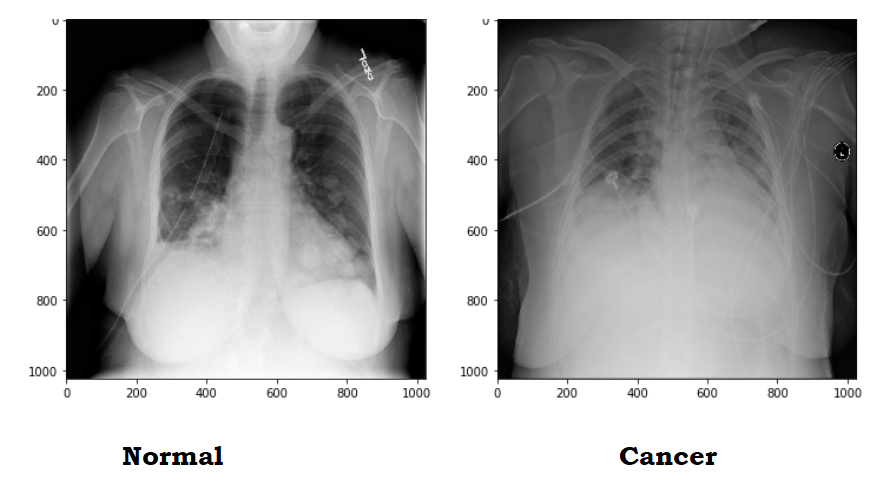
**“LUNG CANCER DETECTION USING IMAGE PROCESSING TECHNIQUES”**

**INTRODUCTION**

Lung cancer is of disease of abnormal cells multiplying and growing into a tumor. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the centre of the chest. Lung cancer often spreads toward the centre of the chest because the natural flow of lymph out of the lung is towards the centre of the chest. Metastasis occurs when a cancer cell leaves the site where it began and moves into a lymph node or to another part of the body through the blood streams. Cancer that starts in the lung is called a primary lung cancer. There are various types of lung cancer such as Carcinoma, Adenocarcinoma and Squamous cell carcinomas. The rank order of cancers for both males and females among Jordanians in 2008 indicated that were 356 cases of lung cancer accounting for (7.7%) of all newly diagnosed cancer cases in 2008. Lung cancer affected 297 (13.1%) males and 59 (2.5%) females with a male to female ratio of 5:1 with a lung cancer ranked second among males and 10th among females. It consists of few stages. The first stage starts with taking a collection of CT images (normal and abnormal) from the available database. The second stage applies the several techniques of image enhancement, to get a best level of quality of clearness. The third stage applies image segmentation algorithms which play a effective role in image processing stages, and the fourth stage obtains the general features from enhanced segmented image which gives indicator of normality or abnormality of images. Lung cancer is the most dangerous and widespread cancer in the world according to stage of discovery of the cancer cells in the lungs, so the process early detection of the disease plays a very important and essential role to avoid the serious advance stages to reduce its percentage of distribution.

**The Challenge** Build an algorithm to automatically identify whether a patient is suffering from pneumonia or not by looking at chest X-ray images. The algorithm had to be extremely accurate because lives of people is at stake.



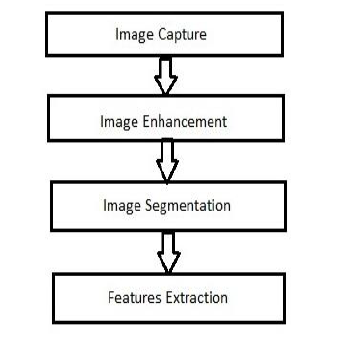
# Data

The dataset can be downloaded from the kaggle website which can be found <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> here.

