Manual for RasterMiner

RasterMiner Toolkit Overview :

The purpose for development of this toolkit is to discover knowledge hidden in raster images. Raster data is a type of digital data that represents information in a grid or pixel-based format. Raster data is commonly used in geographic information systems(GIS) and remote sensing applications to represent things like elevation, temperature, land use and vegetation cover. Raster data operations refer to the various manipulations and analyses that can be performed on raster data. The following are the operations that you can perform on raster data using raster Miner toolkit.

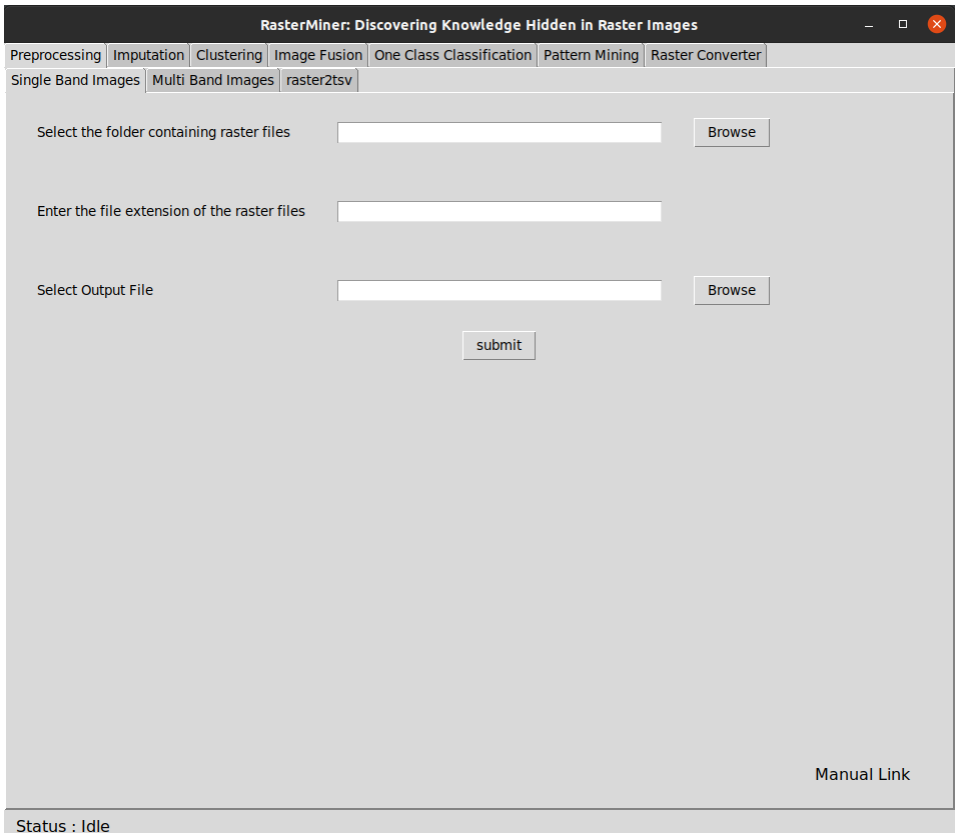
- Data Preprocessing –Data conversion.
- Imputation – Handling corrupted or missing data.
- Clustering – cluster similar character pixels.
- Image Fusion - Predict image with better spectral and spatial resolution.
- One Class Classification – classify specific target class.
- Pattern Mining
- Raster Converter - convert tsv file back into tif format.

File Format : TIFF, JPEG, PNG, Geo TIFF, etc.

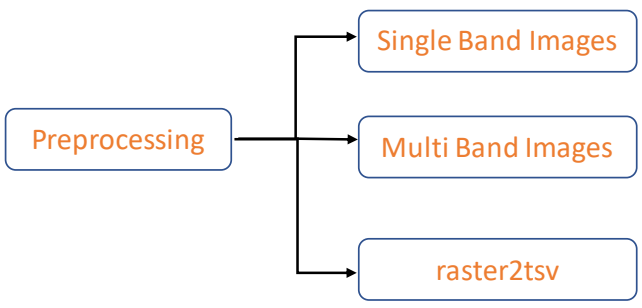
Software language : Python

Working with raster data can be computationally intensive, requiring significant amount of storage space. Overall, raster Miner is a powerful tool for raster data operations, which are a crucial aspect, as they allow users to manipulate and analyze the data in ways that can help to understand and make productive decisions.

Raster Miner toolkit GUI



Preprocessing – Data preprocessing is a crucial step in preparing raster data for analysis. Raster data can be stored in a variety of file formats, which may not be compatible with the software or analysis techniques being used. In these cases, data conversion is necessary to convert the data into a more suitable format. Here any raster data is processed and stored in 'tsv' format.



Converts all single band raster files in the selected folder into single tsv file. (vertical Expansion)

Converts all multi band raster files in the selected folder into single tsv file. (Horizontal Expansion)

Creates separate tsv file for each multi band raster file in the selected folder.

Input : Raster File (Any format)

Output : tsv (tab separated values) File

Preprocessing – Single Band Images

Step 1 : Click on Browse button and select the folder containing single band raster data.

