DYSPLASIA DETECTION IN BARRETT'S ESOPHAGUS

A project report submitted to MALLA REDDY UNIVERSITY
in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLGY

in

COMPUTER SCIENCE & ENGINEERING (AI & ML)

Submitted by

Ch.Tharun reddy : 2011CS020073

Under the Guidance of

Prof. M.Ravinder
Asst. Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (AI & ML)

MALLA REDDY UNIVERSITY
(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

2023

MALLA REDDY UNIVERSITY

(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

COLLEGE CERTIFICATE

This is to certify that this is the bonafide record of the application development entitled, "**DYSPLASIA DETECTION IN BARRETT'S ESOPHAGUS**" Submitted by **C.Sahithi (2011CS020071), Ch.Siddhu Vinay (2011CS020072), Ch.Tharun reddy (2011CS020073), C.Ashritha (2011CS020074)** of B. Tech III year I semester, Department of CSE (AI&ML) during the year 2022-23. The results embodied in the report have not been submitted to any other university or institute for the award of any degree or diploma.

INTERNAL GUIDE	**HEAD OF THE DEPARTMENT**
Prof. M. Ravinder	**Dr. Thayyaba Khatoon**
CSE(AI&ML)	**CSE(AI&ML)**

ACKNOWLEDGEMENT

I would like to express my gratitude to all those who extended their support and suggestions to come up with this application. Special Thanks to my mentor Prof.M.Ravinder whose help and stimulating suggestions and encouragement helped me all time in the due course of project development.

I sincerely thank my HOD Dr. Thayyaba Khatoon for her constant support and motivation all the time. A special acknowledgement goes to a friend who enthused me from the back stage. Last but not the least my sincere appreciation goes to my family who has been tolerant understanding our moods, and extending timely support

ABSTRACT

Barrett’s esophagus (BE) is a condition that develops as a consequence of chronic gastroesophageal reflux disease in which stratified squamous epithelium is replaced by metaplastic columnar epithelium which in turn predisposes to the development of adenocarcinoma of esophagus. In this review article, we discuss recent advances in the endoscopic imaging techniques for the detection of dysplasia and early carcinoma in BE. This will include some of the current available novel technologies as well as future applications specifically concentrating on high-resolution endoscopy, narrow band imaging, chromoendoscopy, confocal laser endomicroscopy and autofluorescence imaging.

CONTENTS

INTRODUCTION

1.1 Problem definition

Barrett's esophagus is associated with an increased risk of developing esophageal cancer. Although the risk of developing esophageal cancer is small, it's important to have regular checkups with careful imaging and extensive biopsies of the esophagus to check for precancerous cells (dysplasia) .It is severe disease which is rapidly increasing in western population .Endoscopyis used to detect the barrett’s esophagus .To make it easy we are going to building a model to detectthe disease using machine learning algorithm. BE dysplasia can be performed through manual or automated segmentation through machine learning techniques.

1.2 Objective of the project

- The objective of the project is to detect the dyplasia in barrett’s esophagus by using machine learning algorithm.

1.3 Limitations of project

- Time constraints in learning It is impossible to make immediate accurate predictions with a machine learning system.
- More number of images may also lead to less accuracy of the model.
- Time consuming
- Limited Dataset
- Complicated process

ANALYSIS

2.1 INTRODUCTION

- Barrett’s esophagus is a condition that develops as a consequence of chronic gastroesophageal reflex
- Barrett’s esophagus is mainly caused by obesity.
- BE is defined as a change in the distal esophageal epithelium of any length that can be recognized as columnar type mucosa at endoscopy and confirmed to have intestinal metaplasia (IM) by biopsies
- It is thought that BE progresses in a step wise manner from low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and finally esophageal adenocarcinoma (EAC) which has been attributed to DNA alterations in the mucosa

2.2 EXISTING SYSTEM

- Endoscopy is generally used to determine barrett’s esophagus
- A light tube with camera at the end is passed down and checks the changes in esophagus tissue
- Detection is end up with alot of scanning

2.3 PROPOSED SYSTEM

- It also uses Convolution neural networks (CNN). Its built-in convolutional layer reduces the high dimensionality of images without losing its information.
- Image processing technique are applied to extract features using different parameters of images

Algorithms

- Proposed system makes use of Convolutional Neural Networks within Deep Learning.
- It uses convolution layer, maxpooling ,Activation relu , fully-connected layer .

