Idea/Approach Details

Technology Bucket	: Life Sciences	Category	: Software
Company Name	: Sun Pharmaceutical Industries Ltd.	Problem Code	: DD1
Team Leader Name	: Shrey Jasuja	College Code	: U-0841

Solution Proposed

There are four types of chromatography methods used by Sun Pharma namely:
1.Ionic Chromatography (IC)
2.Gas Chromatography (GC)
3.High Performance Liquid Chromatography (HPLC)
4.Ultra Performance Liquid Chromatography (UPLC)

The idea is to make a software which is able to predict whether the chromatography output is in the range of the standard output or not. The best way to do that with high accuracy is using Image Recognition method. We are using Convolutional Neural Network Models for Image Recognition which would be able to give high **accuracy of 99% - 99.5%** for the chromatography output.

Overview of Software

We are designing a software which will be made using **tkinter**, Python's Graphic User Interface. It will have a database of all the drugs whose chromatography test is to be done.

It will first ask the user for the type of chromatography method and the drug for which the output range is to be checked. After selecting the drug and method, our software will ask for the pdf file of the chromatograph.

The software will output the **result** whether it is within range or not and will also be able to generate a report on it.

Solution Proposed

Working of Software:

For predicting whether the chromatography output is within the range or not, it will first extract the pdf files for the drug from the dataset. After that, it will convert it into images and finally into array using Numpy and Pandas library. The data-set will be separated in two labelled datasets i.e. “Within range” and “Not in Range”.

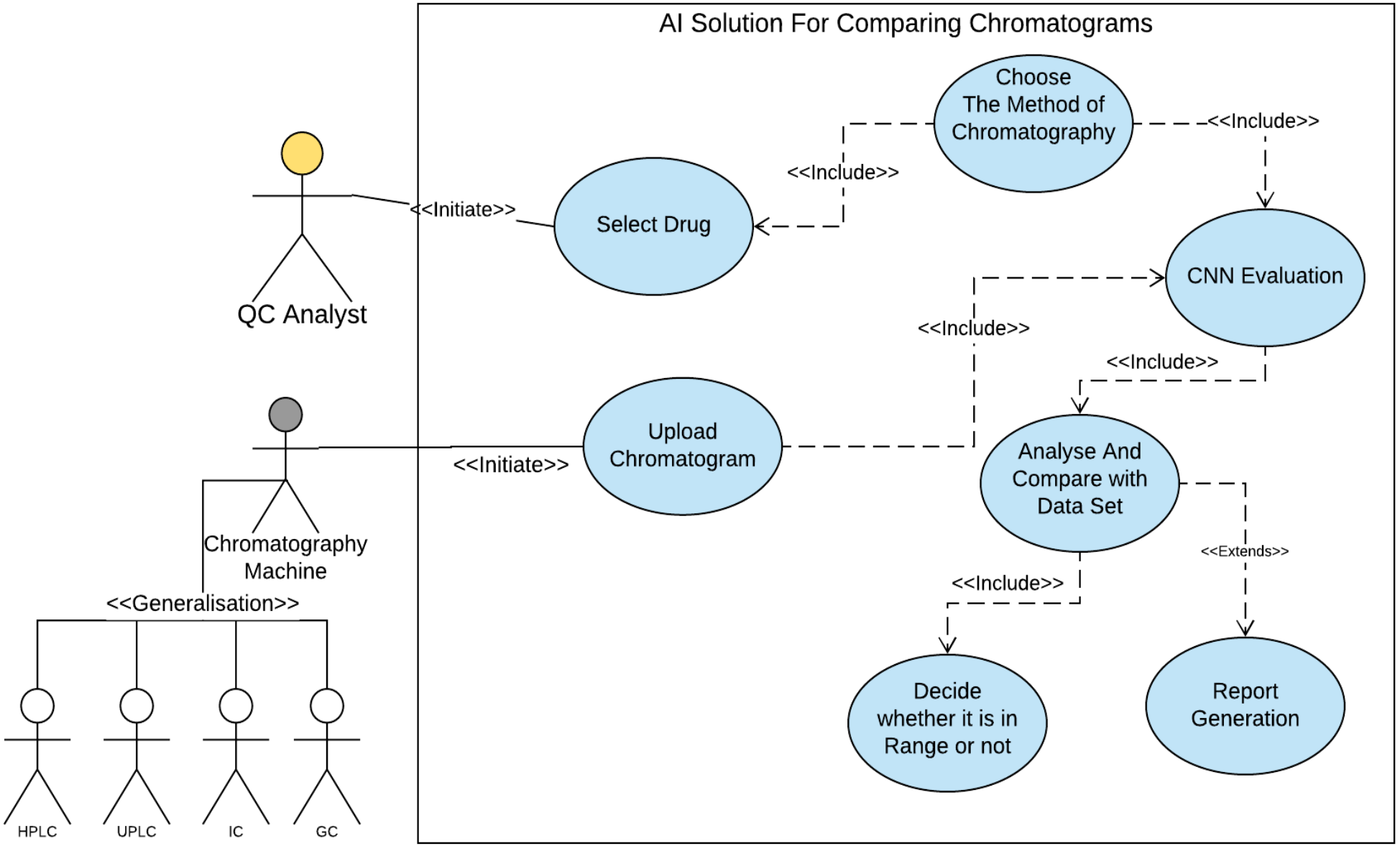
Using Keras and Tensorflow, we will train software to correctly predict the output result. Further, activation functions, optimisers, learning rate annealer and data augmentation will be used to increase the accuracy.

Finally, when a chromatograph output is fed to the software, it will predict whether the output is within the range or not.

Salient Features:

- Easy and convenient UI
- High Accuracy
- Feature for the addition of New Drug
- Shareable to other platforms like Mail, Google Cloud, AWS

Use Case Diagram



Technology Stack

- Image Recognition Model
- Python and its libraries:
  - Tkinter
  - Keras
  - Tensorflow
  - Numpy, Pandas, Matplotlib
  - Scikit Learn

Dependencies / Show Stopper

- Labelled Chromatogram Output Dataset for each drug.
- Guidance of QC Analyst