



Deduction of IDC Using Deep Learning

K S Ramya Devi, M Srenidhi and R M Rani

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 18, 2021

DEDUCTION OF IDC USING DEEP LEARNING

BY

K.S.RAMYA DEVI

M.SRENIDHI

GUIDED BY

DR. RANI

ABSTRACT

Breast cancer is one of the main causes of cancer death worldwide. The diagnosis of biopsy tissue with hematoxylin and eosin stained images is non-trivial and specialists often disagree on the final diagnosis. Computer-aided Diagnosis systems contribute to reduce the cost and increase the efficiency of this process. Conventional classification approaches rely on feature extraction methods designed for a specific problem based on field-knowledge. To overcome the many difficulties of the feature-based approaches, deep learning methods are becoming important alternatives. A method for the classification of hematoxylin and eosin stained breast biopsy images using Convolutional Neural Networks (CNNs) is proposed. Deep learning has the ability to handle such complex situations which include high-dimensional matrix multiplications. Various architectures of CNN were applied and the model with the high generalization accuracy and minimal complexity is selected. The histopathology images are given as input to the CNN network as training models and then finally classified as having IDC or Malignancy. The best model is selected after varying the number of hidden layers and then applied to the dataset for final classification. Manual detection of a cancer cell is a tiresome task and involves human error, and hence computer-aided mechanisms are applied to obtain better results as compared with manual pathological detection systems. In deep learning, this is generally done by extracting features through a convolutional neural network (CNN) and then classifying using a fully connected network. In this project in python, we'll build a classifier to train on 80% of a breast cancer histology image dataset. Of this, we'll keep 10% of the data for validation. Using Keras, we'll define a CNN (Convolutional Neural Network), call it CancerNet, and train it on our images. We'll then derive a confusion matrix to analyze the performance of the model. IDC is Invasive Ductal Carcinoma; cancer that develops in a milk duct and invades the fibrous or fatty breast tissue outside the duct; it is the most common form of breast cancer forming 80% of all breast cancer diagnoses. And histology is the study of the microscopic structure of tissues.

1. Introduction

Developing effective methods for the automated detection of IDC remains a challenging problem for breast cancer diagnosis. The machine learning approach for detection of invasive ductal carcinoma (IDC) from whole image slides containing breast histological cancer cells. Of all phenotypes of breast cancer, invasive ductal carcinoma (IDC) is the most common and comprising of nearly 80% of all diagnosis. Their method, based on a Convolutional Neural Network (CNN), does not need to handcraft features from images and this work has the potential of revolutionizing cancer detection, promoting further research and development in this exciting direction. Developing effective methods for the automated detection of IDC remains a challenging problem for breast cancer diagnosis.

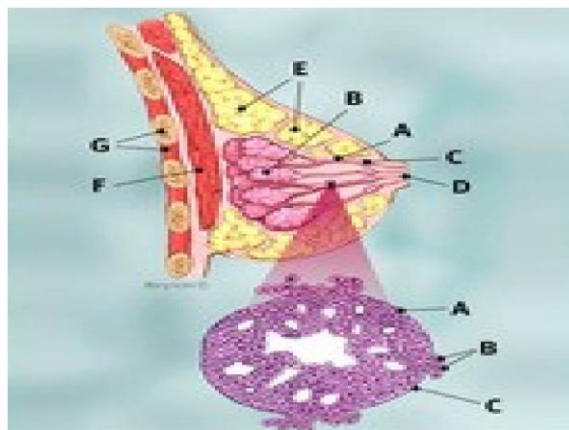
Recently, Cruz and his coworkers proposed a machine learning approach for detection of invasive ductal carcinoma (IDC) from whole image slides containing breast cancer cells. Their method, based on a Convolutional Neural Network (CNN), does not need to handcraft features from images. Their work has the potential of revolutionizing cancer detection, promoting further research and development in this exciting direction. Inspired by Cruz's work, our team investigated various CNN architectures for automated detection of breast cancer. The detection of IDC using deep learning is new existing field of research in medical and technological field. A detailed knowledge of the cancer cells at a place is an important prerequisite for cancer detection and research. Analyzing the affected areas of body cells are useful in various fields like detection, treatment and preventing us from future cancer invasion.

The major task associated with this work is automatic classification between cancerous and non-cancerous tissue histopathology microscopic images, which would be a valuable computer-aided diagnosis tool for the clinician. Python is a programming language which is used for statistical analysis and graphical methods. Charts or graphs are very much helpful to visualize and understand large amounts of complex data. As cancer analysis is the computational task it automatically predict the future effects of breast cancer in same or different patients using the model build on the basis of Convolutional Neural Network (CNN). By monitoring the patient's cancer cells and their histological images we can able to either solve the problems related invasion of ductal carcinoma or we can come up with some new techniques that help us. Deep learning techniques have the power to automatically extract features, retrieve information from data automatically, and learn advanced abstract representations of data.