2.4 SOFTWARE REQUIREMENT SPECIFICATION

2.4.1 Software requirement

Windows operating system will be used during development process. The system will be implemented using in python language. For image processing part we use OpenCV, NumPy and pandas libraries will be used respectively.

2.4.2 Hardware requirement

- System : Pentium i3 Processor.
- Hard Disk : 500 GB
- Monitor : 15” LED
- Input Devices : Keyword, Mouse
- Ram : 4GB

2.5 MODULES

- **Import the required libraries:**

  Here we will be making use of the Keras library for creating our model and training it. We also use Matplotlib and Seaborn for visualizing our dataset to gain a better understanding of the images we are going to be handling
- **Loading the data:**

  Next, let’s define the path to our data. Let’s define a function called get_data() that makes it easier for us to create our train and validation dataset
- **Visualize the data:**

  Let’s visualize our data and see what exactly we are working with. We use seaborn to plot the number of images in both the classes and you can see what the output looks like.

- **Data Preprocessing and Data Augmentation:**

  Next, we perform some Data Preprocessing and Data Augmentation before we can proceed with building the model.

- **Define the Model :**

  Let’s define a simple CNN model with 3 Convolutional layers followed by max-pooling layers. A dropout layer is added after the 3rd maxpool operation to avoid overfitting.

- **Evaluating the result:**

  We will plot our training and validation accuracy along with training and validation loss.