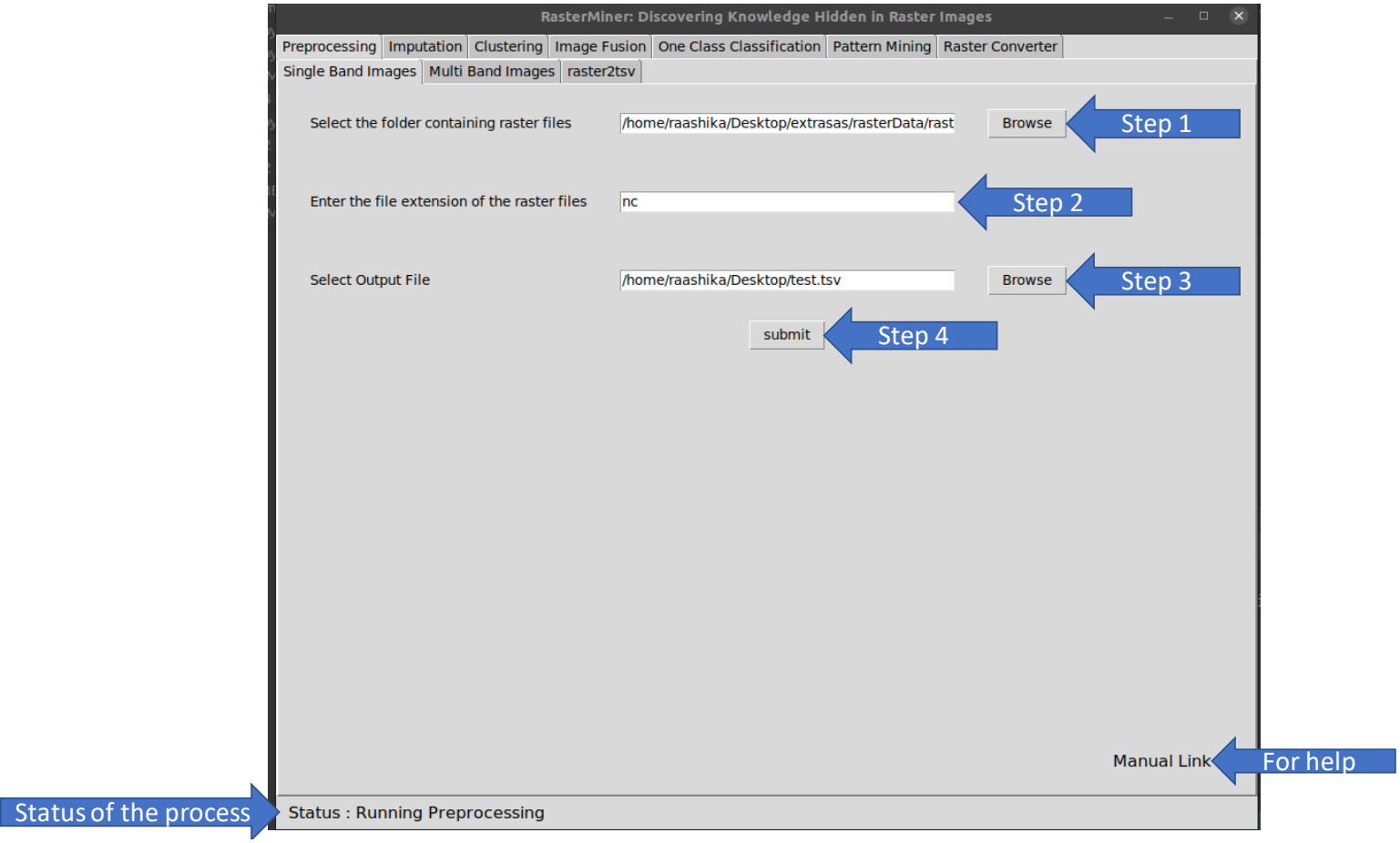
Step 2 : Enter the file extension of the raster files.

For example : "nc"

Step 3 : Click on Browse button and select the output directory and enter the output file name.

Step 4 : Click on the submit button to complete data conversion.

Note : "Status" Bar displays the status of the process – processing / completed. And notification is displayed after successfully completing the process.



Preprocessing – Multi Band Images

Step 1 : Click on Browse button and select the folder containing multi band raster data.

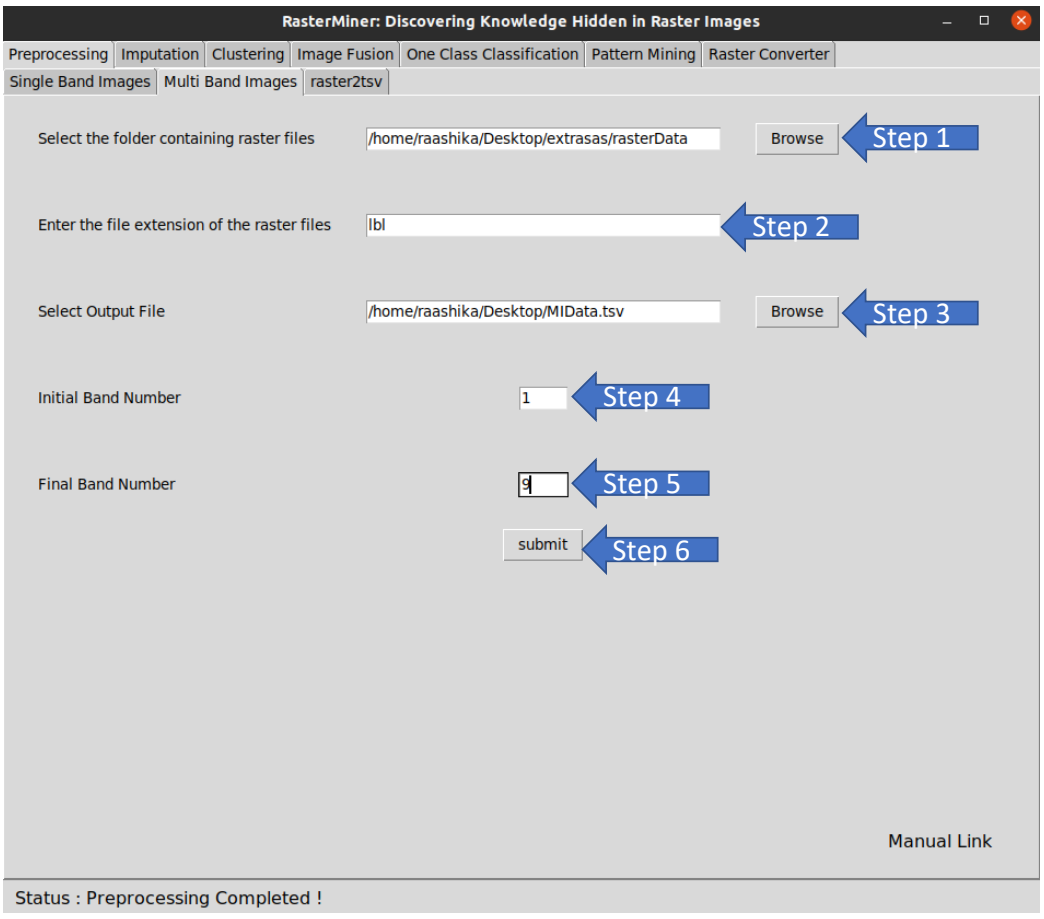
Step 2 : Enter the file extension of the raster files.
Example : "lbl"

Step 3 : Click on Browse button and select the output directory and enter the output file name.

Step 4 : Enter the start band number (Example : If file has 9 bands enter 1).

Step 5: Enter end band number (Example : If raster file has 9 bands enter 9).

Step 6 : Click on the submit button to complete data conversion.



Preprocessing - raster2tsv

Step 1 : Click on Browse button and select the folder containing multi band raster data.

