Code Listing:

1. dataset_manager.py:

   Converts Cryo-EM maps (.mrc files) to datasets: training sets and validation sets.

2. model.py:

   Contains the convolutional neural net model and functions to train, evaluate, load and save the model.

3. helix_prediction.py:

   Splits an input protein to 32x32x32 patches, runs the model on them and returns a helix prediction for the whole protein.

4. helix_predictor_UI.py:

   Manages the user interface of the program.


We are submitting two more code files, which we used to generate the datasets. Those scripts are not part of any pipeline in the application and are not required in order to use it.

5. get_pdb_datasets.py:

   Crawls over the all alpha and a/b classes on https://scop.berkeley.edu/, and downloads .pdb files of proteins.

6. get_raw_data.py:

   Script that runs inside Chimera and synthesizes Cryo-EM map and helix mask for proteins (.pdb files)