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 seperate 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)

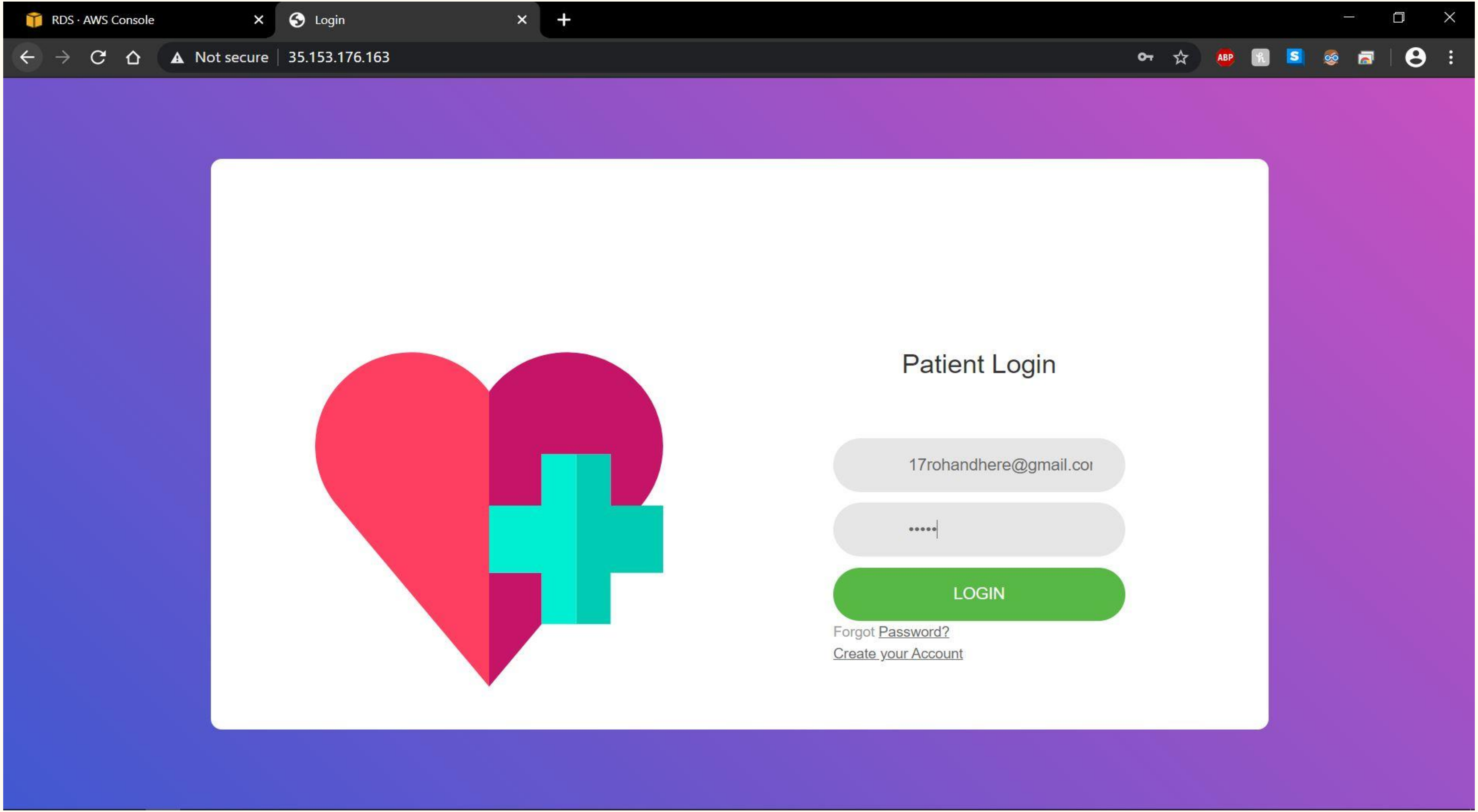
- Front End.
- Dataset.
- CNN Model.
- Back End.