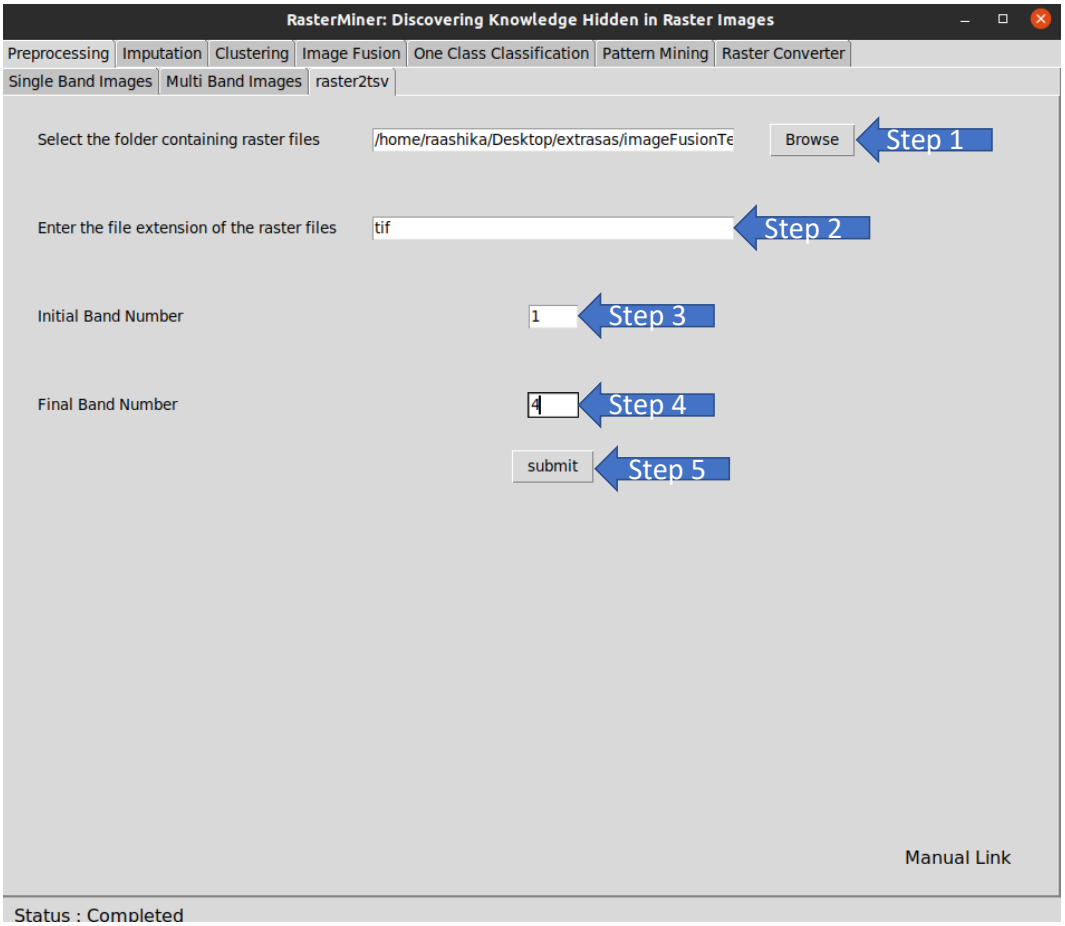
Step 2 : Enter the file extension of the raster files.
Example : "lbl"

Step 3 : Enter the start band number (Example : If file has 9 bands enter 1).

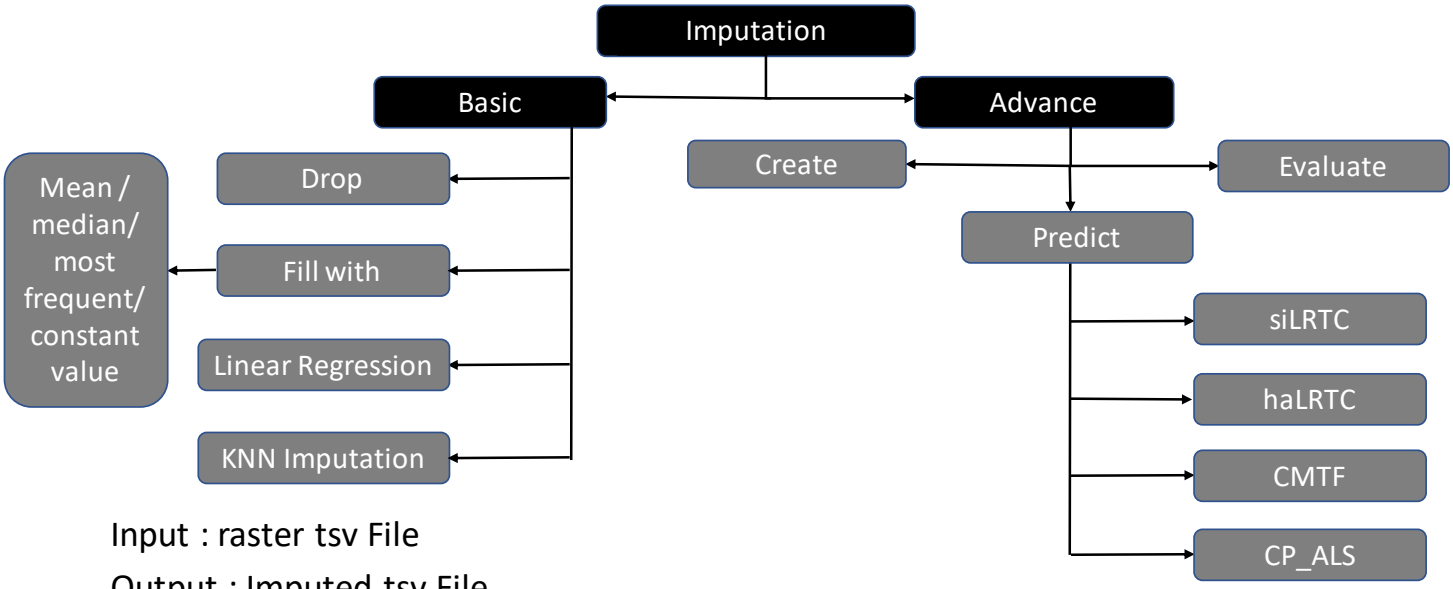
Step 4: Enter end band number (Example : If raster file has 9 bands enter 9).

Step 5 : Click on the submit button to complete data conversion.

NOTE : The output file is stored in the input with input file names.



Imputation – This involves checking the raster data for errors or inconsistencies, such as missing or corrupted values, and addressing any issues that are found by predicting missing values. This step ensures that the data is accurate and suitable for analysis.



Input : raster tsv File

Output : Imputed tsv File

Basic Imputation

Step 1 : Click on Browse button and select the raster file containing missing data.
(Note : only tsv file)

Step 2 : Click on Browse button and select the output directory and enter the output file name.

Step 3 : Select the imputation technique.

Options : Drop – drops the missing values from the data.

Fill type – Categorical Values – If data is categorical select this.