2.6 ARCHITECTURE

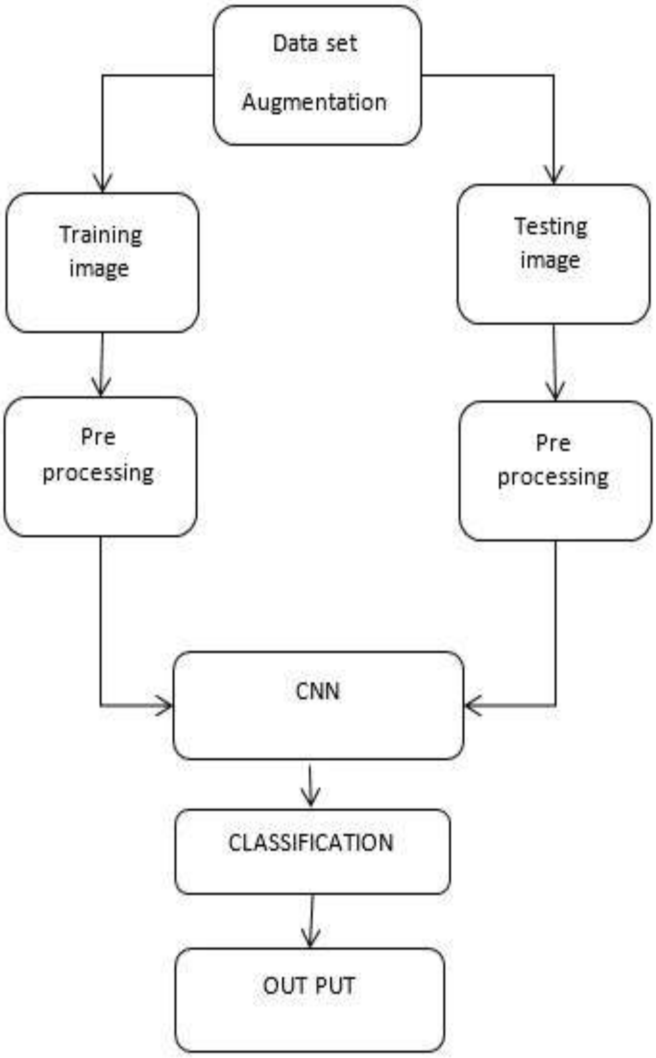


Fig : Process Flow

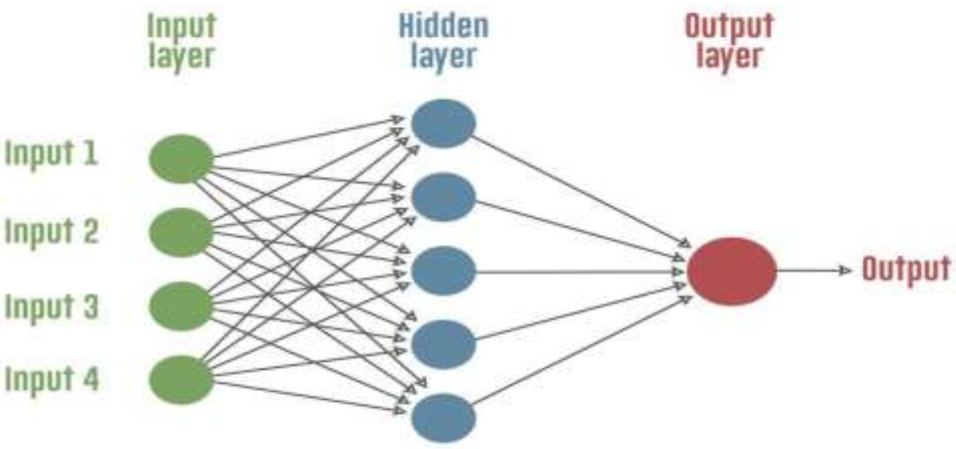
DESIGN

3.1 INTRODUCTION

ML machines perform independently intelligent tasks that should have previously been solved by humans using authority complicated scientific and analytical equipment. This concept of automating combined responsibilities must generate` a great deal of interest in specific networking concerns, approaching every expectation that numerous activities involved in the design and evolution of intelligence arrangements will be offloaded to tools.

Remarkable packages from ML toward special networking fields must previously match the one's expectancies in regions that include intrusion networking regions that have earily coordinated those expectancies in acreage which consist of obtrusion espial, website influx kind, and intellectual radios. Under the aforementioned article, humans use consciousness on ML for ocular chaining among various grid regions. Because of their high capacity, low cost, and a multitude of appealingproperties, optical networks are the basic framework of all massive company networks worldwide.

3.2 UML DIAGRAM



3.3 DATA SET DESCRIPTIONS

In this Project, We have the Dataset containing two types, they are divided into two varieties training and validation images. Further each training and validation have the Forest and Sea satellite images. These are very important dataset and they have their specific function to train the model. In combination of Train and Validation dataset there are a total of 2000 Images. The Image name corresponding to It's label like Forest and Sea are identified.

3.4 DATA PRE-PROCESSING TECHNIQUES

Data Cleaning:

Dirty data can cause confusion and results in unreliable and poor output. Hence first step in Data Pre-processing is Data Cleaning. Cleaning of data is done by filling in missing values, smoothing noisy data by identifying and/or removing outliers, and removing inconsistencies.

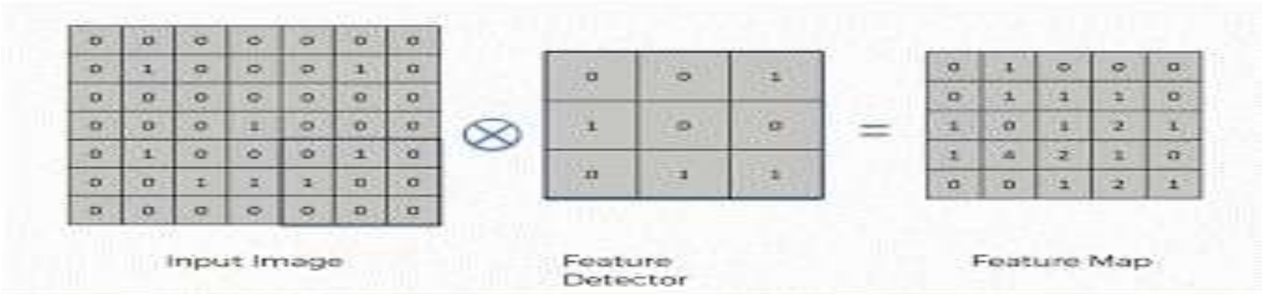
Data Transformation:

Data Transformation involves converting data from one format into another. It involves transforming actual values from one representation to the target representation.

