

# Computer Engineering Department

## A.P. Shah Institute of Technology

G.B.Road, Kasarvadavli, Thane(W), Mumbai-400615  
UNIVERSITY OF MUMBAI  
Academic Year 2019-2020

A Project Report on  
**Histopathologic Cancer Detection**  
Submitted in partial fulfillment of the degree of  
Bachelor of Engineering in Computer Engineering

By  
Mr. Ketan Muddalkar (16102022)  
Mr. Rohan Dhere (15102057)

Under the Guidance of  
Prof. Pravin Adivarekar

---

# 1. Project Conception and Initiation

- Cancer is a group of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body.
- Cancer is the second most common disease in India responsible for maximum mortality with about 0.3 million deaths per year
- • Purpose of our project is to develop a system that can automatically detect cancer from the blood cell images.

# 1.1 Abstract

- In this project we aim to classify cancer images with maximum accuracy using deep learning techniques like Convolutional Neural Network(CNN).
- After achieving desired accuracy we deployed the model on the web using Amazon Web Services.
- The dataset which we are using contains 2 classes, class one consist of images of cancer-free patients and class two consist of images of patient having cancer.
- We are working with 3 models which are ensembled to get better accuracy

## 1.2 Objectives

- Create Deep Learning model that will accurately detect cancer from the histopathologic images.
- To make process smoother and fluent propose a model which will help doctors in taking decision correctly & accurately.
- Deploy the model on internet which will serve as a base for hospitals in the direction of fully digitizing their organization.
- Generating a medical report based on the result predicted by the model

# 1.3 Literature Review

- The performance of the deep learning network models depends on overall task at hand, dataset used, evaluation setup and much more.
- The most prevalent way to perform transfer learning is to employ a pre-trained model and to fine tune it with data at hand.
- Other approach is to use a pre-trained model as a feature extractor and perform further classification with a separate classifier.
- One of the transfer learning approach is to use an architecture that has done well in other tasks and to train it from scratch.

# 1.4 Problem Definition

- Cancer arises from the transformation of normal cells into tumour cells in a multistage process that generally progresses from a precancerous lesion to a malignant tumour if undetected.
- Cancer cells are hard to detect for human eyes even for a professional pathologist.
- Hence with the help of this deep learning method we intend to speed-up the process of cancer detection with high accuracy and hence contribute towards reduction in number of deaths caused due to cancer.

# 1.5 Scope

- The project will involve gathering data from hospitals or websites.
- Planning to deploy the model on a website which can act as an example for hospitals to create a similar portal for the benefits mentioned.
- The scope of this project is to deliver a model, with world-class accuracy, which can detect cancer from histopathology images and produce top notch results.



# 1.6 Technology stack

- Languages  
PHP, Python, HTML, CSS, MySQL
- Libraries  
Tensorflow, Keras, Pandas, NumPy, OpenCV, Bootstrap
- Environments/ Tools  
Apache Server, Google Colab, Nvidia DGX, Amazon Web Services.

# 1.7 Benefits for environment & Society

- Training several large machine learning models with approximately 100 million parameters will have carbon footprint of roughly 600 kg also tuning them will require significant amounts of energy.
- To reduce this amount we'll be using an approach of transfer learning where pretrained models can be used to train on a new dataset.
- This method is computationally inexpensive which can save a lot of energy and help the environment.

## 2. Project Design

—

## 2.1 Proposed System

- The image which is submitted on the webpage by a patient will be saved at backend after storing it this image will be preprocessed too
- For the training we are using deep learning techniques like Convolutional Neural Network(CNN).
- We are working with 3 best models which are ensembled to get better accuracy compared to a single model.
- A final report is getting generated based on the results predicted by the model and this report is saved at the backend

## 2.2 Design(Modules)

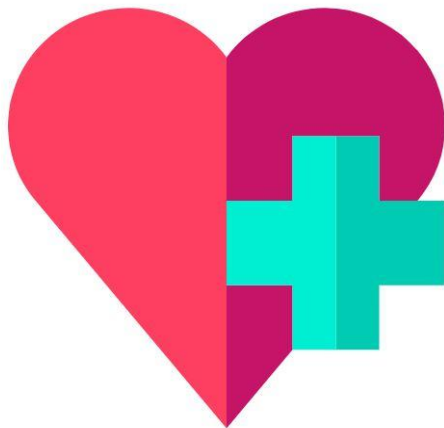
- Front End.
- Dataset.
- CNN Model.
- Back End.