Fill type – Numerical Values – If data is numerical select this. If constant is selected enter the value with which you want to fill.

Linear Regression – Imputation is performed using linear regression technique.

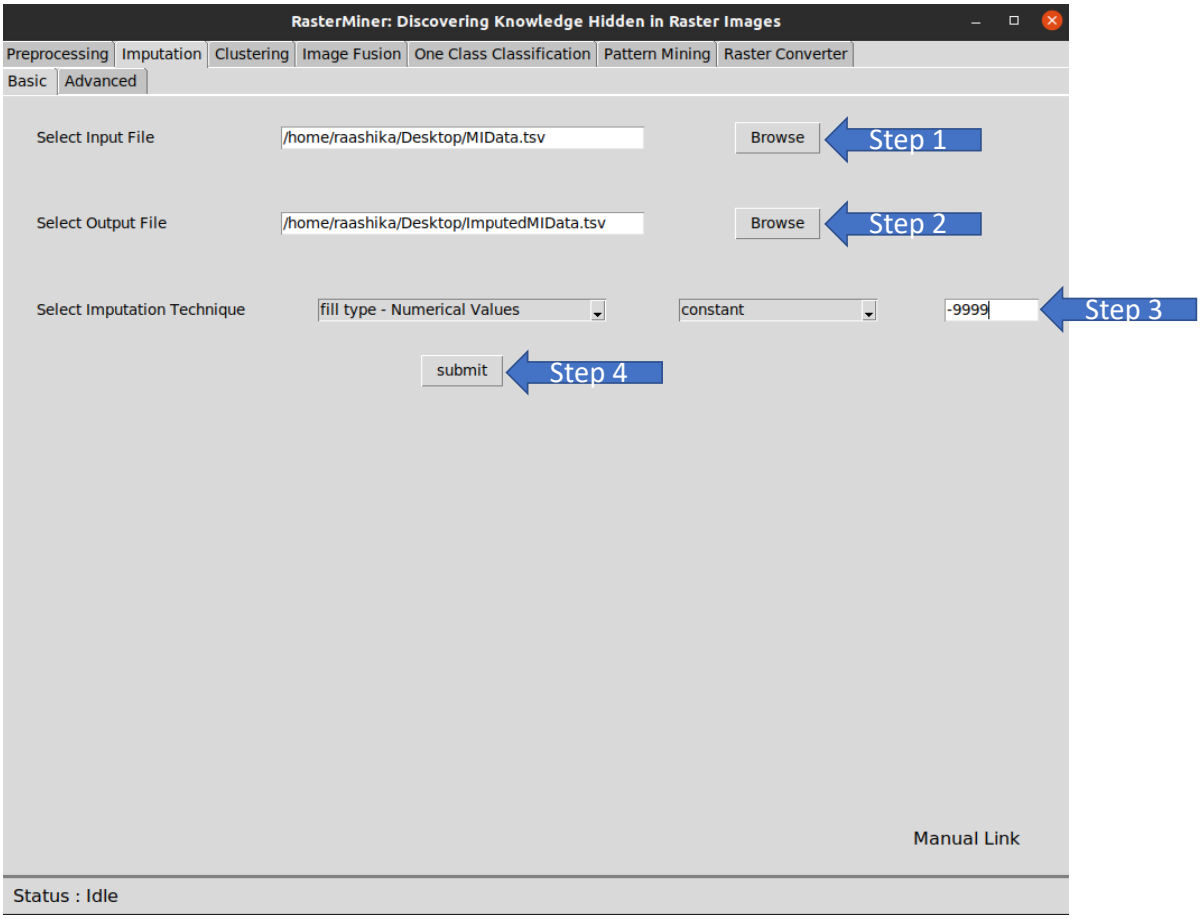
KNN Imputation – K-nearest neighbour technique is used to predict missing values.

For detailed information : https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop.html

Step 4 : Click on the submit button to perform Imputation and check status.

Input – tsv file

Output – tsv file



Advanced Imputation Techniques

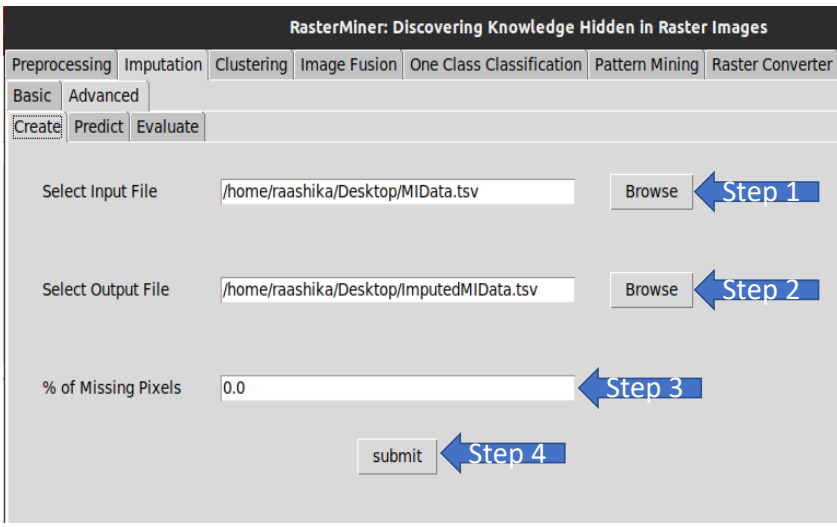
Create - Creates selected percentage of missing pixels in the input file.

Step 1 : Select Input raster file.

Step 2 : Select Output directory and enter output file name.

Step 3 : Enter percentage of missing pixels to be created in the input file.

Step 4 : Click on submit button to create missing pixels.



Input format : tsv File, Output format : tsv File

Predict – predicts missing values or corrupted images.

Step 1 : Select input raster file.
(Either previously created file or any corrupted raster file)

Step 2 : Select output directory and enter output filename.