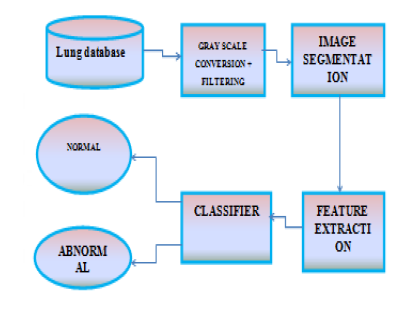
**Environment and tools**

1. [scikit-learn](https://scikit-learn.org/stable/)
2. [keras](https://keras.io/)
3. [numpy](https://www.numpy.org/)
4. [pandas](https://pandas.pydata.org/)
5. [matplotlib](https://matplotlib.org/)

**Flow Chart**



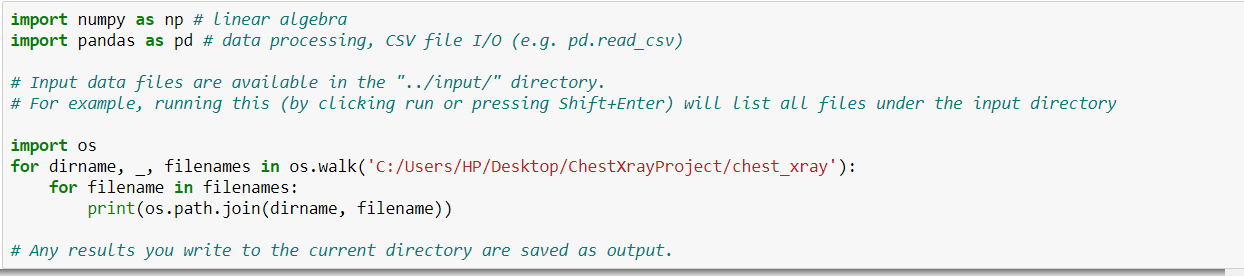
***Figure 1.*** *Lung cancer image processing stages*

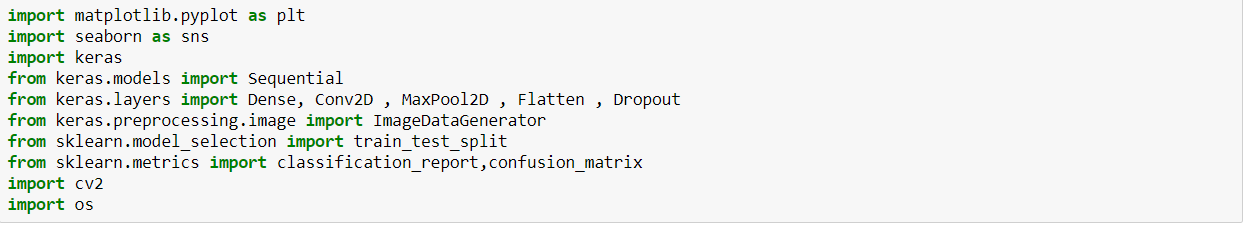


***Figure 2.*** *Block Diagram**Lung cancer image processing stages*

**Source Code**

Lets import required Library at Beginning and read datasets from Target Folder.





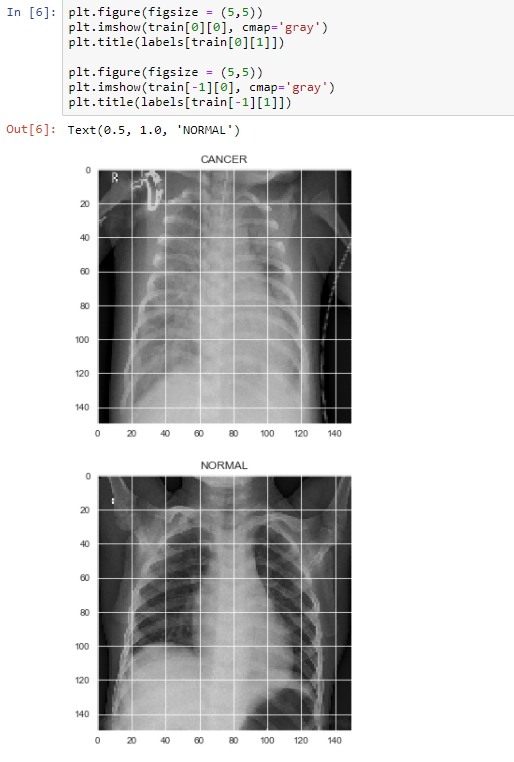
Label Images as Normal and Cancer ,Resizing images and then append it into array.



