# Process of getting training dataset from shapefile

To automate the process of getting the image for the features and labels, FeatureClip.py was developed to use the shapefile with the polygons of the features and generate the images that are needed for the machine learning phase.

There are 5 parameters that the user is required to provide: checkboxlabel, datafolder, infc, saveDir, and workspace. Since the image for features and labels are different, checkboxlabel is a Boolean variable that indicate the type of input feature class that the program is handling. True will represent that the input feature class is a shapefile containing polygons of the label of the features, which only generates an image of historical topographic map (HTM) for each polygon; false will represent that the shapefile contains polygons of the feature itself, which generates image of elevation data, NAIP imagery, and Historical Topo Map. Datafolder is a variable that indicates where the user stores their unclipped DEMs and HTM images if they have any; it is also used as the output folder after the program downloads the data from The Nation Map API (TNM API). Infc is a variable of the path for the feature class file that contains the feature polygons or the features’ label polygons. SaveDir is the folder path to where the user wants to save the output result of the clipped images. Workspace is a temporary folder that stores the junk files generated from running the program.

Once the user provides all the input parameters, the program will go through the features within the input feature class file one by one to generate the clipped images. First, the program will generate a rectangular minimum bounding box around the feature using its center point and store the values of the coordinates of the corners of the minimum bounding box. Then it will use those points as a parameter to send a request to the TNM API and download the image files according to the type of input feature class that it analyzes. Once the program receives the data from the TNM API, it checks whether the user has the file already in their data folder and downloads and stores it in the data folder if not. Finally, the program will clip the image that was downloaded from the TNM API to intersect a polygon that is 1.5 times the area of the minimum bounding box of the features. That images is saved in the directory that is provided as the SaveDir with the name of the feature plus ‘FeatureClipedDEM’ or ‘FeatureClipHT’ depending on the type of input feature class.