Step 3 : Select algorithm for prediction.
(Recommended : CMTF/ CP_ALS)

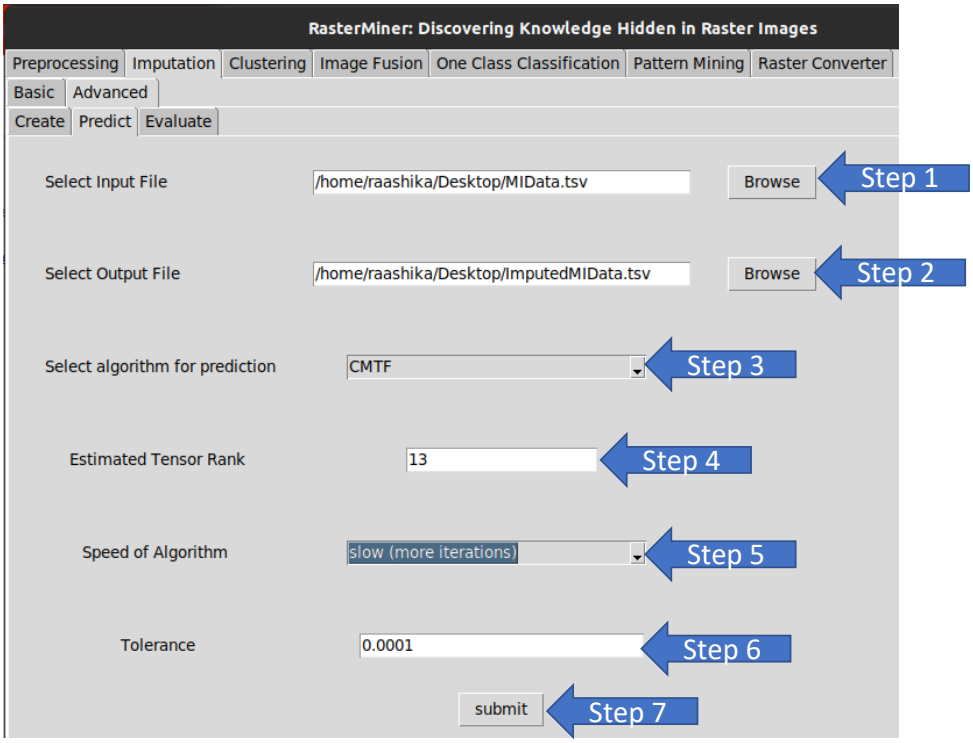
Step 4 : Enter estimated Tensor Rank.(Default : 13)

Step 5 : Select speed of Algorithm.
(slow : For better results, but takes more time)

Step 6 : Enter error tolerance. (Default : 0.0001)

Step 7 : Click on submit button to predict values.

Input format : tsv file, Output format : tsv file.

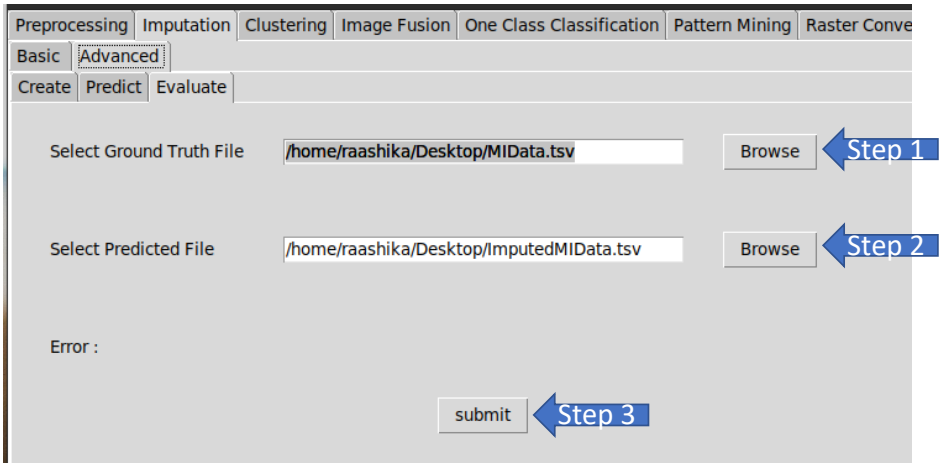


Evaluate – To evaluate the performance of the prediction algorithm. This can only be done if the user has ground truth file without missing values or corrupted data.

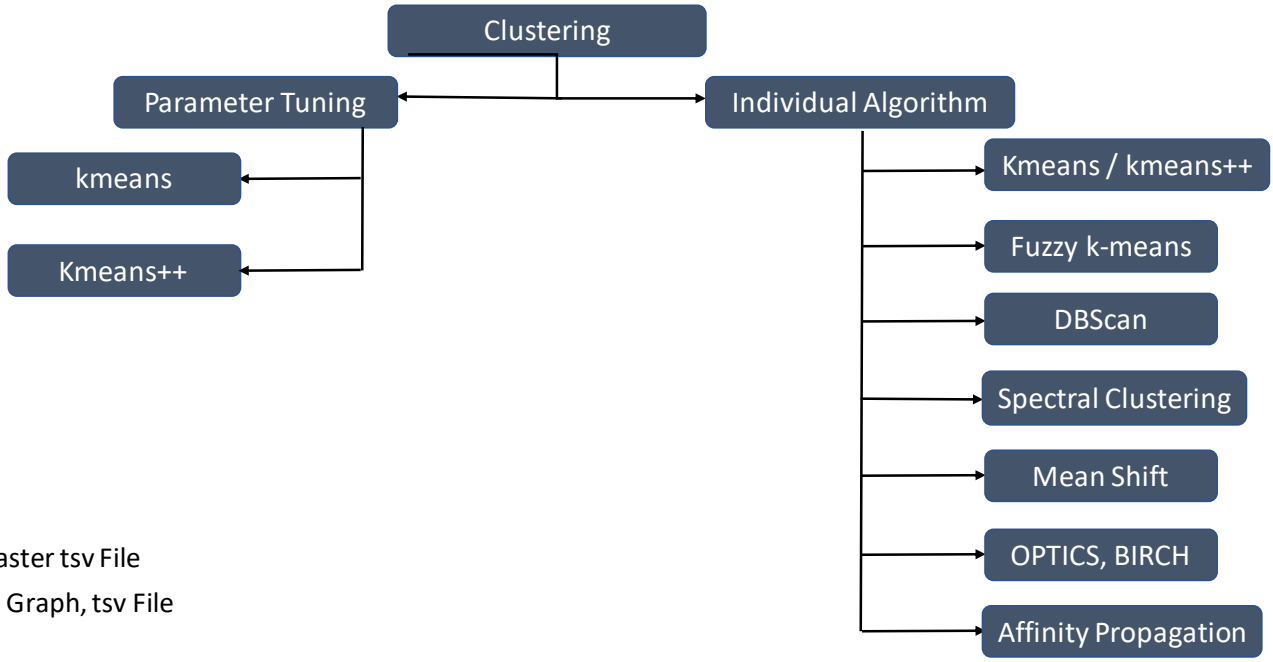
Step 1 : Select ground truth file.

Step 2 : Select predicted file.

Step 3 : Click on submit button to evaluate.
RSE Error is displayed.



Clustering – Clustering is a data analysis technique that can be used on raster data to group similar pixels or regions together based on their values. Clustering algorithms aim to identify clusters of pixels that have similar characteristics, which can be useful for tasks such as land cover classification or object detection by identifying patterns or structures that may not be immediately apparent from visual inspection of data.