3.5 METHODS & ALGORITHMS

Convolution:

Convolution is a linear operation involving the multiplication of weights with the input. The multiplication is performed between an array of input data and a 2D array of weights known as filter or kernel. The filter is always smaller than input data and the dot product is performed between input and filter array.



Activation:

The activation function is added to help ANN learn complex patterns in the data. The main need for activation function is to add non-linearity into the neural network.

Pooling:

The pooling operation provides spatial variance making the system capable of recognizing an object with some varied appearance. It involves adding a 2Dfilter over each channel of the feature map and thus summarise features lying in that region covered by the filter.

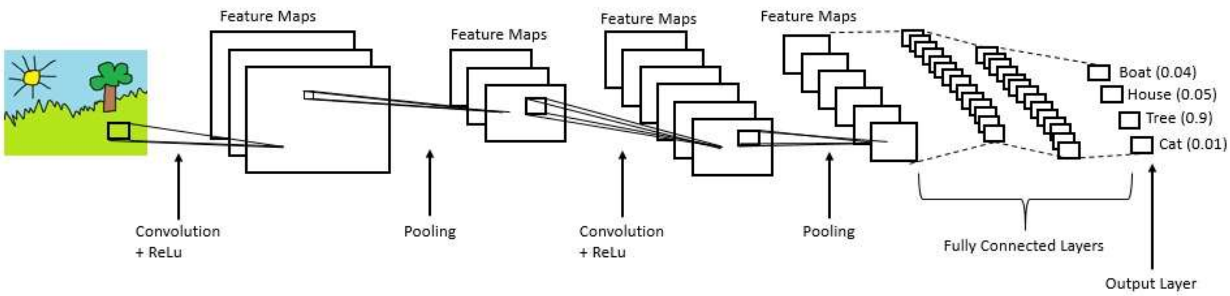
So, pooling basically helps reduce the number of parameters and computations present in the network. It progressively reduces the spatial size of the network and thus controls overfitting. There are two types of operations in this layer; Average pooling and Maximum pooling. Here, we are using max-pooling which according to its name will only take out the maximum from a pool. This is possible with the help of filters sliding through the input and at each stride, the maximum parameter will be taken out and the rest will be dropped.

Fully Connected:

The output from the final Pooling layer which is flattened is the input of the fully connected layer.

The Full Connection process practically works as follows:

The neurons present in the fully connected layer detect a certain feature and preserves its value then communicates the value to both the dog and cat classes who then check out the feature and decide if the feature is relevant to them.



3.6 BUILDING A MODEL

We have used Convolutional Neural Network (CNN). A Convolutional Neural Network is category of deep neural networks, where the machine learns on its own and divide the data provided into the levels of prediction and in a very short period of time gives the accurate results.

CNN stands out among all alternative algorithms in classifying images. Crucial characteristics are sparse Connectivity, Shared Weights and Pooling Feature so as to extract the best features.

3.7 EVALUATION

```
scores=cnn.evaluate(test_set)
```

```
1/1 [==============================] - 0s 195ms/step - loss: 0.3082 - accuracy: 0.9167
```

Here after the evaluation we have got the accuracy of about 90% and loss is 0.2491 which is very less.

DEPLOYMENT AND RESULT

4.1 INTRODUCTION

- Barrett's esophagus is a condition that develops as a consequence of chronic gastroesophageal reflex
- Barrett's esophagus is mainly caused by obesity.
- BE is defined as a change in the distal esophageal epithelium of any length that can be recognizedas columnar type mucosa at endoscopy and confirmed to have intestinal metaplasia (IM) by biopsies
- It is thought that BE progresses in a step wise manner from low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and finally esophageal adenocarcinoma (EAC) which has been attributedto DNA alterations in the mucosa

