



D Y PATIL INTERNATIONAL UNIVERSITY
AKURDI PUNE

Sector 29, Pradhikaran, Akurdi, Pune - Maharashtra, INDIA 411044
(Establishment by Maharashtra Act No. LXIII of 2017)

School of Computer Science, Engg. & Applications

Team Member's name:

- 1) Sahil Dinesh Chavan: 20190802042
- 2) Mayuresh Bhagwan Nehe: 20190802076

Email Id:

- 1) Sahil email-id: 20190802042@dypiu.ac.in
- 2) Mayuresh email-id: 20190802076@dypiu.ac.in

Phone Number:

- 1) Sahil Contact No: 9834020445
- 2) Mayuresh Contact No: 9527683038

Title: Breast Cancer Classification with Deep Learning.

Abstract:

Cancer-related fatal illnesses affect both developed and developing countries globally. Particularly, the incidence of female breast cancer cases is increasing every day, in part due to underdiagnosis and late detection. An efficient first-line treatment for breast cancer can only be administered by accurately recognizing and diagnosing cancer in its very early stages of development.

The use of computer-aided diagnosis and medical image analysis technologies may accelerate and automate the identification and categorization of cancer by instructing and supporting less experienced clinicians. For large datasets of medical images, convolutional neural networks considerably aid in the precise identification and categorization of cancer.

Keywords:

- 1) Electronic diagnostics
- 2) Mammary cancer
- 3) Deep Neural Network
- 4) Transfer learning
- 5) Classification

Literature survey:

Breast cancer develops when certain breast cells start to grow abnormally, according to a method outlined in the article Ultrasound characterization of Breast Masses by S. Gokhale.

These cells multiply and spread more quickly than healthy cells do, continue to amass, and eventually form a lump or mass that may begin to hurt.

Rapid cell spread from your breast to your lymph nodes or to other areas of your body is possible. Due to their family history, way of life, obesity, exposure to radiation, and reproductive variables, certain women may be more susceptible to developing breast cancer.

Breast cancer prediction is an active topic of study, according to Pragya Chauhan and Amit Swami's paper, Breast Cancer Prediction Using Genetic Algorithm Based Ensemble Approach.

Problem Definition & Objectives:

The second most common cancer in both men and women worldwide is breast cancer. It accounted for roughly 12% of all new cancer cases and 25% of all cancers in women in 2012.

Breast cells begin to grow out of control, which is how breast cancer develops. These cells typically develop into tumors that are commonly seen on x-rays or are thought to be lumps. If the tumor's cells have the ability to spread (metastasize) to other parts of the body or enlarge into surrounding tissues, the tumor is considered malignant (cancer).

Methodology:

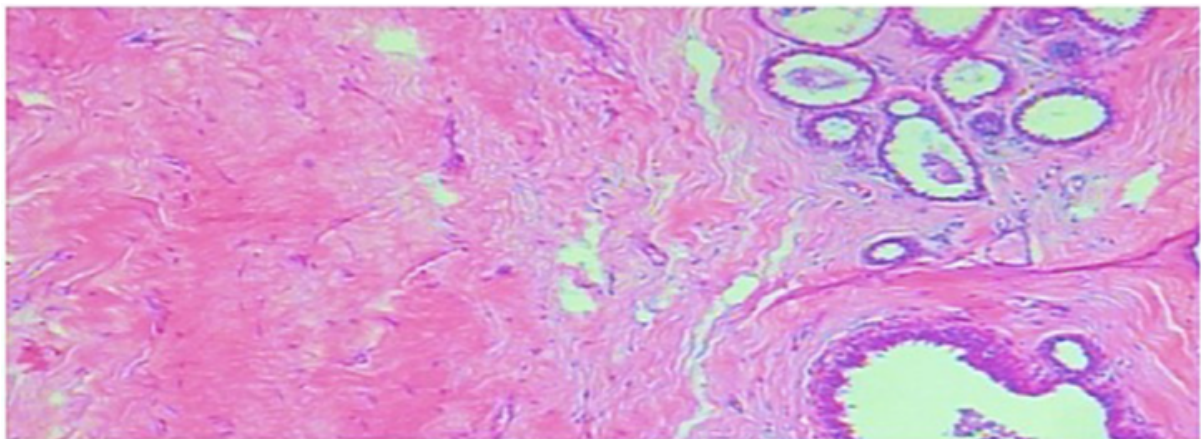
There is a tonne of pictures of cancer sufferers accessible. The dataset, often referred to as the Breast Cancer Histopathological imaging Database (BreakHis), was downloaded by completing the online form on the vision laboratory's website.

