FDA approval for this technology, so that it can actually be used in a medical setting authentically. I also plan to develop a business model, so as to figure out how to sell and market this technology to clinicians.

## XI. ACKNOWLEDGMENTS

I would like to graciously thank the curators of the well maintained open datasets I've utilized, without which this project would not have been possible. Namely, I commend the 15th International Conference on Image Analysis and Recognition for providing the BACH dataset, the University of South Florida for creating the CBIS1-DDSM dataset, the Mammographic Image Analysis Society for preparing the MIAS-M dataset, and the University of California Irvine for their Breast Cancer Wisconsin (Original & Diagnostic) datasets.

I'd also like to appreciate Dalhousie University for providing a network of high-end NVIDIA GPUs to perform all of my models' training, development, and deployment.

## REFERENCES

### A. Books

[1] Barh, D., Carpi, A., Verma, M., & Gunduz, M. (2014). Cancer biomarkers: Minimal and noninvasive early diagnosis and prognosis. Boca Raton: CRC Press, Taylor & Francis Group.
[2] *** Casil, A. S. (2009). Breast cancer: Current and emerging trends in detection and treatment. New York: Rosen Pub.
[3] Chakraborty, S. (2011). Novel insights into the molecular diagnosis and therapy of breast cancer.
[4] Dugdale, D. C., & Eisenberg, M. S. (1992). Medical diagnostics. Philadelphia: W.B. Saunders.
[5] *** Hancke, S. (1981). Mammograms in the diagnosis of breast cancer: Screening & Staging. Stockholm: Almqvist & Wiksell International.
[6] Marsland, S. (2015). Machine Learning: An algorithmic perspective. Boca Raton, FL: CRC Press.
[7] Taktak, A. F., & Fisher, A. C. (2007). Outcome prediction in cancer. Amsterdam: Elsevier.

### B. Websites

[8] Bradley, A. P. (2001, June 07). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Retrieved November 03, 2018, from https://www.sciencedirect.com/science/article/abs/pii/S0031320396001422
[9] Jeerez, J. M. (2010, July 16). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Retrieved November 03, 2018, from https://www.sciencedirect.com/science/article/pii/S0933336571000067 9
[10] Polat, K., & Gunes, S. (2006, November 27). Breast cancer diagnosis using least square support vector machine. Retrieved November 03, 2018, from https://www.sciencedirect.com/science/article/pii/S1051200406001461
[11] Wolberg, W. H., Street, W. N., &Mangasarian, O. L. (2004, March 17). Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. Retrieved November 04, 2018, from https://www.sciencedirect.com/science/article/pii/030438359490099X
[12] Kourau, K., & Exarchos, T. P. (2014, November 15). Machine learning applications in cancer prognosis and prediction. Retrieved November 04, 2018, from https://www.sciencedirect.com/science/article/pii/S2001037014000464
[13] Petuum, Inc. (2018, August 01). Deep Learning for Breast Cancer Identification from Histopathological Images. Retrieved November 05, 2018, from https://medium.com/@Petuum/deep-learning-for-breast-cancer-identification-from-histopathological-images-f38de0a658a5
[14] Asri, H., Mousannif, H., & Moatassime, H. A. (2016, May 12). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Retrieved November 06, 2018, from https://www.sciencedirect.com/science/article/pii/S1877050916302575
[15] Nahid, Kong, & Yinan. (2017, December 31). Involvement of Machine Learning for Breast Cancer Image Classification: A Survey. Retrieved November 08, 2018, from https://www.hindawi.com/journals/cmmm/2017/3781951/
[16] Fu, M. R., & Wang, Y. (n.d.). Machine learning for the detection of lymphedema among breast cancer survivors. Retrieved November 10, 2018, from http://mhealth.amegroups.com/article/view/19663/19656
[17] Sharma, A., & Kulshsrestha, S. (n.d.). Machine learning approaches for breast cancer diagnosis and prognosis. Retrieved November 10, 2018, from https://ieeexplore.ieee.org/document/8280082
[18] Vazquez, F. (n.d.). KDnuggets. Retrieved November 10, 2018, from https://www.kdnuggets.com/2018/05/detecting-breast-cancer-deep-learning.html
[19] Yue Zhao, Nian Wang, & Xiaoyu Cui (n.d.). IOPScience: Aided diagnosis methods of breast cancer based on machine learning. Retrieved November 11, 2018, from

ISSN (online): 2581-3048
Volume 3, Issue 6, pp 1-24, June-2019

http://iopscience.iop.org/article/10.1088/1742-6596/887/1/012072

[20] Lopes, J. M. (2018, July 11). In Breast Cancer, Machine Learning Offers Early Lymphedema Detection. Retrieved November 14, 2018, from https://breastcancer-news.com/2018/07/11/machine-learning-provides-early-lymphedema-detection-breast-cancer-survivors/

[21] Saha, A., Harowicz, M. R., Grimm, L. J., Kim, C. E., Ghate, S. V., Walsh, R., & Mazurowski, M. A. (2018, July 23). A machine learning approach to radiogenomics of breast cancer: A study of 922 subjects and 529 DCE-MRI features. Retrieved November 14, 2018, from https://www.nature.com/articles/s41416-018-0185-8

[22] O'Conner, M. (2018, June 12). 5 machine learning algorithms ID lymphedema among breast cancer patients. Retrieved November 20, 2018, from https://www.healthimaging.com/topics/artificial-intelligence/5-machine-learning-algorithms-id-lymphedema-among-breast-cancer

[23] Mcgill University. (n.d.). Machine Learning Algorithms for Radio-Frequency Breast Cancer Detection at McGill University on FindAPhD.com. Retrieved November 22, 2018, from https://www.findaphd.com/phds/project/machine-learning-algorithms-for-radio-frequency-breast-cancer-detection/?p93532

[24] Google. (2018, October 22). Google Deep Learning Tool 99% Accurate at Breast Cancer Detection. Retrieved November 22, 2018, from https://healthitanalytics.com/news/google-deep-learning-tool-99-accurate-at-breast-cancer-detection

[25] Bates, M. (2018, March 27). Deep Learning Could Help Detect Breast Cancer. Retrieved November 22, 2018, from https://bioengineeringtoday.org/emerging-tech/deep-learning-could-help-detect-breast-cancer

**Citation of this article:**

Om Agarwal, "An Augmentation in the Diagnostic Potency of Breast Cancer through A Deep Learning Cloud-Based A.I. Framework to Compute Tumor Malignancy & Risk" Published in *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, Volume 3, Issue 6, pp 1-24, June 2019.

*******