Input : raster tsv File

Output : Graph, tsv File

Clustering – Parameter Tuning

Parameter Tuning is an important step in optimizing the performance of algorithms applied to raster data analysis. It involves selecting the best values for the various parameters that control how an algorithm works.

Step 1 : Click on browse button and select the input raster file(Format : tsv).

Step 2 : Click on reference link to learn about the algorithm and understand the parameters for input.

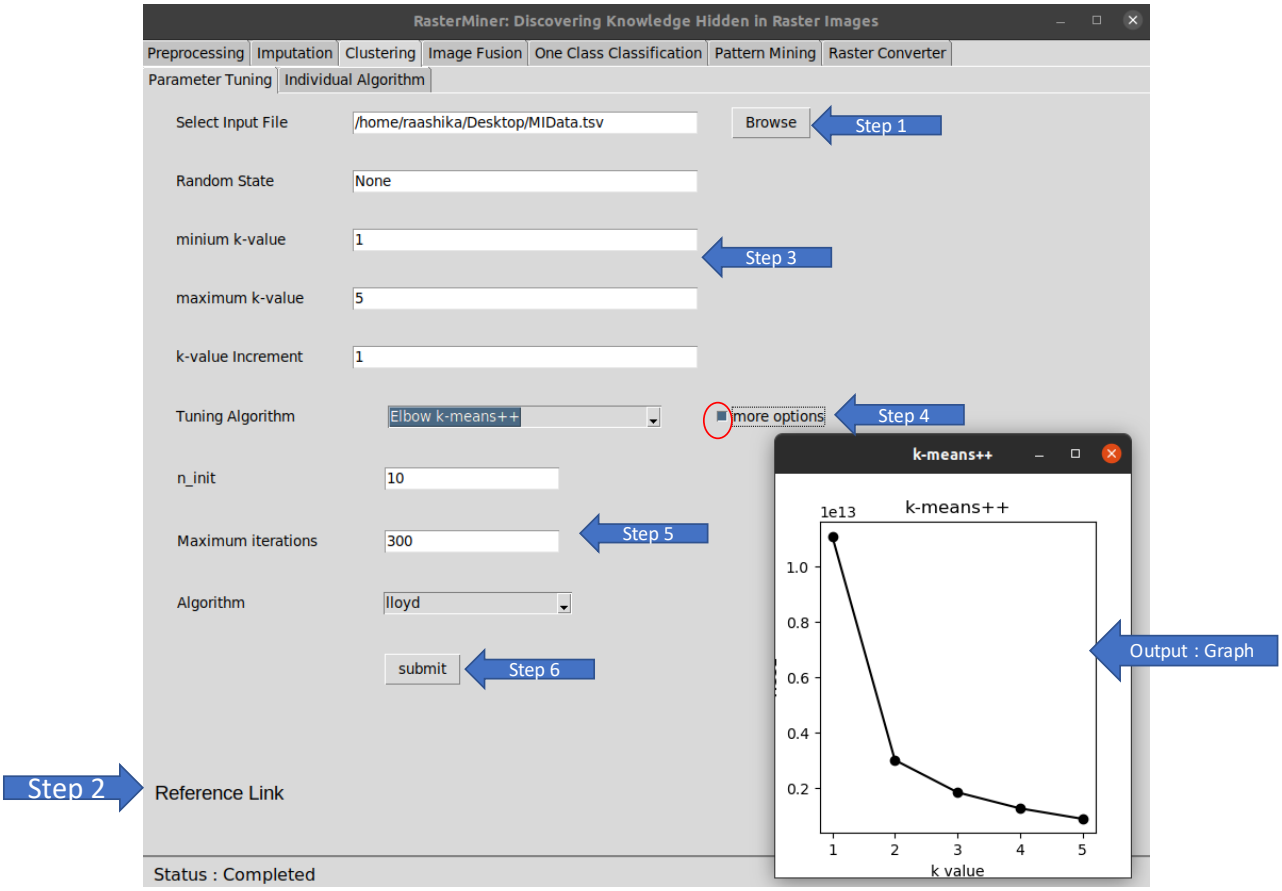
Step 3 : Enter minimum and maximum k-values and increment.

Step 4 : Select tuning algorithm. And click on more options to enter various input parameters.

Step 5 : Enter input parameters, else default values are used.

Step 6 : Click on submit. A graph is displayed after completing the process.

Note : No output file



Clustering – Individual Algorithm

NOTE : The quality of clustering results can depend on the choice of algorithm, the number of clusters chosen, and the quality of the input data.

Prestep : Select algorithm of your choice. The corresponding algorithm GUI appears with input parameters and referencelinks.

Step 1 : Click on reference link to learn about the algorithm and understand the parameters for input.

Step 2 : Click on browse button and select the input raster file(Format : tsv).

Step 3 : Click on browse button and select the output directory.

Step 4 : Enter best k-value(Number of clusters) based on your observation in parameter tuning.

Step 5 : Enter input parameters, else default values are used. And click on more options to enter various input parameters.

Step 6 : Click on submit.

Input Format : tsv file, Output : Two files(cluster centers, cluster data) are saved in the selected output directory.