They can solve the problems of traditional feature extraction and have been successfully applied in computer vision ([He et al., 2015](#); [Xie et al., 2018](#)), biomedical science ([Gulshan et al.,](#)

[2016; Esteva et al., 2017](#)) and many other fields. In view of the powerful feature extraction advantages of deep learning and the challenges in histopathological image analysis of breast cancer, this paper analyzes histopathological images of breast cancer using deep learning techniques. On one hand, we use advanced deep convolutional neural networks, including Inception_V3 ([Szegedy et al., 2016](#)) and Inception_ResNet_V2 ([Szegedy et al., 2017](#)), combined with transfer learning techniques to classify the histopathological images of breast cancer ([Pan and Yang, 2010](#)).

On the other hand, by combining deep learning with clustering and utilizing the dimension-reduction functionality of the autoencoder network ([Hinton and Salakhutdinov, 2006](#)), we propose a new autoencoder network structure to apply non-linear transformations to features in histopathological images of breast cancer extracted by the Inception_ResNet_V2 network. This effectively maps the extracted features to a lower dimensional space. The newly obtained features are then used as input for the classical clustering algorithm known as K-means ([MacQueen, 1967](#)) to perform clustering analysis on histopathological images of breast cancer. Also, we designed a number of comparable experiments to verify the validity of our proposed method of histopathological image analysis of breast cancer images b



Age	In Situ Cases #	In Situ Cases %	Invasive case #	Invasive case %	Death #	Death %
<40	1610	3	11160	4	990	2
40-49	12440	20	39920	15	3480	9
50-59	17680	28	58620	23	7590	19
60-69	17550	28	68070	27	9420	23
70-79	10370	16	47860	19	8220	20
80+	3760	6	30080	12	10910	27
All Ages	63410		252710		40610	

2. Literature Review

STUDIES RELATED TO BREAST CANCER

Soft Computing techniques play an important role for decision in applications with imprecise and uncertain knowledge. The application of soft computing disciplines is rapidly emerging for the diagnosis and prognosis in medical applications. Among the various soft computing techniques, fuzzy expert system takes advantage of fuzzy set theory to provide computing with uncertain words. In a fuzzy expert system, knowledge is represented as a set of explicit linguistic rules. Diagnosis of breast cancer suffers from uncertainty and imprecision associated with imprecise input measures and incompleteness of knowledge of experts. However, there are several technology-oriented studies reported for breast cancer diagnosis, few studies has been reported for the breast cancer prognosis. Fatima. [1] describe a fuzzy expert system for breast cancer prognosis to further support of the process of breast cancer diagnosis. This approach is capable enough to capture ambiguous and imprecise information prevalent in characterization of breast cancer. For this, the paper utilizes a fuzzy reasoning model, which has high interpretability for interacting with human experts during prognosis process and consequently early diagnosis of the diseased. The performance results on real patients dataset reveal the accuracy of the system with an average 95% which shows the superiority of the system in the prognosis process compared to other related works. Breast Cancer Diagnosis and Prognosis were two medical applications, which pose a great challenge to the researchers. The use of machine learning and data mining techniques have revolutionized the whole process of breast cancer Diagnosis and Prognosis. Breast Cancer Diagnosis distinguishes benign from malignant breast lumps and Breast Cancer Prognosis predicts when Breast Cancer are likely to recur in patients that had their cancers excised. Thus, these two problems were mainly in the scope of the classification problems. This study paper summarizes various reviews and technical articles on breast cancer diagnosis and prognosis. Shelly Gupta., et al in [2] describes an overview of the current research being carried out using the data mining techniques to enhance the breast cancer diagnosis and prognosis.

PROJECT DESCRIPTION

3.1 OBJECTIVES

- Preventing the future cancer patients through analysis
- Prevention of huge loss
- Study of cancer
- Diagnosis and analysis of cancer affected patients
- Provides absolute results

3.2 PROBLEM DEFINITION

Breast cancer Prediction and analysis is the application of medical side to predict the state of the benign and malignant type cancer cells. India is an with most number of women cancer patients especially with IDC types. For analyzing and diagnosis of the cancer cells deep learning provides the vast platform in maintaining and analyzing the cancer datasets using histological image datasets of affected patients across country.

Most of the problems occur due to the unawareness of food habits across the country that will arise in future. To protect the people especially women from IDC ,this Deep Learning is a subfield of Machine Learning which uses supervised, unsupervised or semi-supervised learning to automatically form useful information from data. It is similar in structure and function to the human nervous system which processes complex information with the help of a compounded network of interconnected computational units

3.3 PROPOSED SYSTEM

Breast Cancer Dataset (BCD)

Charts or graphs are very much helpful to visualize and understand large amounts of complex data. Python is a programming language which is used for statistical analysis and graphical methods. Python is an interpreted language which is especially for analytical purpose. Anaconda Spyder is a free open source Integrated Development Environment (IDE) for Python programming language. Python is specially for handling complex data and has graphical facilities mainly for visualizing the data. There are several packages performing different functions. The potential of the proposed method is obtained using classification accuracy which was obtained by comparing actual to predicted values. The outcome confirms that the maximum classification accuracy (99.28%) is achieved for this study. The output is in the graphical form. So it is simple for the people from non technical field to understand. The dataset named BraKHis used in this article was published by [Spanhol et al. \(2016a\)](#) in 2016. It is composed of 7,909 histopathological images from 82 clinical breast cancer patients. The database can be accessed through the link <http://web.inf.ufpr.br/vri/breast-cancer-database>. To save the original organization structure and molecular composition, each image was taken by a pathologist from a patient's breast tissue section using a surgical biopsy. Then, the images were collected via haematoxylin and eosin staining. Finally, the real class label was given to each image by pathologists via their observations of the images from a microscope. All the histopathological images of breast cancer are 3 channel RGB micrographs with a size of 700×460 . Since objective lenses of different multiples were used in collecting these histopathological images of breast cancer, the entire dataset comprised four different sub-datasets, namely 40, 100, 200, and 400X. All of these sub-datasets are classified into benign and malignant tumors.