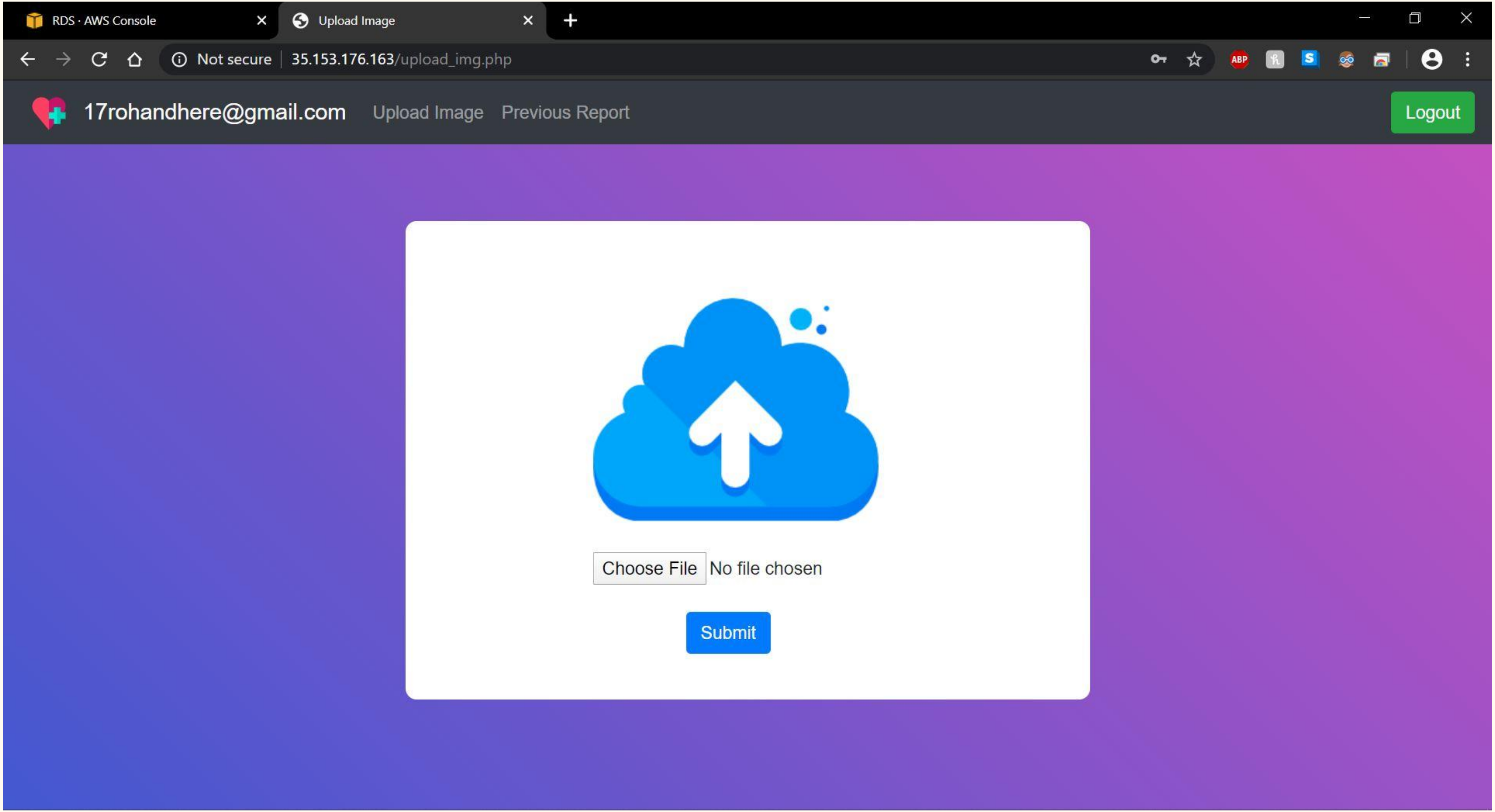
2.3 Description Of Use Case

- 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


RDS · AWS Console
Login
Not secure
35.153.176.163
Patient Login
17rohandhere@gmail.coı
•••••
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

RDS · AWS Console | Report

Not secure | 35.153.176.163/php/upload_img.php

Medical Report

Patitent 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
apsitmlproject@gmail.com