## 2.3 Description Of Use Case

- Login
- Create Login
- Upload Image
- Detection
- View Output/Report
- Logout

## 2.5 Module 1:- Front End

- The website will act as a portal for both hospitals and patients.
- From the account, the user will have the option to upload the reports for diagnosis.
- Front end will include HTML, CSS which is designed using Bootstrap



## Patient Login

LOGIN

Forgot [Password?](#)  
[Create your Account](#)





RDS - AWS Console



Upload Image



Not secure | 35.153.176.163/upload\_img.php



17rohandhere@gmail.com

Upload Image

Previous Report

Logout



Choose File

No file chosen

Submit



---

## Medical Report

---

Patient Name Rohan Dhere  
Address xyz.....  
DATE Of Birth 1998-02-17  
Report Date 2020-04-14  
Report Time 08:30:43pm

APSIT HEALTH CENTER  
Ghodbunder Rd, opp. Hypercity Mall,  
Kasarvadavali, Thane West, Thane,  
Maharashtra 400615  
+91-123456789  
[apsitmproject@gmail.com](mailto:apsitmproject@gmail.com)

---

### Histopathologic Cancer Report For Patient : 303ed

Clinical Diagnosis : Positive  
Date of Diagnosis : 2020-04-14

#### Gross Description:

The specimen is received in two parts. They are labeled #1, "biopsy bladder tumor", and #2, "scalene node, left". Part #1 consists of multiple fragments of gray-brown tissue which appear slightly hemorrhagic. They are submitted in their entirety for processing. Part #2 consists of multiple fragments of fatty yellow tissue which range in size from 0.2 to 1.0 cm in diameter. They are submitted in their entirety for processing.

#### Microscopic:

Section of bladder contains areas of transitional cell carcinoma. No area of invasion can be identified. A marked acute and chronic inflammatory reaction with eosinophils is noted together with some necrosis. Sections are examined at six levels. Section of lymph node contains normal node with reactive germinal centers. All sections taken radially from the superficial center of the resection site fail to include tumor, indicating the tumor to have originated deep within the breast parenchyma. Similarly, there is no malignancy in the nipple region, or in the lactiferous sinuses. Sections of deep surgical margin demonstrate diffuse tumor infiltration of deep fatty tissues, however, there is no invasion of muscle. Total size of primary tumor is estimated to be 4 cm in greatest dimension.

# Create Account

Name:

Age:

Email:

Date of Birth

Contact:

Blood Group:

Password:

Address:

Re-enter Password:

Sex:

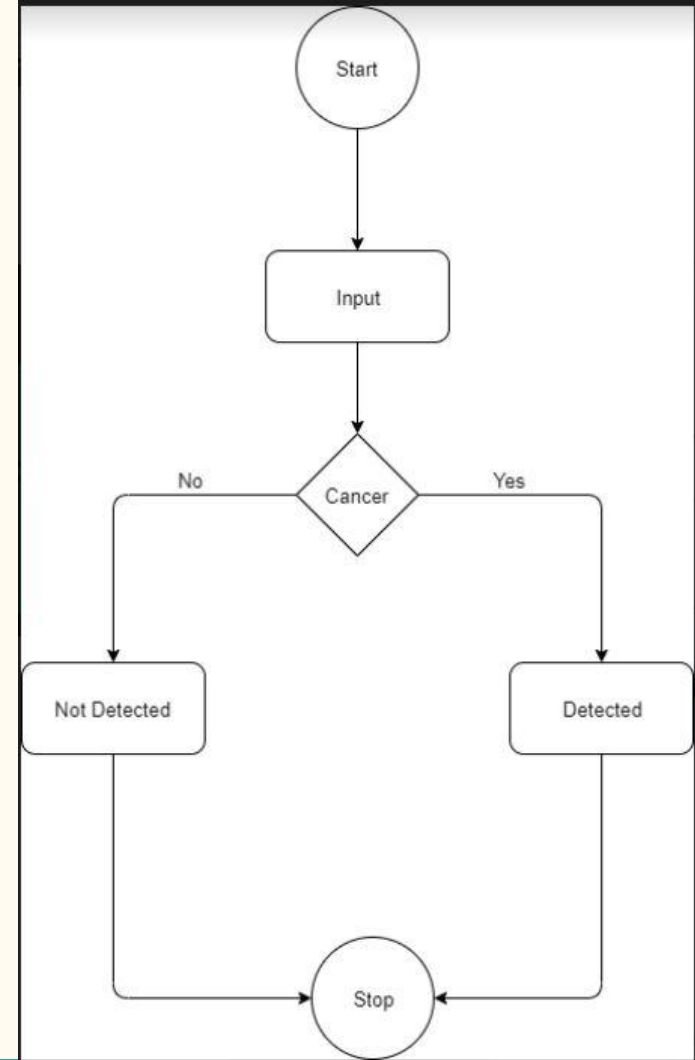
Medical History

Submit

# Forgot Password

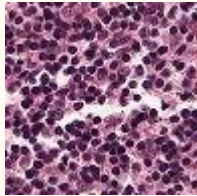
Submit

# Activity diagram

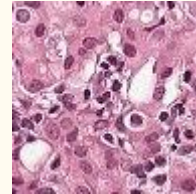


# Module 2.6 :- Dataset

- Understanding the Data
  - Modified version of the PatchCamelyon (PCam) benchmark dataset.
  - The dataset contains a total of 2.2 lakh images of cancer cells.
  - All of the images are 92x92 pixels in size.



Healthy Cell



Cancer Cell

# Data Preprocessing

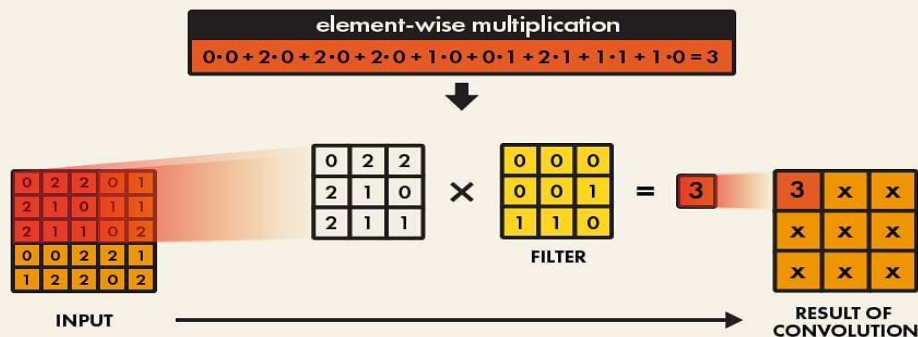
Parameter	Value
Horizontal Flip	True
Vertical Flip	True
Rotation Range	90
Zoom Range	0.2
Width Shift Range	0.1
Height shift Range	0.1
Shear Range	0.05

# Module 2.7 :- ConvNet

- This model will be responsible to predicting the result from the image given by the user
- This is where data pre-processing comes into the picture
- The model will be trained on these images.
- After achieving desired accuracy, the model will be allowed to predict the outcome for the input provided by the user.