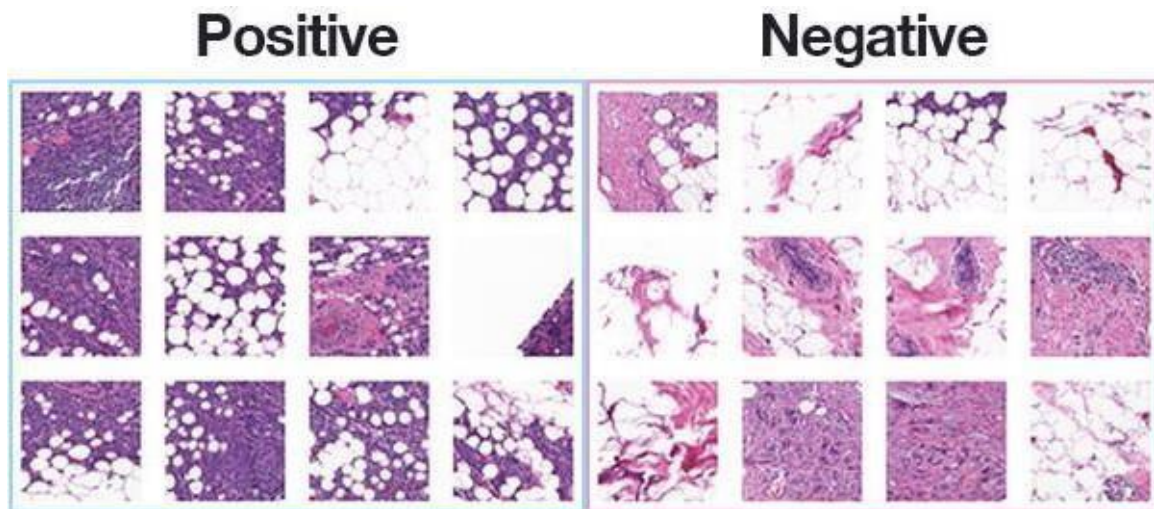


Fig: 3.4.1 Dataset Image

The dataset we are using for today's post is for Invasive Ductal Carcinoma (IDC), the most common of all breast cancer. The dataset was originally curated by [Janowczyk and Madabhushi](#) and [Roa et al.](#) but is available in public domain on [Kaggle's website](#).

The original dataset consisted of 162 slide images scanned at 40x.

Slide images are naturally massive (in terms of spatial dimensions), so in order to make them easier to work with, a total of 277,524 patches of 50x50 pixels were extracted, including: 198,738 negative examples (i.e., no breast cancer)

78,786 positive examples (i.e., indicating breast cancer was found in the patch)

There is clearly an imbalance in the class data with over 2x the number of negative data points than positive data points.

Each image in the dataset has a specific filename structure. An example of an image filename in the dataset can be seen below:

10253_idx5_x1351_y1101_class0.png We can interpret this filename as:

Patient ID: 10253_idx5

x-coordinate of the crop: 1,351 y-coordinate of the crop: 1,101

Class label: 0 (0 indicates no IDC while 1 indicates IDC)

Therefore, both benign and malignant tumors have four different subsets. Benign tumors include Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT), and Tubular Adenoma (TA). Malignant tumors include Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC) each of the histopathological images of breast cancer must be transformed into a 299 x 299 image to match the required input size of the network structure. Some image preprocessing methods in the TensorFlow framework were used in the transforming process, including cutting the border box, adjusting image size, and adjusting saturation, etc. These methods are used to create two classifiers that must discriminate benign from malignant breast lumps. To create the

classifier, the WBCD (Wisconsin Breast Cancer Diagnosis) dataset is employed. proposed diagnosis for BS based on nuclei segmentation of cytological images, using different neural nets and support vector machines, achieved accuracy rate ranging from 76% to 94% on a dataset of 92 images.

4. ANALYSIS AND DESIGN

Invasive Ductal Carcinoma analysis is the computational task of automatically predict the affected cancer patients amount using the model build on the basis of linear regression and lasso regression. Accuracy of the model can be tested by using the test dataset.