Histopathologic Cancer Report For Patitent : 303ed

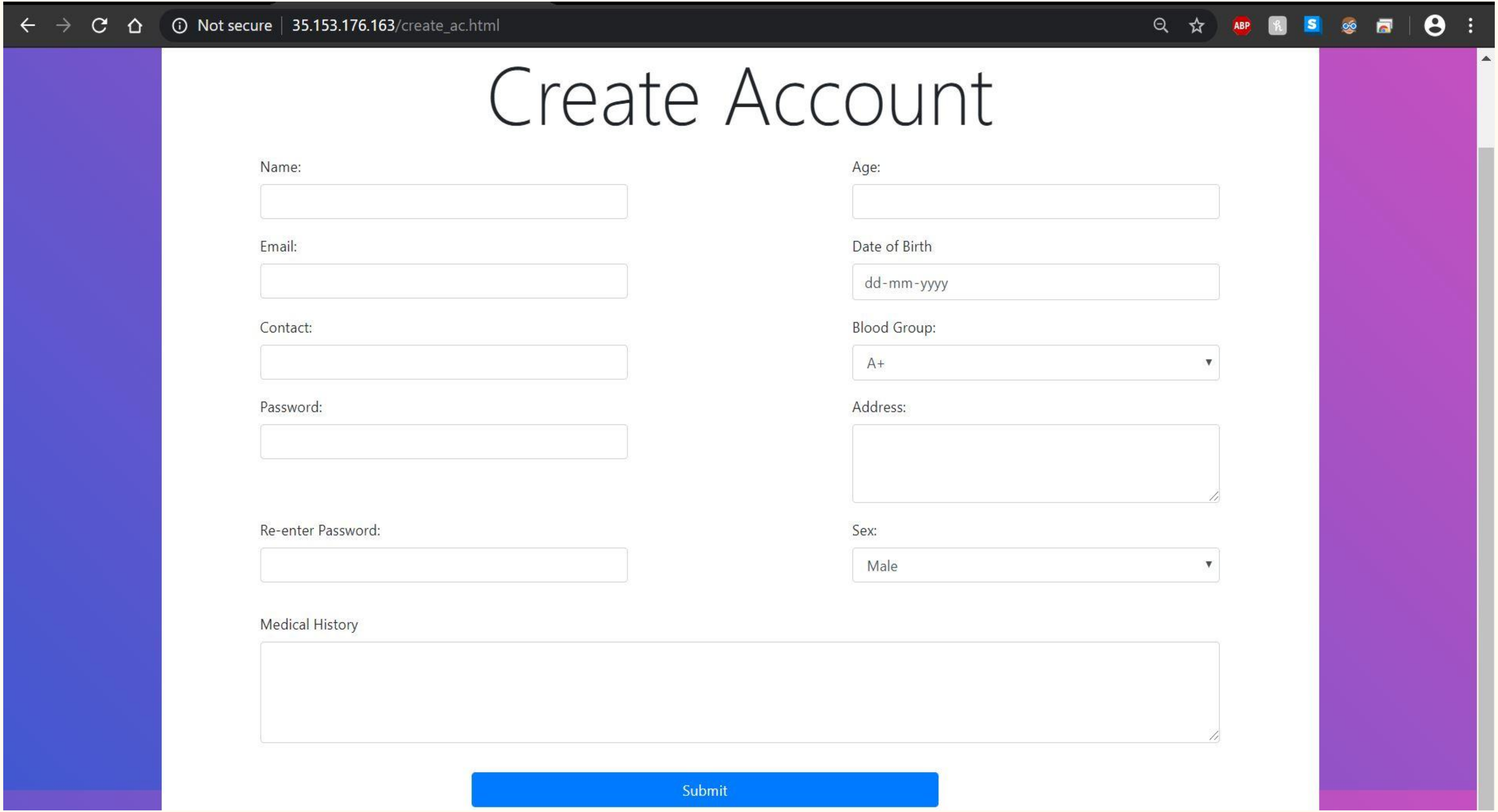
Clinical Diagnosis : Positive
Date of Diagnosis : 2020-04-14

