**Pathology Image Analysis For Lung Cancer Classification Using IBM Watson**

**Problem statement:** The main objective of this project is to find the tumor present in the patient's lung using CNN.

**Software specifications:** Model was constructed on Jupyter Notebook and app was made using Spyder on Anaconda Navigator.

**Dataset and data collection:** This project depends a lot on data. And for that, we have to collect data. CNN also requires images, which then introduces the concept of trainset and testset. It is basically the data set which is used to train the model for performing the objectives of the app. The trainset will be used to building the model and the testset will be used for validating the model.

**Pre processing the images:** As mentioned above, we have collected the necessary data requireed for the model to be built and validated. We can now use this data to train the model but before moving to that step, we have to preprocess the images. For this we use Keras ImageDataGenerator. The whole process of Image Pre-processing contains 3 main steps which are to import the ImageDataGenerator library, to configure its class and to apply it to the collected trainset and testset.

As we import the keras library to the python script, we have to define the parameters for ImageDataGenerator class. The parameters/arguments include adjustments like to rescale, shear\_range, rotation range and zoom range of the image, etc.

After initializing the preprocessing features, we now have to apply those features to the dataset which we have collected.

**Building of the Model:** With the above steps completed, which are acquiring the augmented and pre-processed image data. We can proceed further by beginning the process of building our model. It consists of certain steps like importing model building libraries, initializing the model, configuring, training and testing the model, etc.

We import libraries such as sequential, dense, flatten, etc. Using all of that, we initialize the model. Now as the data is defined and initialized, we configure the learning process by compiling the Sequential model class. After obtaining the trained data and fully configured network, we train the model. We pass data to the model for the training process to begin. In the end, the model is saved.

**Application building:** We build an application that is integrated to the model which we have built and obtained above. Our HTML code consists of the webpage which has the introduction and the image uploader where the tumor will be identified.

A python file is created using spyder and we firstly import necessary libraries, flask module. We then load files used for detection of the disease from the image which we will upload. Finally, we render the html page and bind the url which will then perform the action.

In order to run the app, we use anaconda prompt to run the python app.py and navigate to the localhost to find the webpage of the app.

**Training the model on IBM:** Firstly we register on IBM cloud and we then train the same model which we obtained above on IBM Watson.

**Conclusion:** The model and app are built using Jupyter Notebook using datasets and lastly the model is run on IBM Watson.

**Referrences:** github.com ,

  stackoverflow.com