IMPORTING AND CLEANING

The data are imported into the python base and edit section. Usually the data are in a structured or cleaned histological image format. By using these unordered data we can't build a model. So build a model we need to do some pre-processing which is commonly called as cleaning of data. The histological image of data sets carried cancer affected cells of both benign and malignant type, where 0 indicates the unaffected cancer cells (i.e benign type) and 1 indicates the affected cancer cells (i.e malignant type)

4.1 SPLITTING OF DATA

The dataset that we are using are split into two namely of validate and train. One of dataset is used for training the model and the other is used for checking the accuracy of our trained model. To achieve high level of accuracy, huge amount of data and henceforth computing power is needed to train these networks. However, despite the computational complexity involved Predicting. The validation set is used to evaluate a given model, but this is for frequent evaluation. We as machine learning engineers use this data to fine-tune the model hyper parameters. The challenging task for image classification particularly the microscopic images from histopathological section is due to the large amount of inter-interaction variables, presence of complex geometrical structure, complex textures and minute details in image which can be the region of interest for classification [27] depicts the complex texture found in histopathological images. Here Deep learning provides the possibilities of learning features directly from input data and process it through its hidden layers. In this paper, the CNN model is used to classify image samples into cancerous or non-cancerous tissues and the accuracy is compared with other classifier models

4.2 SYSTEM ARCHITECTURE

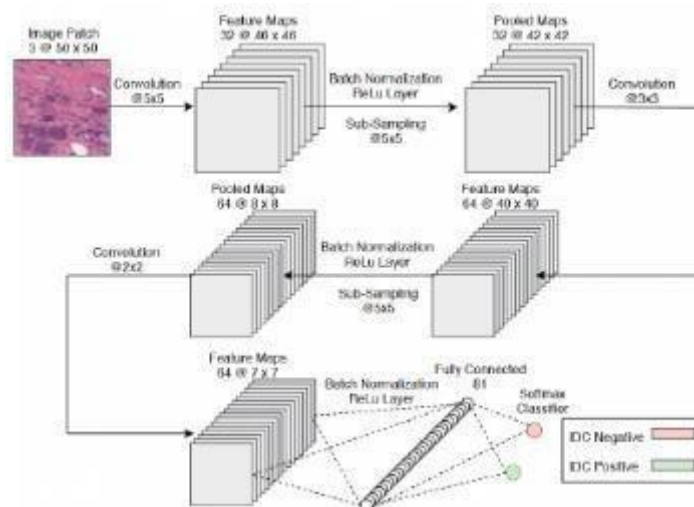


Fig: 4.2.1 Base CNN Architecture

IMAGE CLASSIFICATION USING CNN

The CNN model is used to classify image samples into cancerous or non-cancerous tissues and the accuracy is compared with other classifier models. Pre-processing It is the neural network which is used in image recognition, object recognition, image classifications, face recognition etc. In CNN the input image is passed through a series of convolutional layers with kernels, pooling layer and fully connected layer before the final image classification. Convolution layer is the first stage where the features are extracted from the input image. Convolution is performed with different filters to perform functions like edge detection, sharpening or blurring. When the filters cannot accommodate the input image, padding is used so that the output image is same size as the input image. There are two types of padding commonly used 1) Zero padding and 2) Valid Padding. The input image is then passed through a non-linear activation function like ReLU or Sigmoid or Tan Function. When the number of parameters in a large image is high, pooling is used to reduce the dimension retaining the important information. Different architectures of CNN were applied to the dataset consisting of 5547 patches of images that were extracted from the original IDC dataset of 277,524 samples. The accuracies of different architectures are compared in Table 4. It can be observed that most of the architectures suffer from bias. The best performing model gives an accuracy of 76% with negligible bias. The Break His dataset is split into two parts, training and validation sets. Different transform techniques like random scaling, cropping and flipping were applied on training set. Since a pre-trained network RESNET-152 is used [29], so the input image was resized as required by the network. Also normalization methods were applied on each color channel to center it at 0 to 1 range. Based on the pre-trained network, new untrained feed-

forward network acts as a classifier using ReLU as the activation function. The Adam optimizer with a learning rate of 0.005 is used for the training set. The total time taken to complete the training was 18 min 20s and the accuracy was found to be 89%.

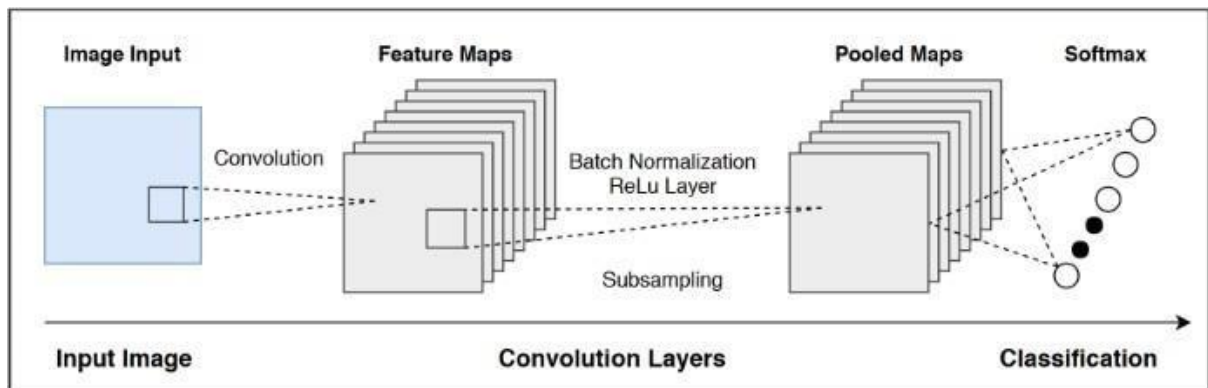


Fig: 4.2.2 Generic Architecture of CNN

POOLING

Pooling can be 1) Max pooling 2) Sum pooling 3) Average pooling. The last step includes feeding out image into a Fully Connected layer after flattening it into a vector. Finally, the flattened image is passed through the activation function to classify the output.