The 7909 total samples that are currently accessible were taken from 82 different patients. Breast cancer photos, both benign and

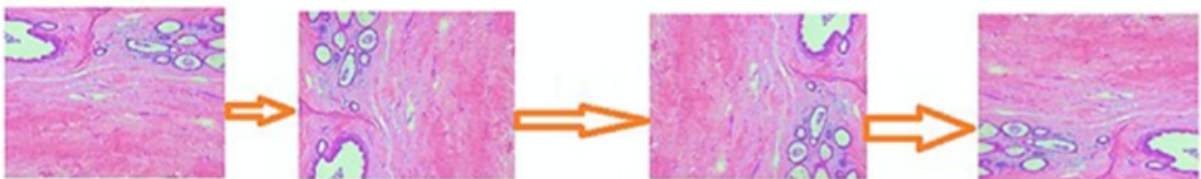
malignant, are included in the dataset utilized in this study. There are four possible magnification factors for each of the accessible photos in the collection.

The dataset contains 2480 benign lesions and 5429 malignant lesions, respectively. Adenosis (AE), Fibroadenoma (FA), Phyllodes Tumor (PT), and Tubular Adenoma are four classifications that are covered by benign pictures (TA).

Architecture:



Original Infected Image



Augmented Images

Algorithm:

a) Convolution Neural Network:

Convolutional neural networks are a form of artificial neural networks used most frequently in deep learning to interpret visual vision.

i) Use 3X3 CONV filters:

A kernel, convolution matrix, or mask is a tiny matrix used in image processing. It can be used for edge detection, embossing, sharpening, and more. By performing a convolution between a kernel and an image, this is achieved.

ii) Stack these filters on top of each other

iii) Perform Max-Pooling:

A pooling operation known as maximum pooling, also known as max pooling, determines the maximum or largest value in each patch of each feature map. The outcomes are down-sampled or pooled feature maps that emphasize the feature that is most prevalent in the patch, as opposed to the feature's average presence in the case of average pooling.

b) Deep Neural Network:

Artificial intelligence (AI) and machine learning techniques called deep learning model how people acquire specific types of information. Data science, which also encompasses statistics and predictive modeling, contains deep learning as a key component.