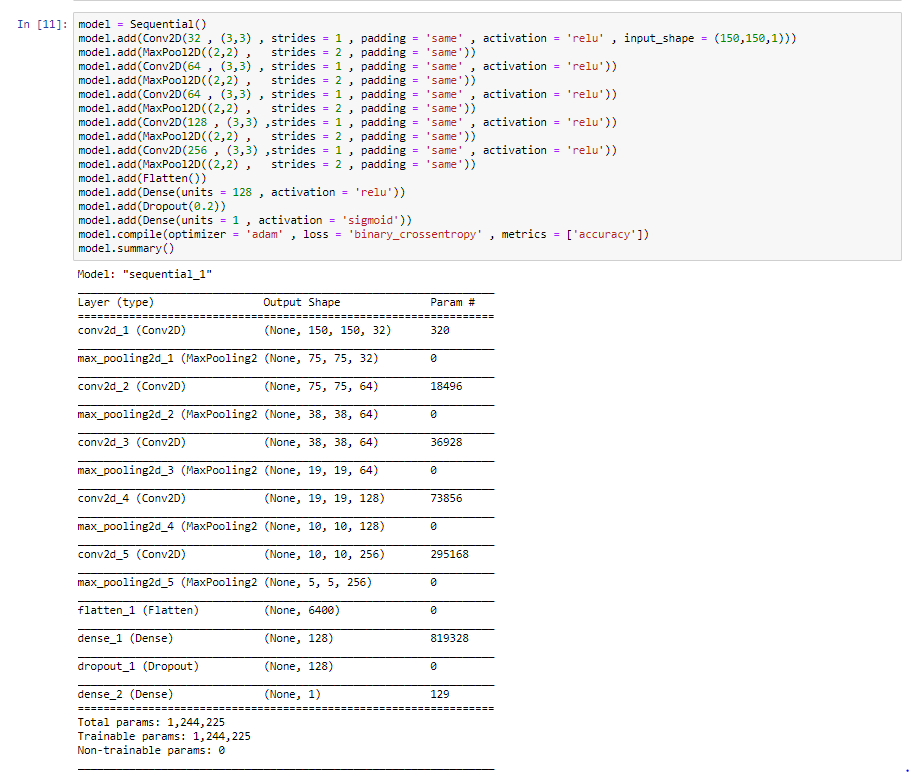
Draws a Graph that Shows number of Cancer and Normal images in Training set

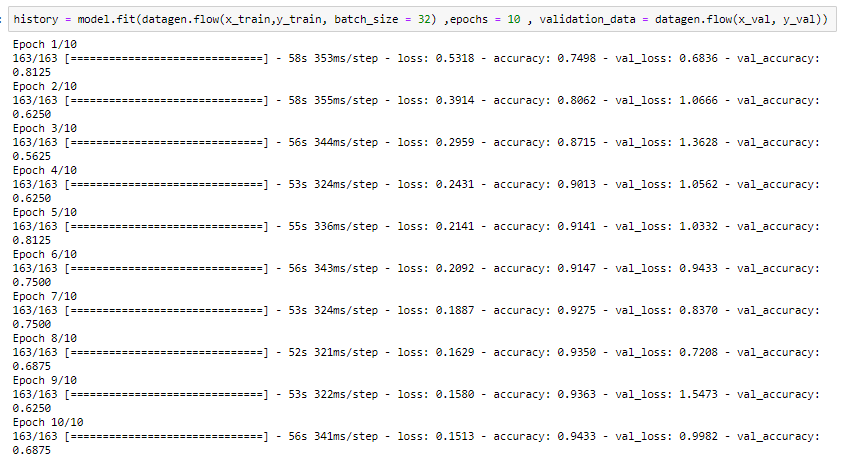


Shows the Difference Between Cancer and Normal Lung X Rays Copy.