4.2 SOURCE CODE

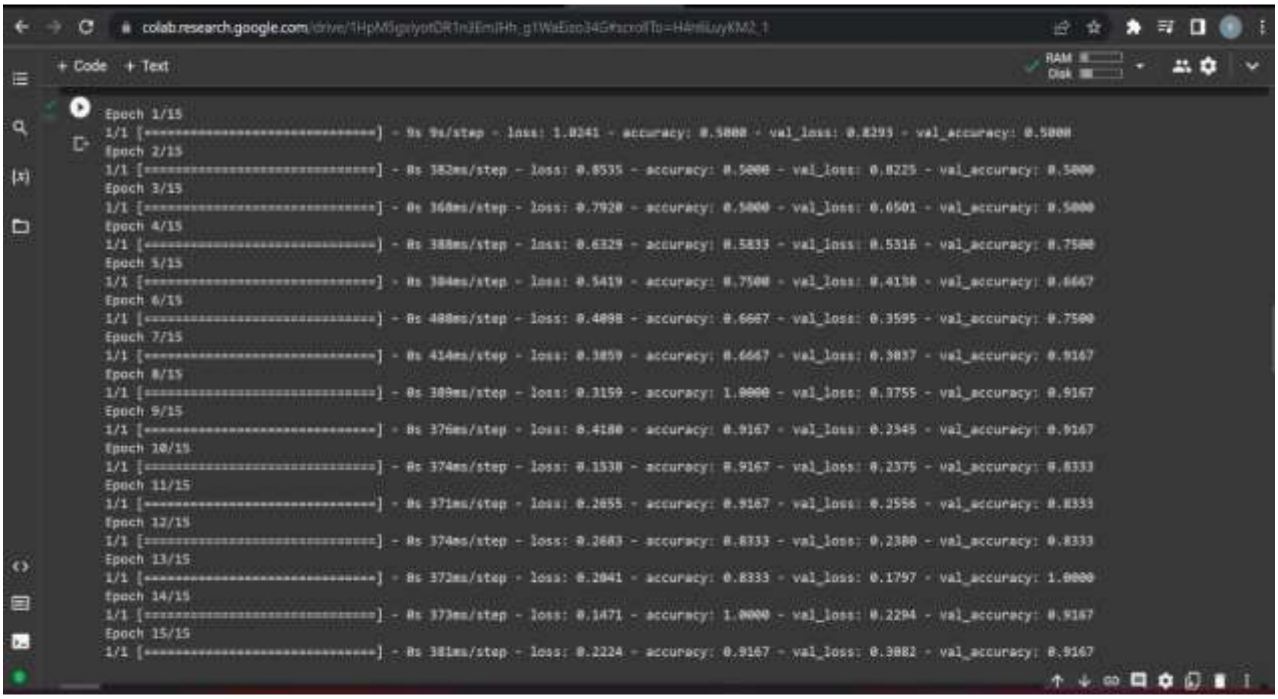
```
        #
import pandas as pd
import numpy as np
import tensorflow as tf

import warnings
warnings.filterwarnings('ignore')

import zipfile
zip_ref = zipfile.ZipFile('/content/project.zip', 'r')
zip_ref.extractall('/content')
zip_ref.close()

from tensorflow.keras.preprocessing.image import ImageDataGenerator

datagen = ImageDataGenerator(rescale=1./255, shear_range=0.2,
zoom_range=0.2, horizontal_flip=True)
```