TRAINING

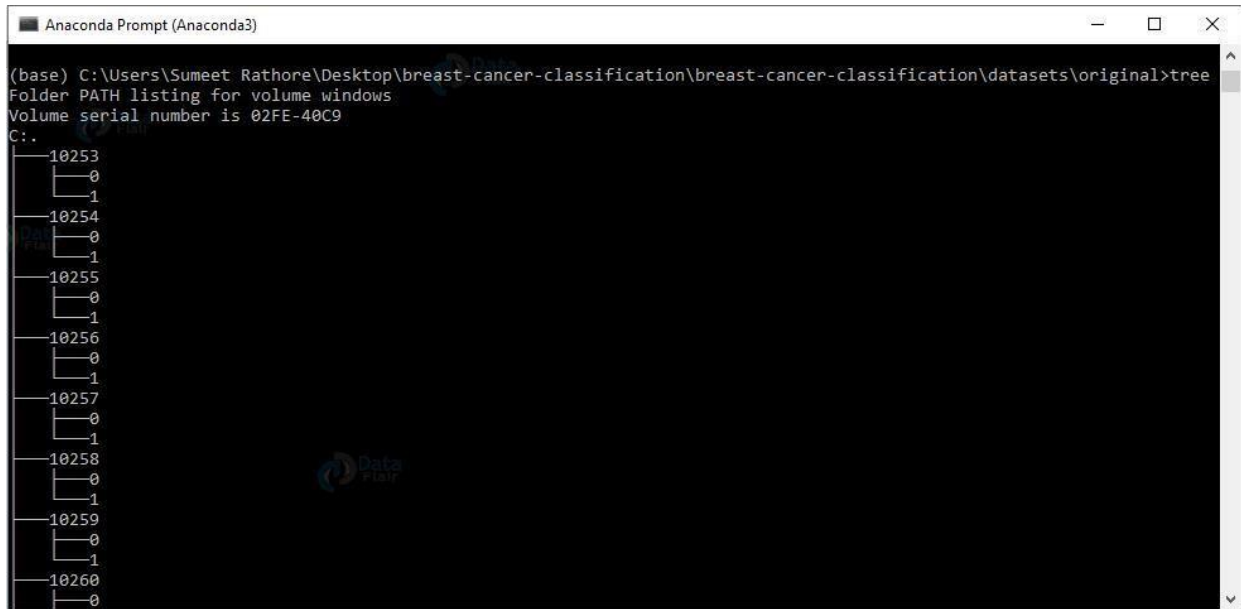
To achieve high level of accuracy, huge amount of data and henceforth computing power is needed to train these networks. However, despite the computational complexity involved Predicting.

VALIDATING

The validation set is used to evaluate a given model, but this is for frequent evaluation. We as machine learning engineers use this data to fine-tune the model hyper parameters