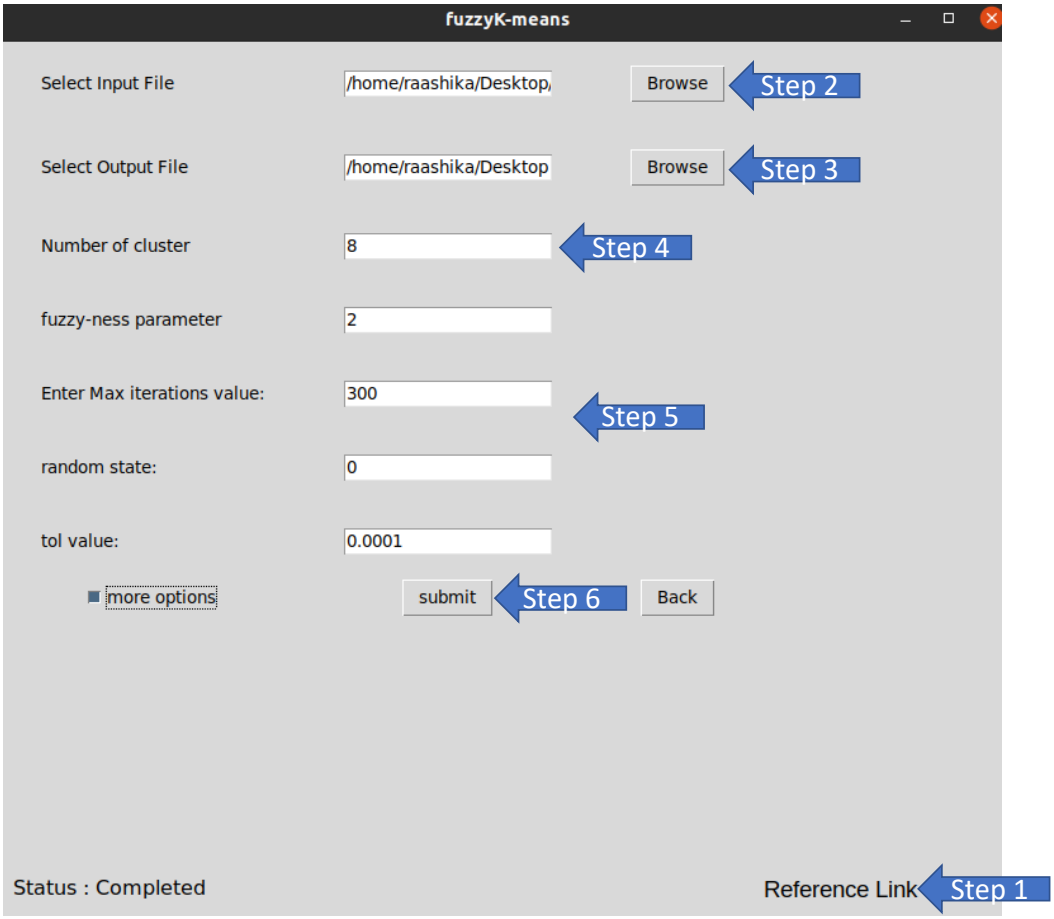


Image Fusion – Image Fusion is a technique used to combine two or more images of the same scene, each captured by different sensors or platforms, into a single image that contains more information than any of the individual images. In raster data analysis, image fusion is commonly used in remote sensing applications, where different types of sensors may be used to capture images of the same area at different wavelengths or resolutions. Image Fusion can provide several benefits in raster data analysis, such as increased spatial or spectral resolution, enhanced image quality, and improved detection or classification accuracy.

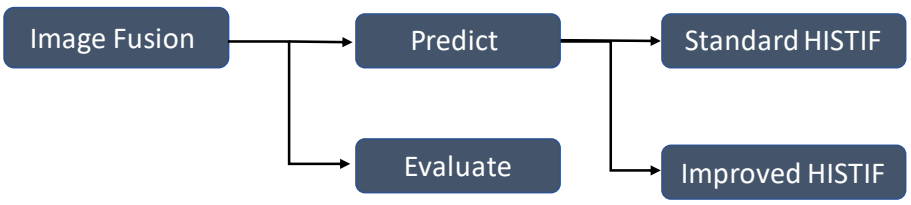


Image Fusion – Predict

NOTE : The quality of Image Fusion results can depend on the quality of the input images, the choice of the fusion method, and specific application being considered.

Step 1 : Select Algorithm for prediction of raster image.

Step 2 : Select input files as specified.

Coarse Image at t0 & t1– low resolution image at t0 & t1 time stamps respectively.

Fine Image at t0 & t1 – High resolution images at t0 & t1 time stamps respectively.

Step 3 : Click on browse button and select the output directory and enter the output file name.

Step 4 : Enter input parameters (necessary : scale factor, iterations, neighbors) and click on "Use Recommended Values" to use default input parameters for others.

Step 5 : Click on submit. (Takes more time to complete the process check status)