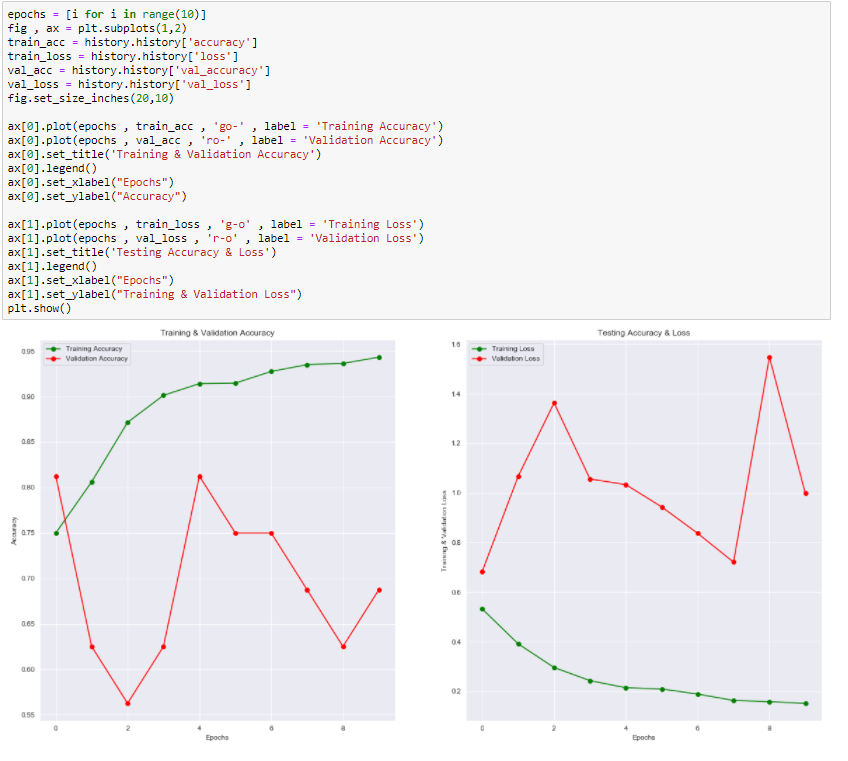


Next I wrote a function in which I did some data augmentation, fed the training and test set images to the network. Also I created labels for the images.The practice of data augmentationis an effective way to increase the size of the training set. Augmenting the training examples allow the network to “see” more diversified, but still representative, data points during training.Then I defined a couple of data generators: one for training data, and the other for validation data. A data generatoris capable of loading the required amount of data (a mini batch of images) directly from the source folder, convert them into training data (fed to the model) and training targets (a vector of attributes — the supervision signal).For my experiments, I usually set the batch\_size = 32. In general a value between 32 and 128 should work well. Usually we should increase/decrease the batch size according to computational resources and model’s performances.

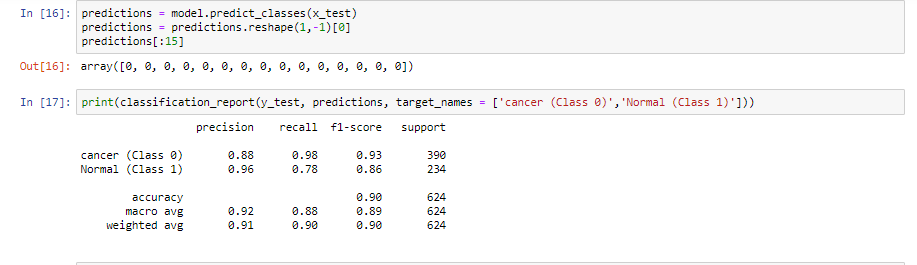




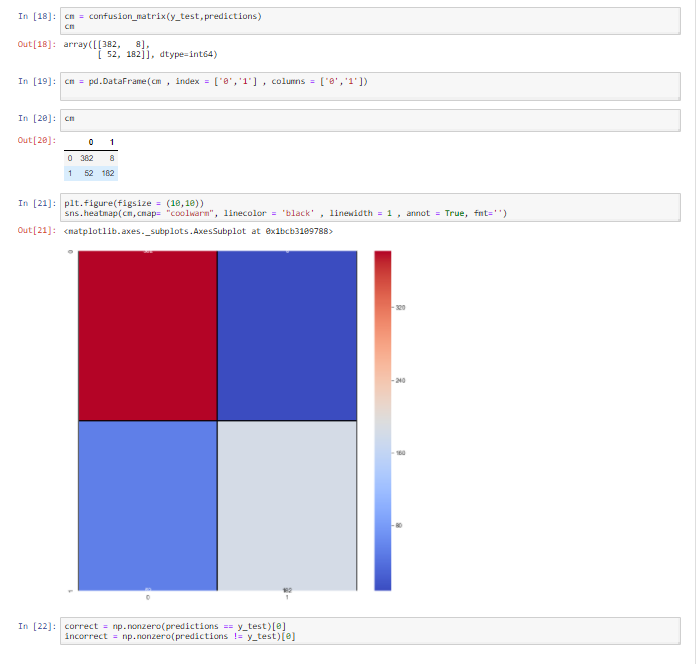
Shows Training Vs Validation Accuracy Graph and Testing Accuracy Vs Loss Graph