# Convolution Function



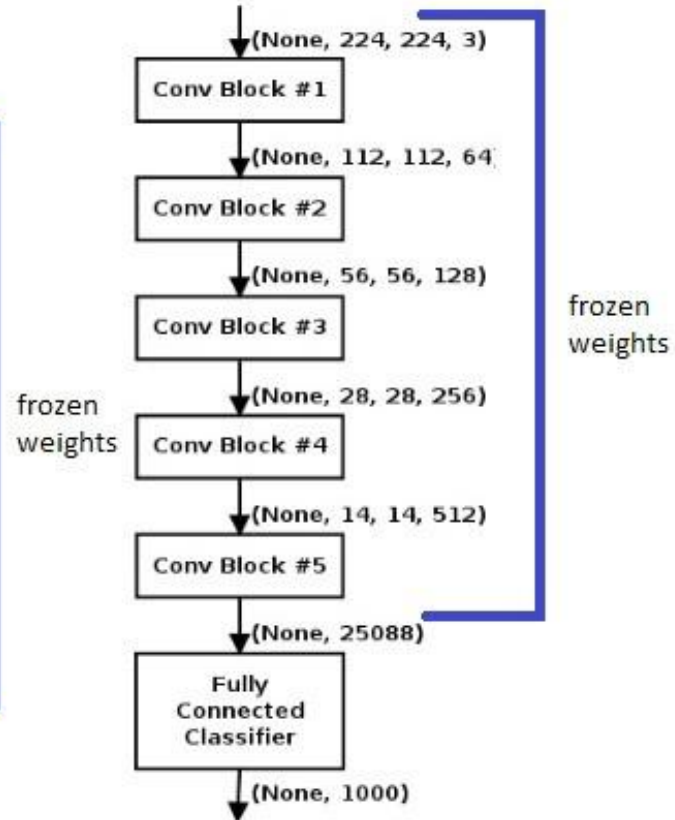
## Simple Convolution

Start with a simple filter, which is simply a small matrix of weights. This filter “convolving” around the input image, performing an element-wise multiplication with the original pixel value of the image, and then summing up the results into a single output as convolved feature.

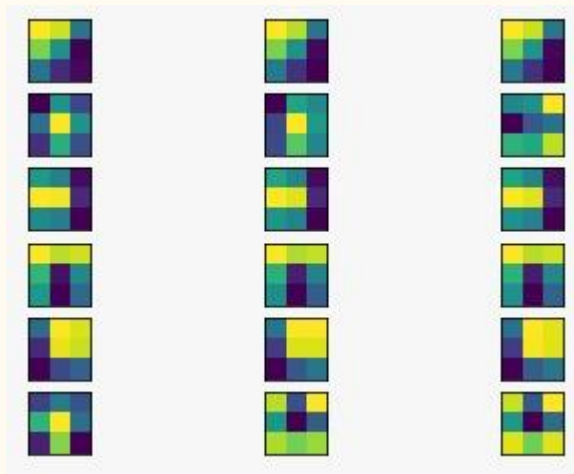
# Architecture of ConvNet

## Keras VGG-16 Model

```
( 0, 'input_6',      (None, 224, 224, 3))
( 1, 'block1_conv1', (None, 224, 224, 64))
( 2, 'block1_conv2', (None, 224, 224, 64))
( 3, 'block1_pool',  (None, 112, 112, 64))
( 4, 'block2_conv1', (None, 112, 112, 128))
( 5, 'block2_conv2', (None, 112, 112, 128))
( 6, 'block2_pool',  (None, 56, 56, 128))
( 7, 'block3_conv1', (None, 56, 56, 256))
( 8, 'block3_conv2', (None, 56, 56, 256))
( 9, 'block3_conv3', (None, 56, 56, 256))
(10, 'block3_pool',  (None, 28, 28, 256))
(11, 'block4_conv1', (None, 28, 28, 512))
(12, 'block4_conv2', (None, 28, 28, 512))
(13, 'block4_conv3', (None, 28, 28, 512))
(14, 'block4_pool',  (None, 14, 14, 512))
(15, 'block5_conv1', (None, 14, 14, 512))
(16, 'block5_conv2', (None, 14, 14, 512))
(17, 'block5_conv3', (None, 14, 14, 512))
(18, 'block5_pool',  (None, 7, 7, 512))
(19, 'flatten',      (None, 25088))
(20, 'fc1',          (None, 4096))
(21, 'fc2',          (None, 4096))
(22, 'predictions', (None, 1000))
```



# Visualization



Filters Visualization



Layers Visualization

# Module 2.7 :- Back-end

- This part consists of a web server, model, the database with all the user information and generation of user's medical report.
- AWS Services Used :
  - Ec2 instance (Elastic Compute Cloud).
  - Amazon RDS(Amazon Relational Database Service).
  - VPC (Virtual Private Cloud).