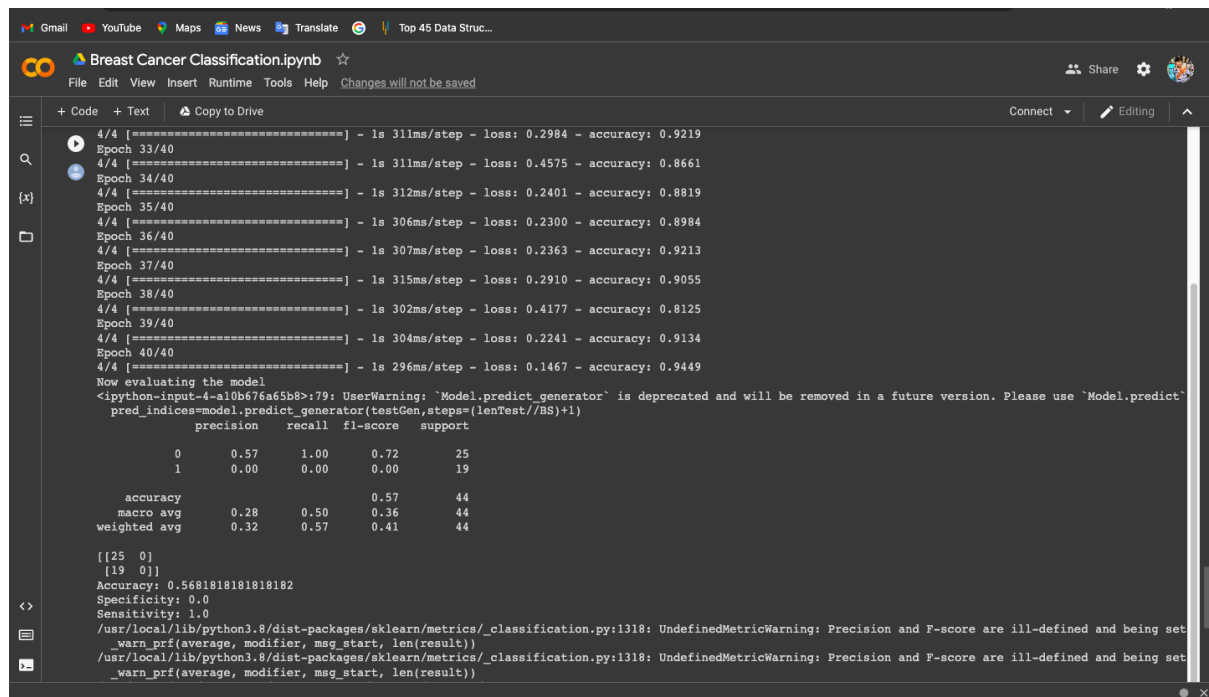
Datasets:

We'll use the IDC_regular dataset (the breast cancer histology image dataset) from Kaggle.

This dataset holds 2,77,524 patches of size 50×50 extracted from 162 whole-mount slide images of breast cancer specimens scanned at 40x. Of these, 1,98,738 tested negative, and 78,786 tested positive with IDC.

Experiment Result:

Here, we are attaching a screenshot of the experiment result of our project which tells us about the accuracy and number of epochs.



The screenshot shows a Jupyter Notebook interface with the title 'Breast Cancer Classification.ipynb'. The notebook contains a series of training epochs and their corresponding loss and accuracy values. The training progress is as follows:

Epoch	Loss	Accuracy
33/40	0.2984	0.9219
34/40	0.4575	0.8661
35/40	0.2401	0.8819
36/40	0.2300	0.8984
37/40	0.2363	0.9213
38/40	0.2910	0.9055
39/40	0.4177	0.8125
40/40	0.2241	0.9134
41/40	0.1467	0.9449

After training, the model is evaluated. The evaluation results are shown in a table:

	precision	recall	f1-score	support
0	0.57	1.00	0.72	25
1	0.00	0.00	0.00	19
accuracy			0.57	44
macro avg	0.28	0.50	0.36	44
weighted avg	0.32	0.57	0.41	44

The notebook also displays the confusion matrix and the accuracy score:

```
[[25  0]
 [19  0]]
Accuracy: 0.5681818181818182
Specificity: 0.0
Sensitivity: 1.0
```

Warnings are also shown at the bottom of the notebook:

```
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to nan due to no predicted samples
warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to nan due to no predicted samples
warn_prf(average, modifier, msg_start, len(result))
```

