**User Manual:**

The project folder contains below mentioned folder and files.

Programs folder: It contains all the programs.

Data folder: It contains all the data files.

Presentation ppt: Project PPT

Result\_analysis excel sheet.

IEEE Paper : FinalReport

The Programs folder contains all the MATLAB programs and python program.

**Step1**: rfeprogram.py contains the code for Recursive feature elimination technique.

To run it, open the file in any python IDE, input the number of features to be selected at Line number 29 in the program. Input the training data filename and training label file name at line numbers 13 and 14 in the program.

After the program runs successfully, type selector.get\_support(‘true’) in python shell. It prints an array of indices of features. Index starts from zero.

**Step2:** generateFCancerData.m takes trainingdata file and selected features indices file as inputs and outputs the training data file with only selected features.

**Step3:** accuracy\_rfe.m takes above generated training data set file and training label file as input and applies 10-fold cross validation logic and outputs the predicted label set.

**Step4:**assessmentRFE.m takes original labels file and predicted labels file as inputs and outputs all the assessment metrics discussed in report.

generateFCancerData.m, accuracy\_rfe.m and assessmentRFE.m are MATLAB programs.

**Dataset files:**

Project zip has a folder names Data. It contains all the data files and predicted label files.

The naming convention followed for the data files is explained below.

cancerdata.txt: it has the complete training data with all the features. It has 569 samples and 30 features.

Cancerlabels.txt: It has 569 labels corresponding to 569 samples.

**n\_sfindex:** this file contains the selected feature numbers. For example, 5\_sfindex contains the top 5 selected feature numbers.

**n\_fcd:** this file contains the training data with n features only. For example, 5\_fcd contains the training data with 5 selected features only.

**Data** folder contains another folder named **testLabels**. This folder contains the predicted label files. For example, file name **n\_fcd\_testlabels** contains the predicted labels for the training data with only **n** features.