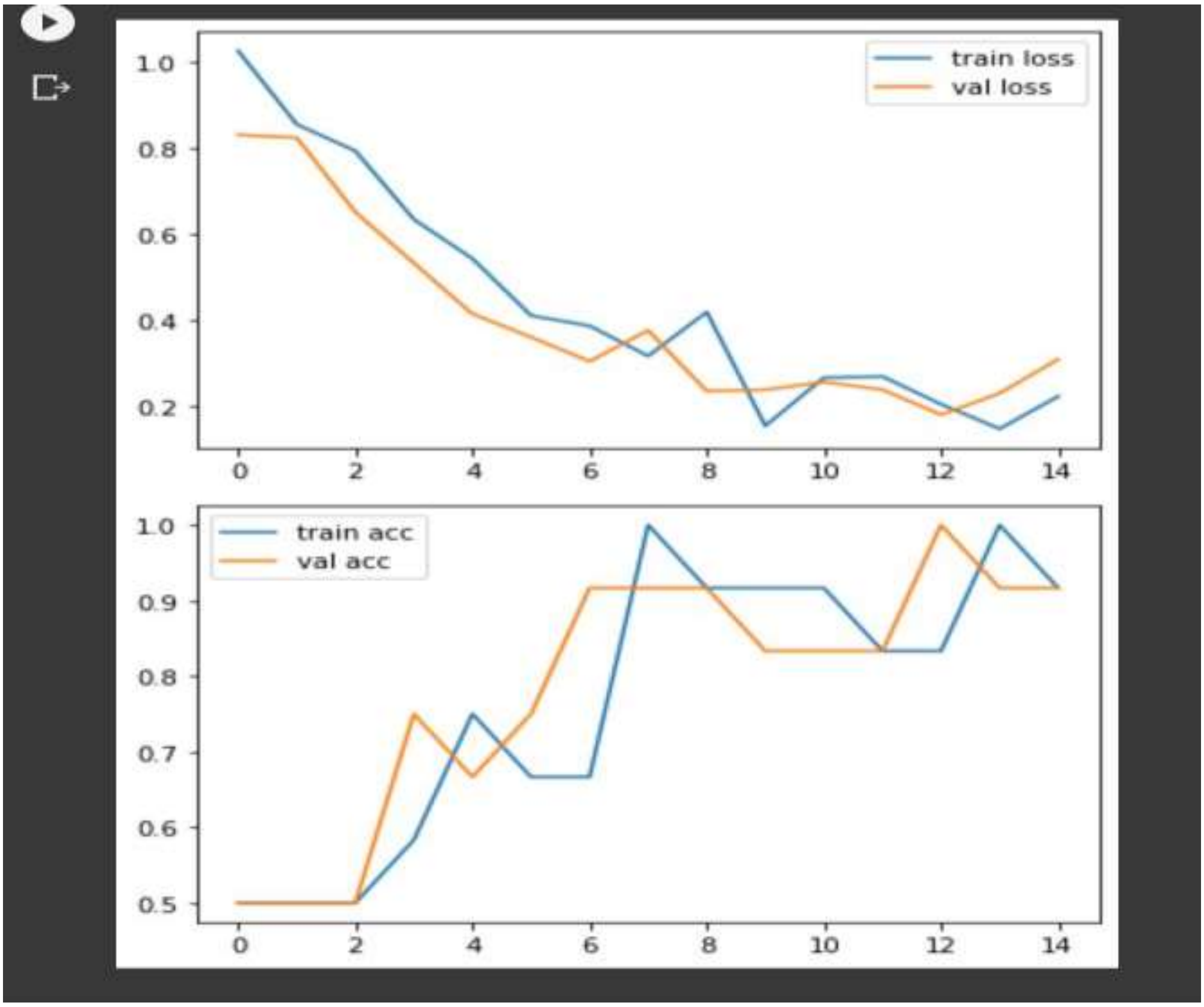
```
training_set = datagen.flow_from_directory(
        "/content/project/Train",
        target_size=(64, 64),
        batch_size=32,
        class_mode="binary"
      )

datagen1 = ImageDataGenerator(rescale=1./255)

test_set = datagen1.flow_from_directory(
        "/content/project/Train",
        target_size=(64, 64),
        batch_size=32,
        class_mode="binary"
      )

from tensorflow.keras.layers import Conv2D
from tensorflow.keras.layers import Dense

from tensorflow.keras.regularizers import l2

cnn = tf.keras.models.Sequential()

cnn.add(tf.keras.layers.Conv2D(filters=32,padding="same",kernel_size=3,
activation='relu', strides=2, input_shape=[64, 64, 3]))

cnn.add(tf.keras.layers.MaxPool2D(pool_size=2, strides=2))

cnn.add(tf.keras.layers.Conv2D(filters=32,padding='same',kernel_size=3,
activation='relu'))
cnn.add(tf.keras.layers.MaxPool2D(pool_size=2, strides=2))

cnn.add(tf.keras.layers.Flatten())

cnn.add(tf.keras.layers.Dense(units=128, activation='relu'))

cnn.add(Dense(1,
kernel_regularizer=tf.keras.regularizers.l2(0.01),activation
             ='linear'))
```

```
cnn.summary()

cnn.compile(optimizer = 'adam', loss = 'hinge', metrics = ['accuracy'])

r=cnn.fit(x = training_set, validation_data = test_set, epochs = 15)

import matplotlib.pyplot as plt
plt.plot(r.history['loss'], label='train loss')
plt.plot(r.history['val_loss'], label='val loss')
plt.legend()
plt.show()

# plot the accuracy
plt.plot(r.history['accuracy'], label='train acc')
plt.plot(r.history['val_accuracy'], label='val acc')
plt.legend()
plt.show()

from tensorflow.keras.preprocessing import image
import cv2
test_image = image.load_img('/content/abe73788-1a7e-410a-a909-
9d70f9323b06.jpg', target_size = (64,64))

plt.imshow(test_image)

test_image = image.img_to_array(test_image)
test_image=test_image/255
test_image = np.expand_dims(test_image, axis = 0)

result = cnn.predict(test_image)

if result[0]<0:
    print("The image classified is esophagus")
else:
    print("The image classified is non - esophagus ")

from sklearn.metrics import
confusion_matrix,classification_report,accuracy_score

Y_pred = cnn.predict_generator(test_set, 6// 32+1)
y_pred = np.argmax(Y_pred, axis=1)
```

```
print('Confusion Matrix')
print(confusion_matrix(test_set.classes, y_pred))
print(accuracy_score(test_set.classes, y_pred)*100)

print('Classification Report')
target_names = ['esophagus','Non esophagus']
print(classification_report(test_set.classes, y_pred,
target_names=target_names))
score = cnn.evaluate(test_set)
```