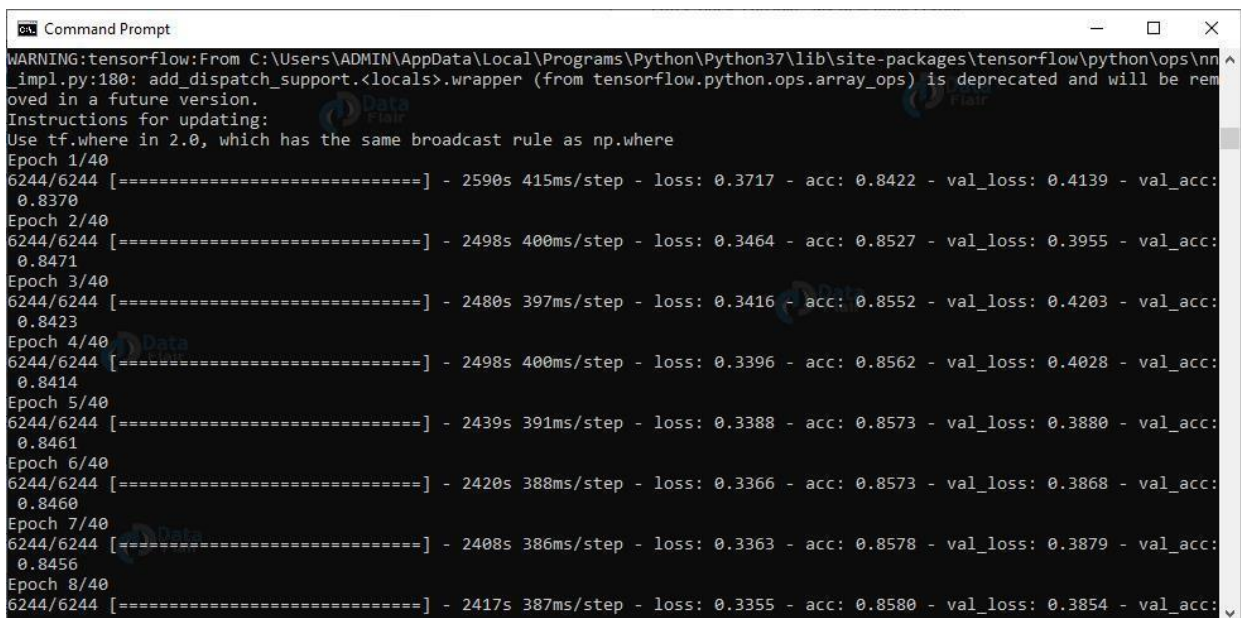
EXPERIMENTAL RESULTS

SCREENSHOTS



```
(base) C:\Users\Sumeet Rathore\Desktop\breast-cancer-classification\breast-cancer-classification\datasets\original>tree
Folder PATH listing for volume windows
Volume serial number is 02FE-40C9
C:
├── 10253
│   ├── 0
│   └── 1
├── 10254
│   ├── 0
│   └── 1
├── 10255
│   ├── 0
│   └── 1
├── 10256
│   ├── 0
│   └── 1
├── 10257
│   ├── 0
│   └── 1
├── 10258
│   ├── 0
│   └── 1
├── 10259
│   ├── 0
│   └── 1
└── 10260
    └── 0
```

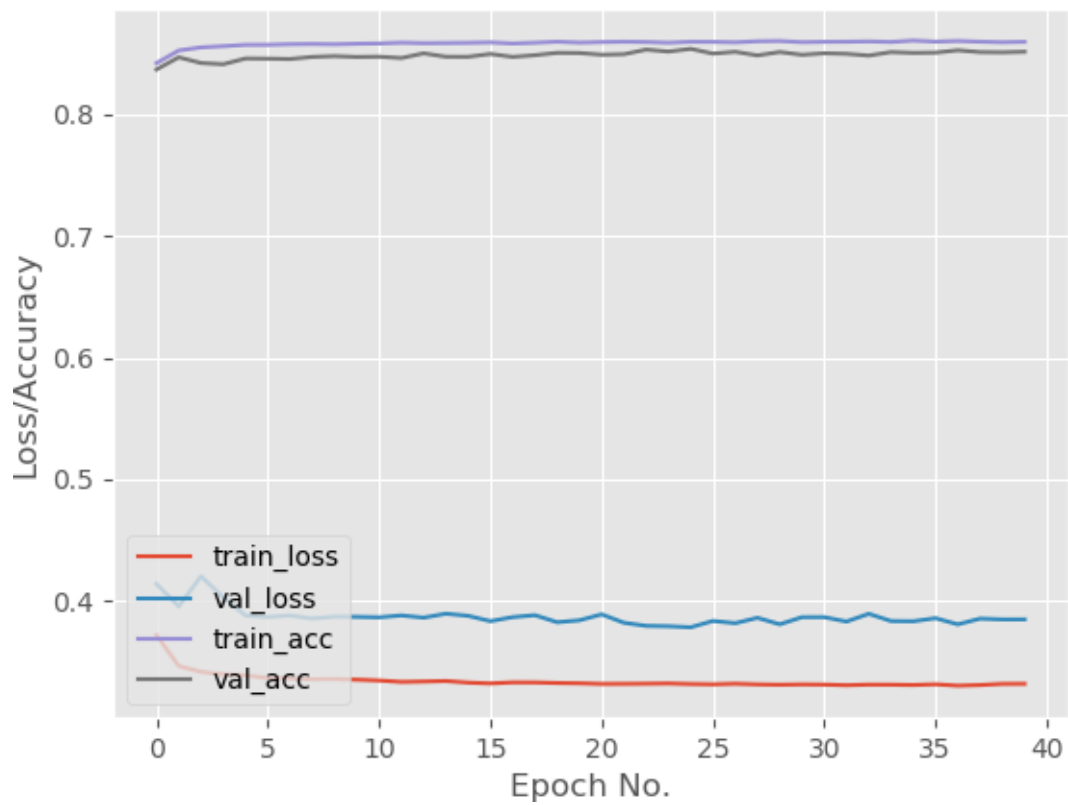
Tree Plot



```
WARNING:tensorflow:From C:\Users\ADMIN\AppData\Local\Programs\Python\Python37\lib\site-packages\tensorflow\python\ops\nn_impl.py:180: add_dispatch_support.<locals>.wrapper (from tensorflow.python.ops.array_ops) is deprecated and will be removed in a future version.
Instructions for updating:
Use tf.where in 2.0, which has the same broadcast rule as np.where
Epoch 1/40
6244/6244 [=====] - 2590s 415ms/step - loss: 0.3717 - acc: 0.8422 - val_loss: 0.4139 - val_acc: 0.8370
Epoch 2/40
6244/6244 [=====] - 2498s 400ms/step - loss: 0.3464 - acc: 0.8527 - val_loss: 0.3955 - val_acc: 0.8471
Epoch 3/40
6244/6244 [=====] - 2480s 397ms/step - loss: 0.3416 - acc: 0.8552 - val_loss: 0.4203 - val_acc: 0.8423
Epoch 4/40
6244/6244 [=====] - 2498s 400ms/step - loss: 0.3396 - acc: 0.8562 - val_loss: 0.4028 - val_acc: 0.8414
Epoch 5/40
6244/6244 [=====] - 2439s 391ms/step - loss: 0.3388 - acc: 0.8573 - val_loss: 0.3880 - val_acc: 0.8461
Epoch 6/40
6244/6244 [=====] - 2420s 388ms/step - loss: 0.3366 - acc: 0.8573 - val_loss: 0.3868 - val_acc: 0.8460
Epoch 7/40
6244/6244 [=====] - 2408s 386ms/step - loss: 0.3363 - acc: 0.8578 - val_loss: 0.3879 - val_acc: 0.8456
Epoch 8/40
6244/6244 [=====] - 2417s 387ms/step - loss: 0.3355 - acc: 0.8580 - val_loss: 0.3854 - val_acc:
```