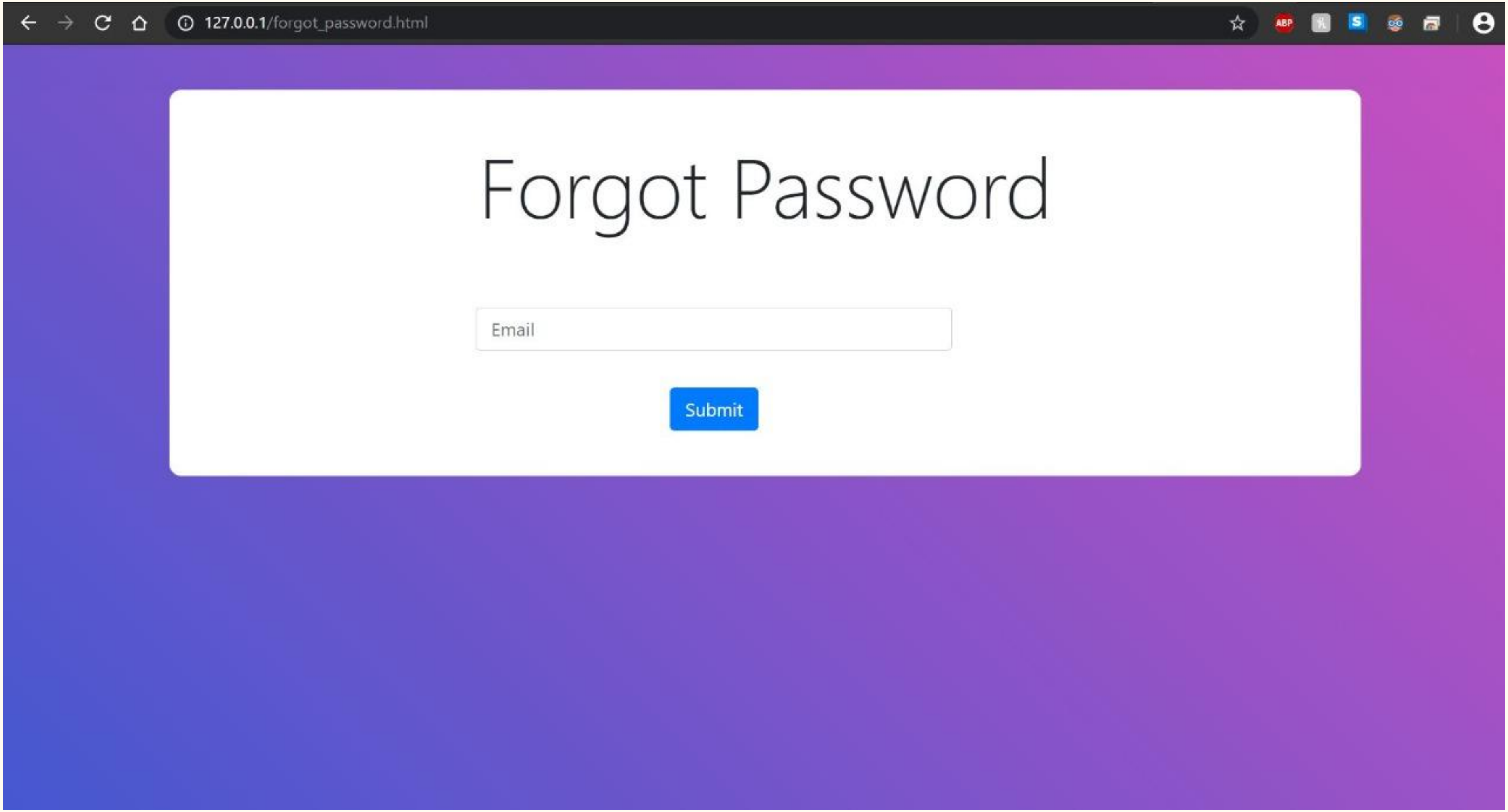
Gross Description:

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.

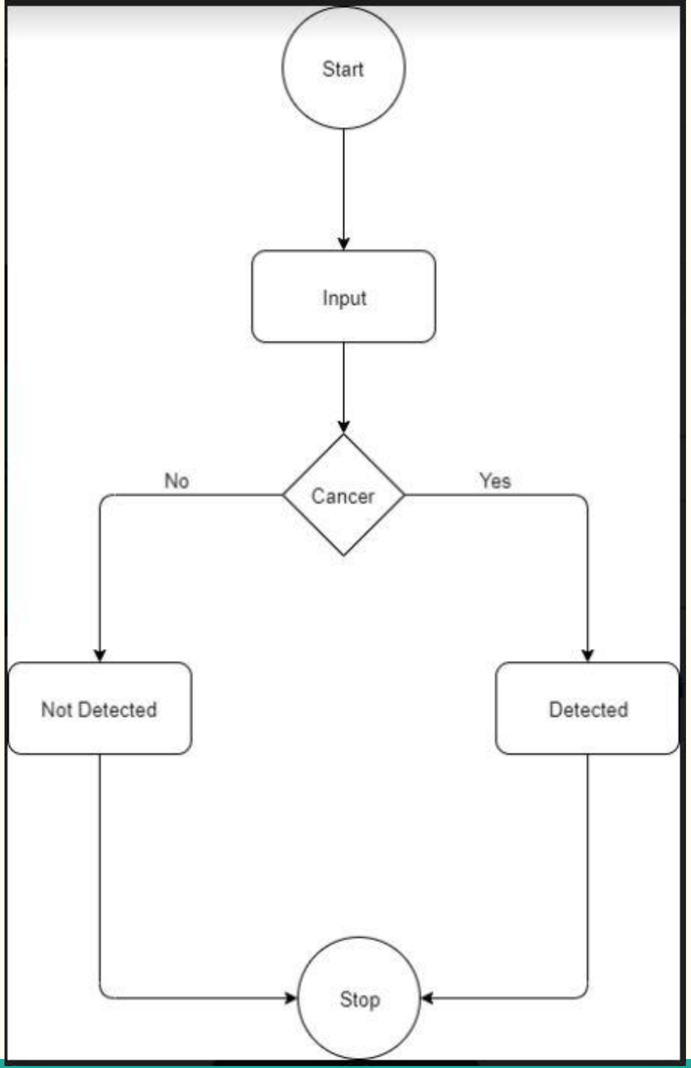
Microscopic:

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.


