Step-by-step guidelines to Run the AutoNSGACytoNet framework:

1. Collect data from **https://xenabrowser.net/datapages/: Collect GDC TCGA gene expression RNAseq TPM data for BRCA, LAML, LUAD, LUSC, COAD, SKCM, GBM and LIHC Pancancer Cohorts.**
   Already downloaded data can be found in: dataset

2. **The Dataset folder contains directory-wise carcinogenic Data. Here, each directory is named after each cancer type, and each directory contains .tsv file format for each cancer type. These initial tsv files need to be Transposed so that we can use the genes as features and merge together to result in a final Dataset. Using the merging script, the final dataset can be achieved.**

PS: Step 2 has one directory-related dependency. We need to set the directory/path of Dataset manually to run the code.

3. **Step 3 is the preprocessing step. Data in file can be used or generated using the merging script. The notebook Preprocessing Data can be used to create preprocessed CSV files. It also has director/path dependency. Need to set the path manually. The outcome is already stored as Preprocessed_8_cancer_genes.csv in Google Drive.**
4. **Step 4 is the autoencoder-based feature selection step. Here, the input is as Preprocessed_8_cancer_genes.csv file. After running the notebook autoencoders-with-cv.ipynb we will get out csv files top_0.5_percent_features_cv.csv, top_0.25_percent_features_cv.csv, top_1.0_percent_features_cv.csv. Again, this notebook has a directory/path dependency. We need to change accordingly.**
5. **The next step is the NSGA-2 Step. This step is implemented in nsga2-with-rf-cv2.ipynb notebook. It will take these files (Preprocessed_8_cancer_genes.csv, top_0.5_percent_features_cv.csv, top_0.25_percent_features_cv.csv, top_1.0_percent_features_cv.csv) as input. We need to set the path manually here as well.**
6. **After running the notebook nsga2-with-rf-cv2.ipynb, we will get three output files NSGA2_77_compression_1.csv, NSGA2_308_compression_3.csv, NSGA2_154_compression_2.csv with gene subsets.**
7. **At step 7, we will need the gene subset from NSGA2_308_compression_3.csv file. We will upload the gene set in STRING database to generate Protein-Protein-Interaction PPI network. The network needs to be downloaded in .tsv format. Which is also available in Google Drive as string_interactions.tsv**
8. **At step 8 we need to upload the string_interactions.tsv network in Cytoscape software with cytohubba dependency. It is an open-source tool. At this step, we will select 13 hub genes. These genes will be used for evaluation purposes.**
9. **This is the final evaluation step. Her,e we need to run the classification.ipynb notebook. This notebook requires Preprocessed_8_cancer_genes.csv Dataset. The gene subset acquired from Cytoscape is already hardcoded in the notebook. After running the notebook, we will get desired evaluation metrics.**