# EC2 Configuration

The screenshot displays the AWS Management Console interface for an EC2 instance. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a star icon. On the right, there are notification and account details: 'APSITCLOUD', 'N. Virginia', and 'Support'. Below the navigation bar, the 'Launch Instance' button is highlighted in blue, with 'Connect' and 'Actions' buttons to its right. A search bar with the placeholder 'Filter by tags and attributes or search by keyword' is present. The main content area shows a table of EC2 instances. The first instance, 'ec2-cloud-m...', is in the 'running' state. Below the table, the 'Description' tab is selected, showing a detailed configuration for the instance. The configuration is organized into two columns of key-value pairs.

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Public IP	IPv6 IP
ec2-cloud-m...	i-033b12374e525618f	t2.micro	us-east-1a	running	2/2 checks ...	None		35.153.176.163	-

**Description** | Status Checks | Monitoring | Tags

Instance ID	i-033b12374e525618f	Public DNS (IPv4)	-
Instance state	running	IPv4 Public IP	35.153.176.163
Instance type	t2.micro	IPv6 IPs	-
Finding	Opt-in to AWS Compute Optimizer for recommendations. <a href="#">Learn more</a>	Elastic IPs	
Private DNS	ip-10-0-0-30.ec2.internal	Availability zone	us-east-1a
Private IPs	10.0.0.30	Security groups	<a href="#">default</a> , <a href="#">view inbound rules</a> , <a href="#">view outbound rules</a>
Secondary private IPs		Scheduled events	No scheduled events
VPC ID	vpc-000a6c61df5545ca8 (cloud-mafia-vpc)	AMI ID	ubuntu/images/hvm-ssd/ubuntu-bionic-18.04-amd64-server-20200112 (ami-07ebfd5b3428b6f4d)
Subnet ID	subnet-07c9cba92b6af9099 (public-cloud-mafia)	Platform details	-
Network interfaces	eth0	Usage operation	-
IAM role	-	Source/dest. check	True
Key pair name	latest-pair	T2/T3 Unlimited	Disabled
Owner	169261769675	EBS-optimized	False
Launch time	April 13, 2020 at 11:47:17 PM UTC+5:30 (25 hours)	Root device type	ebs
Termination protection	False	Root device	/dev/sda1

# Virtual Private Cloud Configuration

The screenshot displays the AWS Management Console interface for configuring a Virtual Private Cloud (VPC). The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a search icon. The right side of the header shows a notification bell, 'APSTCLOUD', 'N. Virginia', and 'Support'.

Below the header, there are buttons for 'Create VPC' and 'Actions'. A search bar prompts the user to 'Filter by tags and attributes or search by keyword'. A table lists VPCs, with the selected VPC 'cloud-mafia...' highlighted. The table columns are: Name, VPC ID, State, IPv4 CIDR, IPv6 CIDR, DHCP options set, Main Route table, and Main Network ACL.

The details for VPC 'vpc-000a6c61df5545ca8' are shown below the table. The 'Description' tab is active, displaying various configuration parameters:

VPC ID	vpc-000a6c61df5545ca8	Tenancy	default
State	available	Default VPC	No
IPv4 CIDR	10.0.0.0/26	Classic link	Disabled
IPv6 CIDR	-	IPv6 Pool	-
DNS resolution	Enabled	Network ACL	acl-081a11e2762ecd19a
DNS hostnames	Disabled	DHCP options set	dopt-0f417630fa6f4d57e
ClassicLink DNS Support	Disabled	Route table	rtb-0d9cd905a78475db3
Owner	169261769675		

# Relational Database Configuration

The screenshot displays the Amazon RDS console interface. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a search icon. On the right, there are links for 'APSITCLOUD', 'N. Virginia', and 'Support'. The left sidebar shows the 'Amazon RDS' menu with options like Dashboard, Databases (highlighted), Query Editor, Performance Insights, Snapshots, Automated backups, Reserved instances, Proxies, Subnet groups, Parameter groups, Option groups, Custom Availability Zones, Events, Event subscriptions, Recommendations (with a '2' badge), and Certificate update.