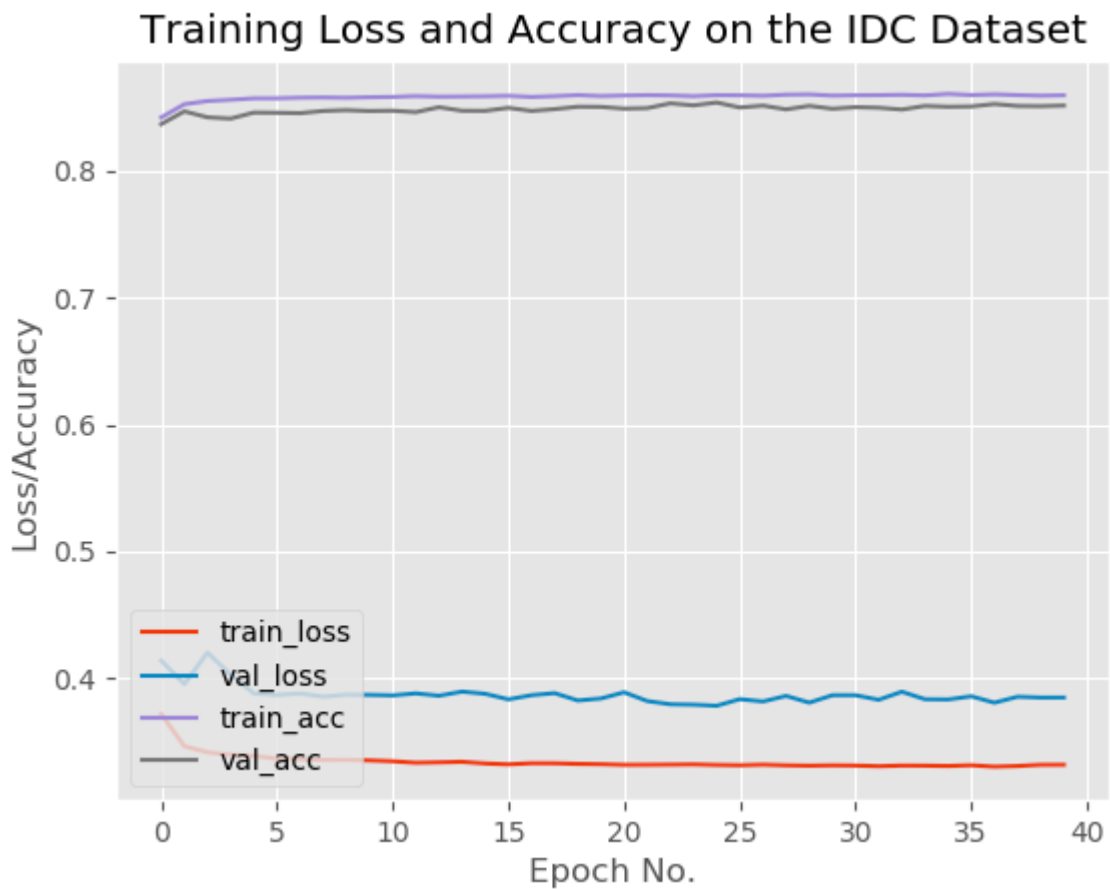
Conclusion:

In this Python project, we developed the network CancerNet and learned how to build a breast cancer classifier using the IDC dataset (containing histology images for Invasive Ductal Carcinoma). To do the same, we made advantage of Keras. I hope this Python project was enjoyable.

Result:

We have successfully trained our model using the IDC dataset. Then, we evaluate the performance of the model using a confusion matrix

and we got 56% of accuracy.



Future Scope:

End-to-end Breast Mass Classification

The main objective of an end-to-end computer-aided diagnosis system is to improve classification accuracy while minimizing false positives. One can improve our architecture to better simulate the diagnosis procedure followed by radiologists.

Reference:

- [1] World Health Organization, WHO Position Paper on Mammography Screening, World Health Organization, 2014.

- [2] S. Ahmad, S. Ur Rehman, A. Iqbal, R. K. Farooq, A. Shahid, M. I. Ullah, Breast cancer research in Pakistan: a bibliometric analysis:, 11(3) (2021).

- [3] M.T. McCann, J.A. Ozolek, C.A. Castro, B. Parvin, J. Kovacevic', Automated histology analysis: opportunities for signal processing, IEEE Signal Process.

- [4] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, J. Fan, Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis.