Epoch Iteration

Training Loss and Accuracy on the IDC Dataset



Validation Loss and Gain and Train Loss and Gain

SOURCE CODE

MANIFEST CODE FOR BUILDING IDC DATA SET:

```
import os
INPUT_DATASET = "datasets/original"
BASE_PATH = "datasets/idc"
TRAIN_PATH = os.path.sep.join([BASE_PATH, "training"])
VAL_PATH = os.path.sep.join([BASE_PATH, "validation"])
TEST_PATH = os.path.sep.join([BASE_PATH, "testing"])
TRAIN_SPLIT = 0.8
VAL_SPLIT = 0.1
```

MANIFEST CODE FOR BUILDING IDC DATA SET:

```
from cancernet import config
from imutils import paths
import random, shutil, os
originalPaths=list(paths.list_images(config.INPUT_DATASET))
random.seed(7)
random.shuffle(originalPaths)
index=int(len(originalPaths)*config.TRAIN_SPLIT)
trainPaths=originalPaths[:index]
testPaths=originalPaths[index:]
index=int(len(trainPaths)*config.VAL_SPLIT)
valPaths=trainPaths[:index]
trainPaths=trainPaths[index:]
datasets=[("training", trainPaths, config.TRAIN_PATH),
("validation", valPaths, config.VAL_PATH),
("testing", testPaths, config.TEST_PATH)
]
for (setType, originalPaths, basePath) in datasets:
print(f'Building {setType} set')
if not os.path.exists(basePath):
print(f'Building directory {basePath}')
os.makedirs(basePath)
for path in originalPaths:
file=path.split(os.path.sep)[-1]
label=file[-5:-4]
labelPath=os.path.sep.join([basePath,label])
```

```

if not os.path.exists(labelPath):
print(f'Building directory {labelPath}')
os.makedirs(labelPath)
newPath=os.path.sep.join([labelPath, file])
shutil.copy2(path, newPath)
MANIFEST CODE FOR PLOTTING AND PREDICTION OF RESULT:
import matplotlib
matplotlib.use("Agg")
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import LearningRateScheduler
from keras.optimizers import Adagrad
from keras.utils import np_utils
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from cancernet.cancernet import CancerNet
from cancernet import config
from imutils import paths
import matplotlib.pyplot as plt
import numpy as np
import os
NUM_EPOCHS=1;
INIT_LR=1e-2;
BS=32
trainPaths=list(paths.list_images(config.TRAIN_PATH))
lenTrain=len(trainPaths)
lenVal=len(list(paths.list_images(config.VAL_PATH)))
lenTest=len(list(paths.list_images(config.TEST_PATH)))
trainLabels=[int(p.split(os.path.sep)[-2]) for p in trainPaths]
trainLabels=np_utils.to_categorical(trainLabels)
classTotals=trainLabels.sum(axis=0)
classWeight=classTotals.max()/classTotals
trainAug = ImageDataGenerator(
rescale=1/255.0,
rotation_range=20,
zoom_range=0.05,
width_shift_range=0.1,

```

```

height_shift_range=0.1,
shear_range=0.05,
horizontal_flip=True,
vertical_flip=True,
fill_mode="nearest")
valAug=ImageDataGenerator(rescale=1 / 255.0)
trainGen = trainAug.flow_from_directory(
config.TRAIN_PATH,
class_mode="categorical",
target_size=(48,48),
color_mode="rgb",
shuffle=True,
batch_size=BS)
valGen = valAug.flow_from_directory(
config.VAL_PATH,
class_mode="categorical",
target_size=(48,48),
color_mode="rgb",
shuffle=False,
batch_size=BS)
testGen = valAug.flow_from_directory(
config.TEST_PATH,
class_mode="categorical",
target_size=(48,48),
color_mode="rgb",
shuffle=False,
batch_size=BS)
model=CancerNet.build(width=48,height=48,depth=3,classes=2)
opt=Adagrad(lr=INIT_LR,decay=INIT_LR/NUM_EPOCHS)
model.compile(loss="binary_crossentropy",optimizer=opt,metrics=
["accuracy"])
M=model.fit_generator(
trainGen,
steps_per_epoch=lenTrain//BS,
validation_data=valGen,
validation_steps=lenVal//BS,