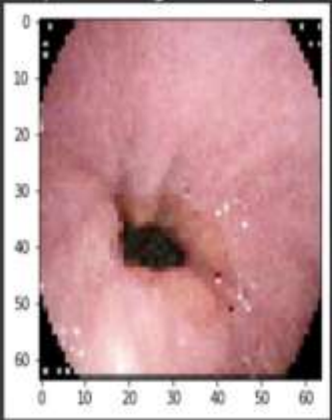
4.3 FINAL RESULT




1.0
0.8
0.6
0.4
0.2
0
2
4
6
8
10
12
14
train loss
val loss
1.0
0.9
0.8
0.7
0.6
0.5
0
2
4
6
8
10
12
14
train acc
val acc

```python
from tensorflow.keras.preprocessing import image
import cv2
test_image = image.load_img('/content/21cf34af-23b5-4b57-8bd8-71a536726796.jpg', target_size = (64,64))

plt.imshow(test_image)
```

```
<matplotlib.image.AxesImage at 0x7f702007b280>
```




```python
if result[0]<0:
    print("The image classified is esophagus")
else:
    print("The image classified is non - esophagus ")
```

```
The image classified is esophagus
```

```
print('Classification Report')
target_names = ['esophagus','Non esophagus']
print(classification_report(test_set.classes, y_pred, target_names=target_names))
```

```
Classification Report
               precision    recall  f1-score   support

    esophagus       0.50      1.00      0.67         6
Non esophagus       0.00      0.00      0.00         6

     accuracy                           0.50        12
    macro avg       0.25      0.50      0.33        12
 weighted avg       0.25      0.50      0.33        12
```

```
scores=cnn.evaluate(test_set)
```

```
1/1 [==============================] - 0s 195ms/step - loss: 0.3082 - accuracy: 0.9167
```

CONCLUSION

5.1 PROJECT CONCLUSION

We learned a great deal in this project, from learning to find image data to create a CNN model that was able to achieve reasonable performance. We also learned the application of transfer learning to further improve our performance.

That is not the end, we saw that our models were misclassifying a lot of images which means that is still room for improvement. We could begin with finding more data or even implementing better and latest architectures that might be better at identifying the features.

5.2 FUTURE SCOPE

That is not the end, we saw that our models were misclassifying a lot of images which means that
is still room for improvement. We could begin with finding more data or even implementing better and latest architectures that might be better at identifying the features

5.3 REFERENCES

- https://www.analyticsvidhya.com/blog/2020/10/create-image-classification-model-python-keras/

- https://www.pantechsolutions.net/satellite-image-analysis-using-convolutional-neural-network

- https://www.researchgate.net/publication/346345285_Classification_of_Satellite_Images