Input Format : tsv file, Output : tsv file

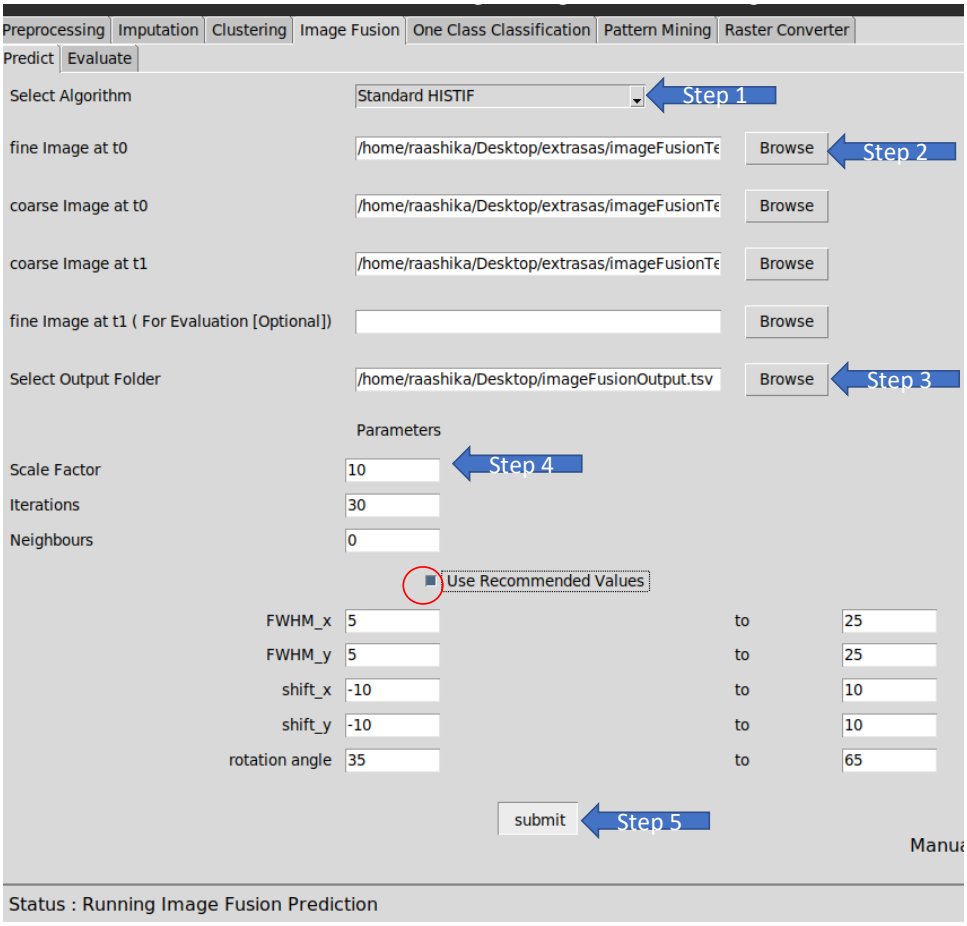


Image Fusion – Evaluate

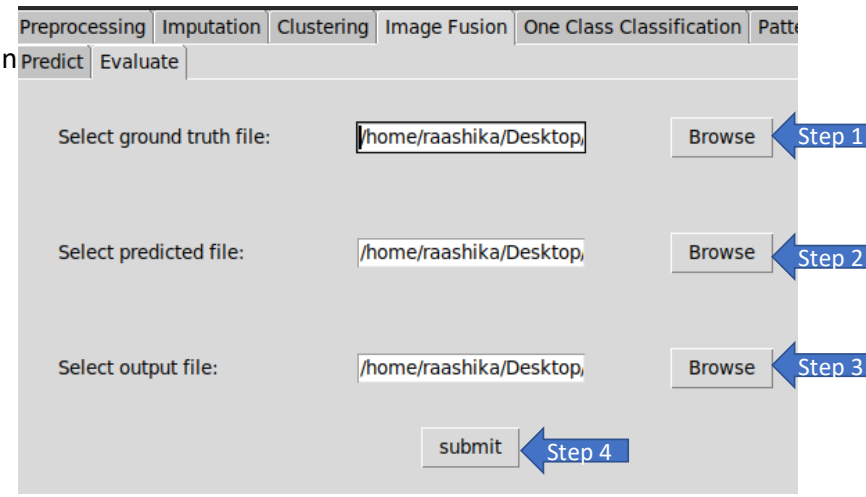
NOTE : This operation is possible only if you have original image (High resolution image at t1 time stamp)

Step 1 : Select ground truth file(high resolution image at t1 timestamp).

Step 2 : Selected predicted file (predicted high resolution image at t1 timestamp).

Step 3 : Click on browse button and select the output directory and enter the output file name.

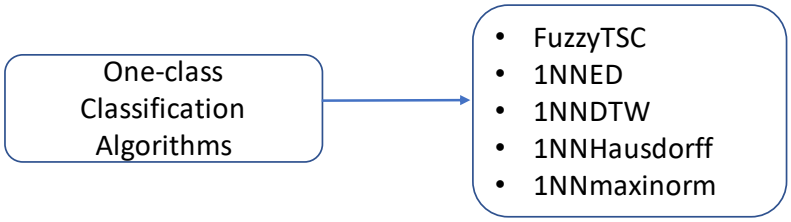
Step 4 : Click on submit. Output file with evaluation metrics for each band is saved in the selected directory.



Evaluation Metrics : RMSE, Mean Absolute Error

Input Format : tsv file, Output : tsv file

- **One Class Classification** – One-class classification is a type of machine learning problem where the goal is to learn a model that can distinguish between a specific target class and all other classes in a raster dataset. This is often used when there is only one class of interest, and the objective is to identify instances of that class in new, unseen raster data.



Step 1 : Select respective algorithm checkbox to run the algorithm. (Can select multiple)

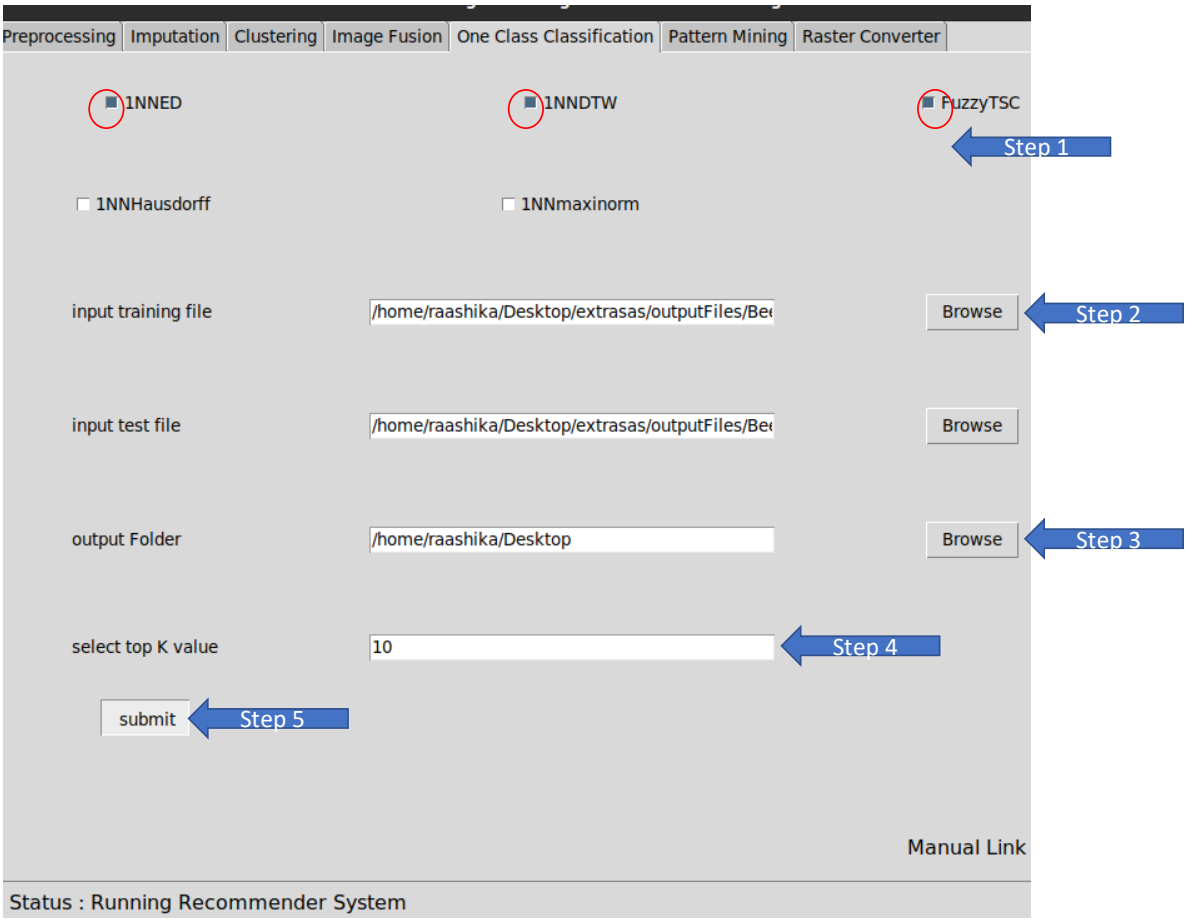
Step 2 : Select input train and test files.

Step 3 : Select output directory.

Step 4 : Select Top-k value.

Step 5 : Click on submit button.

Input Format : tsv file, Output : Two files (Final samples, Top-k Samples) with default names are saved in the selected output directory.



Raster Converter – Raster data is converted into tsv format to make it compatible with software and analysis techniques being used. Raster Converter is used to convert back tsv file into tif format.(contains some limitations)

Step 1 : Select raster tsv file.

Step 2 : Select output directory and enter output file name with "tif" extension.

Step 3 : Click on submit button.