```

```

class_weight=classWeight,
epochs=NUM_EPOCHS)
print("Now evaluating the model")
testGen.reset()
pred_indices=model.predict_generator(testGen,steps=(lenTest//BS)+1)
pred_indices=np.argmax(pred_indices,axis=1)
print(classification_report(testGen.classes,
pred_indices,target_names=testGen.class_indices.keys()))
cm=confusion_matrix(testGen.classes,pred_indices)
total=sum(sum(cm))
accuracy=(cm[0,0]+cm[1,1])/total
specificity=cm[1,1]/(cm[1,0]+cm[1,1])
sensitivity=cm[0,0]/(cm[0,0]+cm[0,1])
print(cm)
print(f'Accuracy: {accuracy}')
print(f'Specificity: {specificity}')
print(f'Sensitivity: {sensitivity}')
N = NUM_EPOCHS
plt.style.use("ggplot")
plt.figure()
plt.plot(np.arange(0,N), M.history["loss"], label="train_loss")
plt.plot(np.arange(0,N), M.history["val_loss"], label="val_loss")
plt.plot(np.arange(0,N), M.history["accuracy"], label="train_acc")
plt.plot(np.arange(0,N), M.history["val_accuracy"], label="val_acc")
plt.title("Training Loss and Accuracy on the IDC Dataset")
plt.xlabel("Epoch No.")
plt.ylabel("Loss/Accuracy")
plt.legend(loc="lower left")
plt.savefig('plot.png')

```


CONCLUSION AND FUTURE WORK

Breast cancer analysis is a field of study that analyzes affected cancer areas and the invasive amount that will occur in the future. A CNN technology is selected as data and these data are analyzed for cancer prediction. As cancer analysis is a computational task, it automatically predicts the future rainfall amount using the model built on the basis of linear and lasso regression. By monitoring the patient's histological image datasets, we can solve the patients by analyzing and diagnosing the histological images. Machine Learning and Deep Learning approaches have been implemented for Breast cancer classification. Different ML algorithms like Logistic Regression, KNN, SVM and Decision Tree are discussed and their accuracies are compared. Deep learning approaches like CNN are also studied and various performance measures have been evaluated to study the accuracy of the best architecture. The best performing CNN architecture gives an accuracy of 81% which is far more superior than the conventional ML algorithms. Larger datasets consisting of multiple resolutions of images were trained on GPU, leading to an improved performance accuracy of 89%. This paper indicates that Deep learning approaches can efficiently classify breast cancer samples compared to other models discussed in the paper. As a scope of future work, high-resolution images can be trained using GPU like CUDA toolkit or Google Colab which supports free GPU.

REFERENCES

- [1] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 Lyon, France: International Agency for Research on Cancer; 2013
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, and A. Jemal, 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6), pp.394-424.
- [2] Siegel, L. Rebecca, Kimberly D. Miller, and J. Ahmedin. "Cancer statistics, 2019." *CA: a cancer journal for clinicians* 69.1 (2019): 7-34.
- [4] Bonnema, Jorien, et al. "Ultrasound-guided aspiration biopsy for detection of nonpalpable axillary node metastases in breast cancer patients: new diagnostic method." *World journal of surgery* 21.3 (1997): 270-274.
- [5] Ponraj, D. Narain, et al. "A survey on the preprocessing techniques of mammogram for the detection of breast cancer." *Journal of Emerging Trends in Computing and Information Sciences* 2.12 (2011): 656-664.
- [6] M. A., Aswathy, and M. Jagannath. "Detection of breast cancer on digital histopathology images: Present status and future possibilities." *Informatics in Medicine Unlocked* 8 (2017): 74-79.
- [7] Cruz-Roa, Angel, et al. "High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection." *PloS one* 13.5 (2018): e0196828.
- [8] F. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, A Dataset for Breast Cancer Histopathological Image Classification, *IEEE Transactions on Biomedical Engineering (TBME)*, 63(7):1455-1462, 2016.
- [9] Filipczuk, Paweł, et al. "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies." *IEEE Transactions on Medical Imaging* 32.12 (2013): 2169-2178.
- [10] George, Yasmeen Mourice, et al. "Remote computer-aided breast cancer detection and diagnosis system based on cytological images." *IEEE Systems Journal* 8.3 (2014): 949-964.
- [11] Spanhol, Fabio Alexandre, et al. "Breast cancer histopathological image classification using

convolutional neural networks." 2016 international joint conference on neural networks (IJCNN). IEEE, 2016.

[12] Zhang, Yungang, et al. "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles." *Machine vision and applications* 24.7 (2013): 1405-1420.

13. Farahani, Navid, Anil V. Parwani, and LironPantanowitz. "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives." *Pathol Lab Med Int* 7 (2015): 23-33.