Not secure
35.153.176.163/create_ac.html
Create Account
Name:
Age:
Email:
Date of Birth
dd-mm-yyyy
Contact:
Blood Group:
A+
Password:
Address:
Re-enter Password:
Sex:
Male
Medical History
Submit

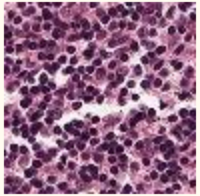

127.0.0.1/forgot_password.html
Forgot Password
Email
Submit

Activity diagram

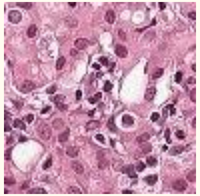


Module 2.6 :- Dataset

- 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

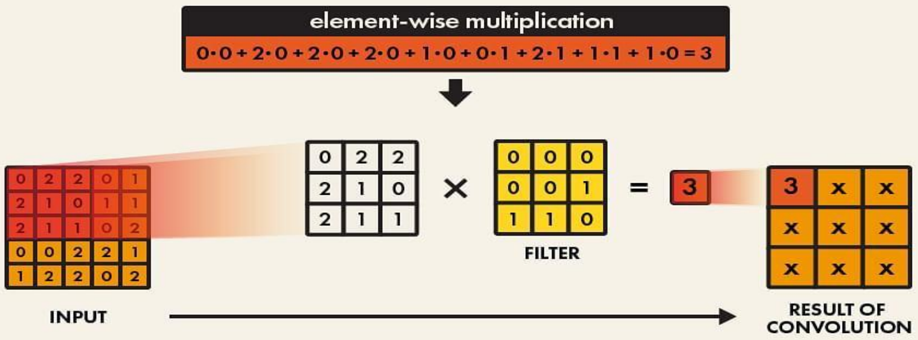
Data Preprocessing

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.

