Work Flowchart (DRAFT - ABRIDGED):
Machine Learning Based Prediction To Detect Anti-Cancer Polypeptides Utilizing Deep Learning, Support Vector Machine Sigmoidal Function and Naive Bayes.

Start

Obtain Tyagi-B (Anti-Cancer Peptide) dataset & Controls

Exploratory Data Analysis -Testing
1. Investigate: 1st, 2nd 3rd moments of AA & DP,
2. Check for Near Zero Variance; feature extraction,
3. Use Boruta RF; feature selection,
4. Check Pearson's Correlation Coefficients for multi-collinearity; feature extraction,
5. Conduct Principle Component Analysis; feature selection
6. K-Means Clustering for EDA

Pre-process 'Anti-Cancer Peptide' & Control dataset based on EDA to obtain 'Working' dataset.

Model Tuning
Scout experimental space of hyper-parameters for the machine learning methods.

Model Generation and Measure Performance
Use repeated cross-validation to evaluate predicted classes and probabilities
using hyper-parameters from Model Tuning.

Identify Miss-Classified Observations
Extract false positive and false negatives from Model Generation.
Investigate specific polypeptides which are miss-classified

Discussion of Results

Discussion of Conclusion