The main content area is titled 'Instance' and displays the configuration for the 'cloud-mafia-database' instance. The configuration is organized into four columns:

- Configuration:**
  - DB instance id: cloud-mafia-database
  - Engine version: 5.7.22
  - DB name: -
  - License model: General Public License
  - Option groups: default:mysql-5-7
  - ARN: arn:aws:rds:us-east-1:169261769675:db:cloud-mafia-database
  - Resource id: db-WYVGNFPPZJOYIFD4XZ2UBFT5O4
  - Created time: Mon Apr 13 2020 23:17:43 GMT+0530 (India Standard Time)
  - Parameter group: default:mysql5.7 (in-sync)
  - Deletion protection: Disabled
- Instance class:**
  - Instance class: db.t2.micro
  - vCPU: 1
  - RAM: 1 GB
  - Availability:
    - Master username: admin
    - IAM db authentication: Not Enabled
    - Multi AZ: No
    - Secondary Zone: -
- Storage:**
  - Encryption: Not Enabled
  - Storage type: General Purpose (SSD)
  - IOPS: -
  - Storage: 20 GiB
  - Storage autoscaling: Disabled
- Performance Insights:**
  - Performance Insights enabled: No

# Tables in Database

report

email	varchar(100)
path	varchar(1000)
date	date
time	varchar(100)
result	varchar(10)

patients

id	varchar(1000)	PK
email_id	varchar(100)	PK
password	varchar(500)	
salt	varchar(500)	
name	varchar(50)	
age	int(11)	
date of birth	date	
sex	varchar(10)	
blood_group	varchar(10)	
contact	bigint(10)	
address	text	
medical_history	text	
cell_image	varchar(100)	



# 3. Results

—

## 3.1 Results

- We have used 3 CNN architecture, VGG16, Efficient Net B0 and Efficient Net B5.

Model	Train Acc	Test Acc
VGG16	95%	94%
EfficientNet B0	96%	94%
EfficientNet B5	95%	94%

- All these 3 models were ensembled and their cumulative accuracy was 96%.

## 4. Conclusion

—

# 4.1 Conclusion

- Individual accuracies of ConvNet models were low compared to the accuracy of the ensembled network's accuracy.
- Compared to a pathologist our system can generate results faster.
- Proposed system should act as an assistant rather than a decision-maker.

## 5. Future Scope

—

# 5.1 Future Scope

- Our system can be deployed on the portal of hospitals and can be used to assist oncologists to detect cancer.
- Similar models can also be trained to detect different types of cancers if the proper dataset is provided.
- An offline application interface can be created to use our system in the case of the unavailability of an internet connection.
- Such a system would be very useful in rural areas.

## 6. References

- Bardou, D., Zhang, K. and Ahmad, S.M., 2018. Classification of breast cancer based on histology
- Images using convolutional neural networks. IEEE Access, 6, pp.24680-24693.  
M. C. Chun. (2018). Breast Cancer: Symptoms, Risk Factors, and Treatment, Medical News Today. Accessed: Mar. 10, 2018.
- L. He, L. R. Long, S. Antani, and G. R. Thoma, “Histology image analysis for carcinoma. detection and grading,” Comput. Methods Programs Biomed., vol. 107, no. 3, pp. 538–556, 2012.
- P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, “Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies,” IEEE Trans. Med. Imag., vol. 32, no. 12, pp. 2169–2178, Dec. 2013.
- M. Veta, J. Pluim, P. van Diest, and M. Viergever, “Breast cancer histopathology image analysis: A review,” IEEE Trans. Biomed. Eng., vol. 61, no. 5, pp. 1400–1411, May 2014.

Thank You

—