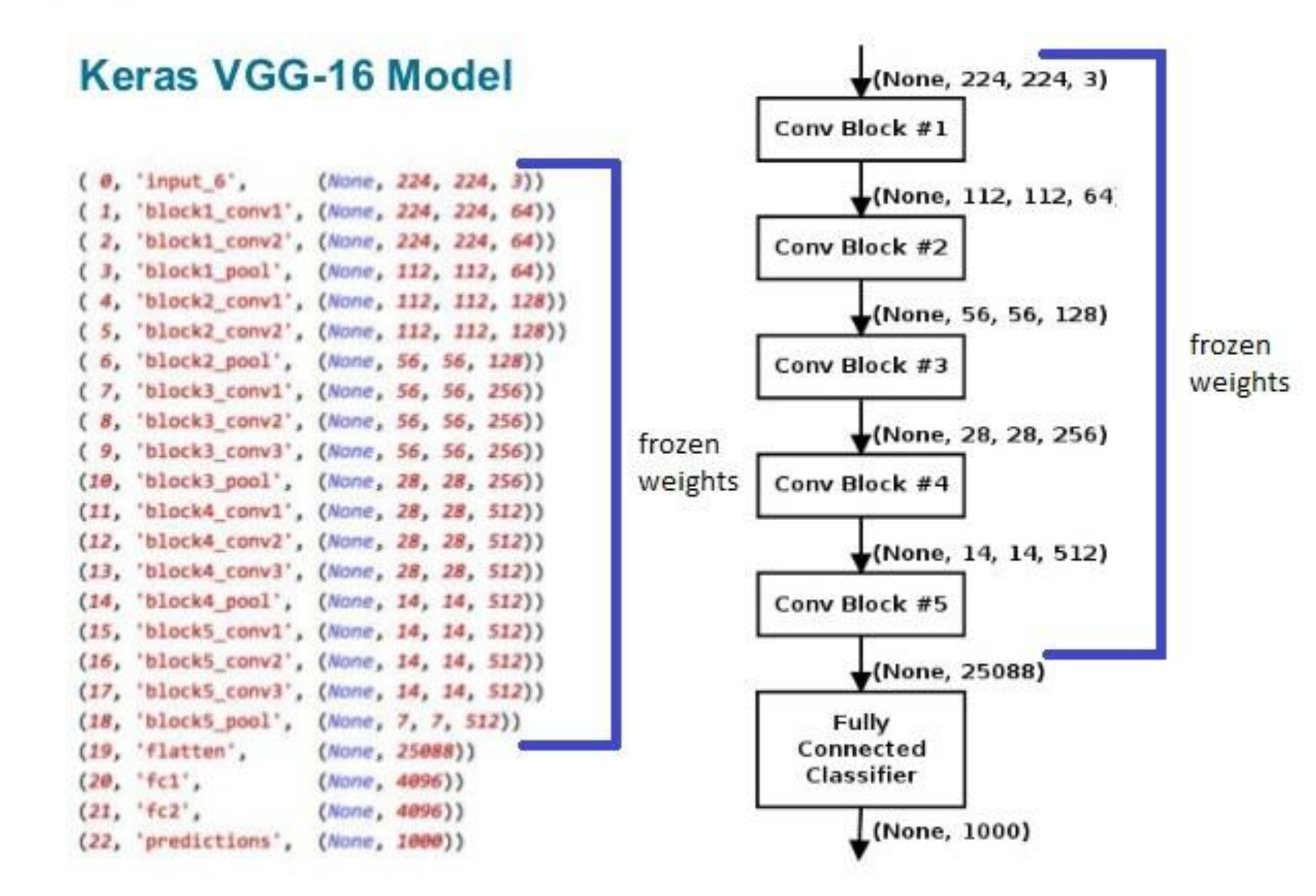
Convolution Function



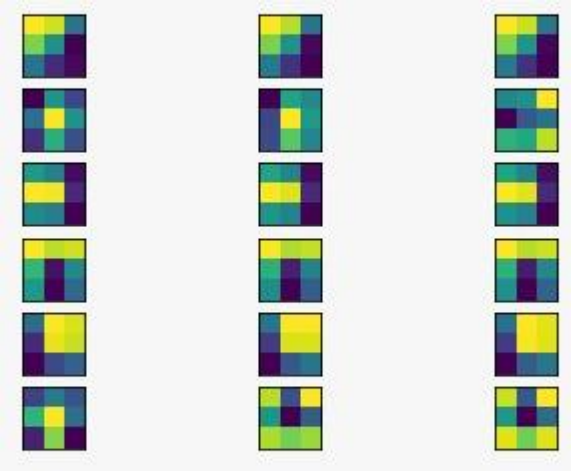
Simple Convolution

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



Visualization



Filters Visualization