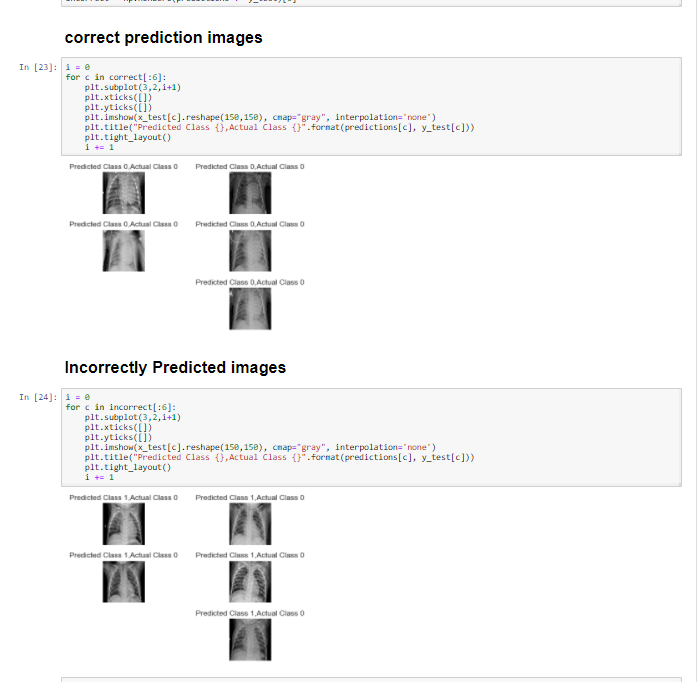
Shows accuracy ,macro average and weighted Average

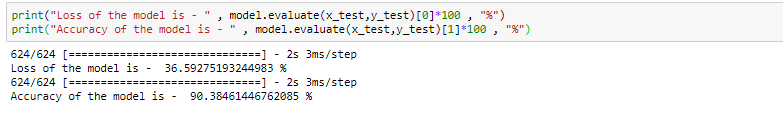


Confusion Matrix Representation with Its output Data



Representation Of Correctly Predicted Images and Wrongly Predicted Images





The model is able to achieve an accuracy of 90.384% which is quite good considering the size of data that is used.

**Advantages**

(a) Early detection of cancer greatly increases the chances for successful treatment.

(b) With the use of this treatment is often simpler and more likely to be effective.

(c)The proposed systems are more efficient and give the better result.

(d)Provides better image quality and accuracy.

**Applications**

[1]It is widely used in many medical areas for early detection of cancer .so the the proper treatment will be provided to the patient.

[2]This image processing technique can also be used to detect other cancer such as breast cancer and tumor in our body part.

# Conclusions

I have demonstrated how to classify positive and negative pneumonia data from a collection of X-ray images. The model was made from scratch, which separates it from other methods that rely heavily on transfer learning approach. In the future this work could be extended to detect and classify X-ray images consisting of lung cancer and pneumonia. Distinguishing X-ray images that contain lung cancer and pneumonia has been a big issue in recent times, and our next approach should be to tackle this problem.

# References

[1] Febr Mokhled S. AL-TARAWNEH, “Lung Cancer Detection Using Image Processing Techniques”, Leonardo Electronic Journal of Practices and Technologies, June 2012.

[2] Muhammad Usman, Muhammad Shoaib and Mohamad Rahal, “Lung Cancer Detection Using Digital Image Processing”, PIERS Proceedings, Stockholm, Sweden, Aug. 12-15, 2013.