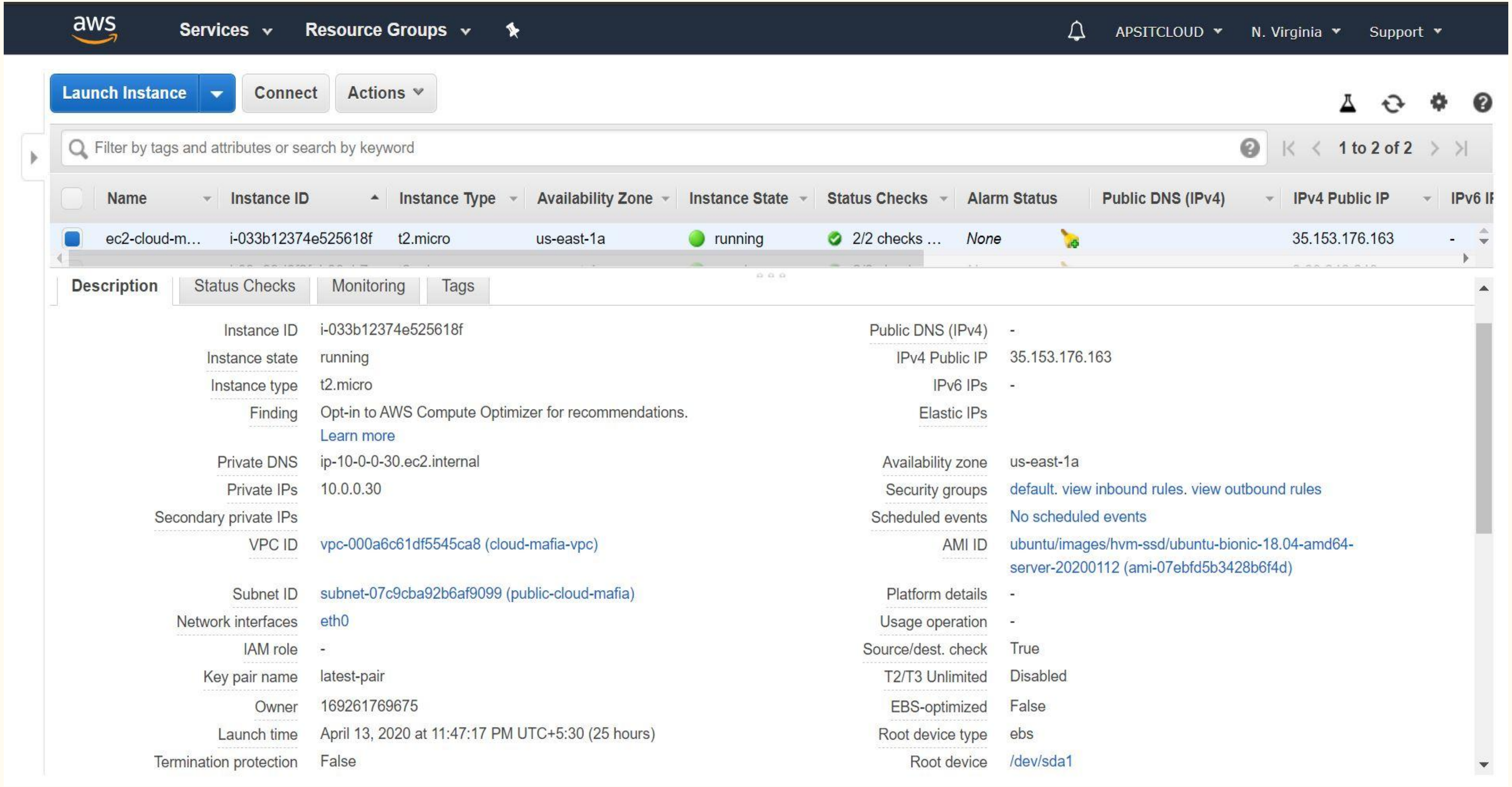
Layers Visualization

Module 2.7 :- Back-end

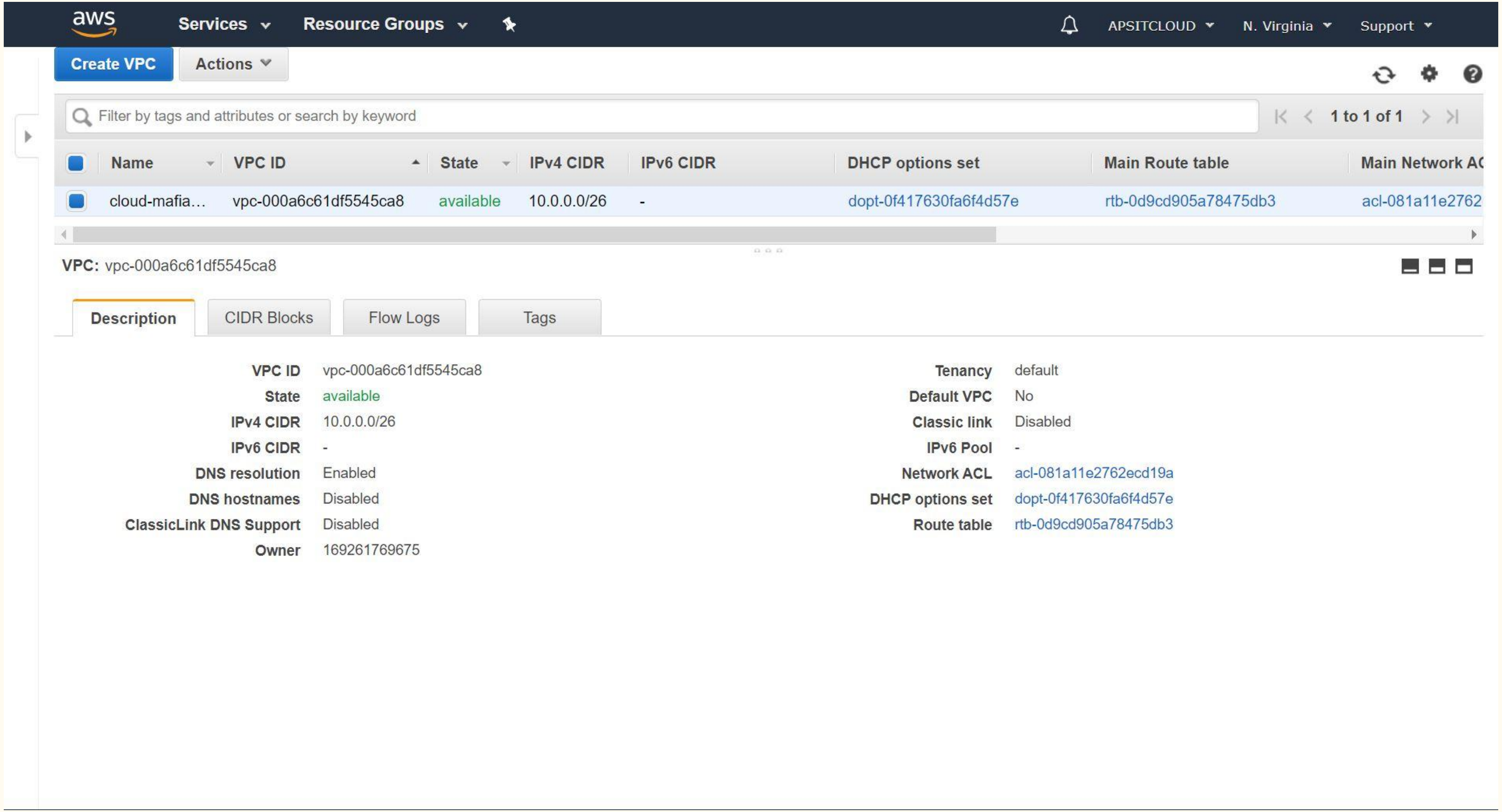
- This part is consists of a web server, model, the database with all the user information and generation of user's medical report.

- AWS Services Used :
    - Ec2 instance (Elastic Compute Cloud).
    - Amazon RDS(Amazon Relational Database Service).
    - VPC (Virtual Private Cloud).

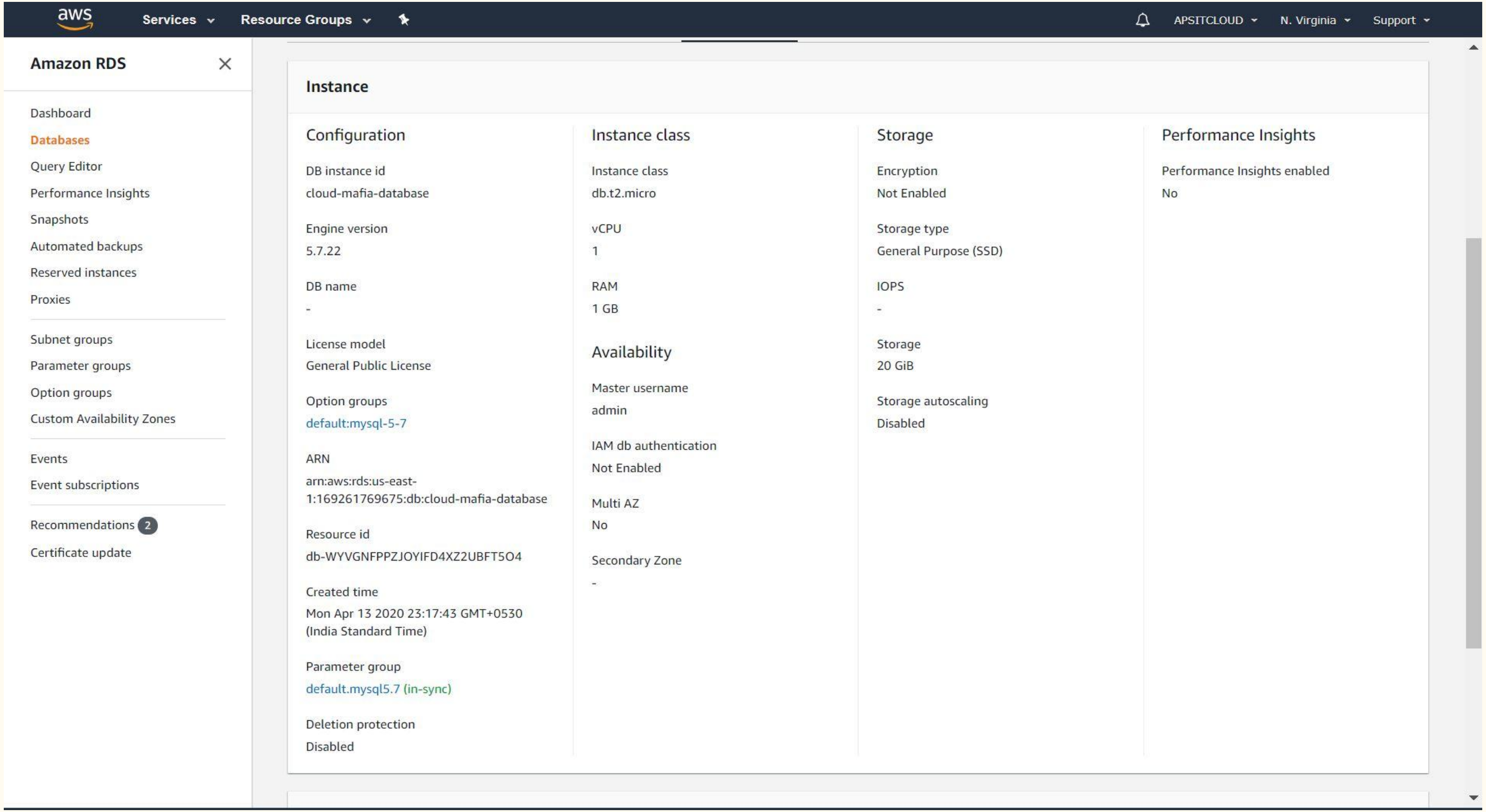
EC2 Configuration



Virtual Private Cloud Configuration



Relational Database Configuration



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, Ecient 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 ensembeled 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