

# 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.

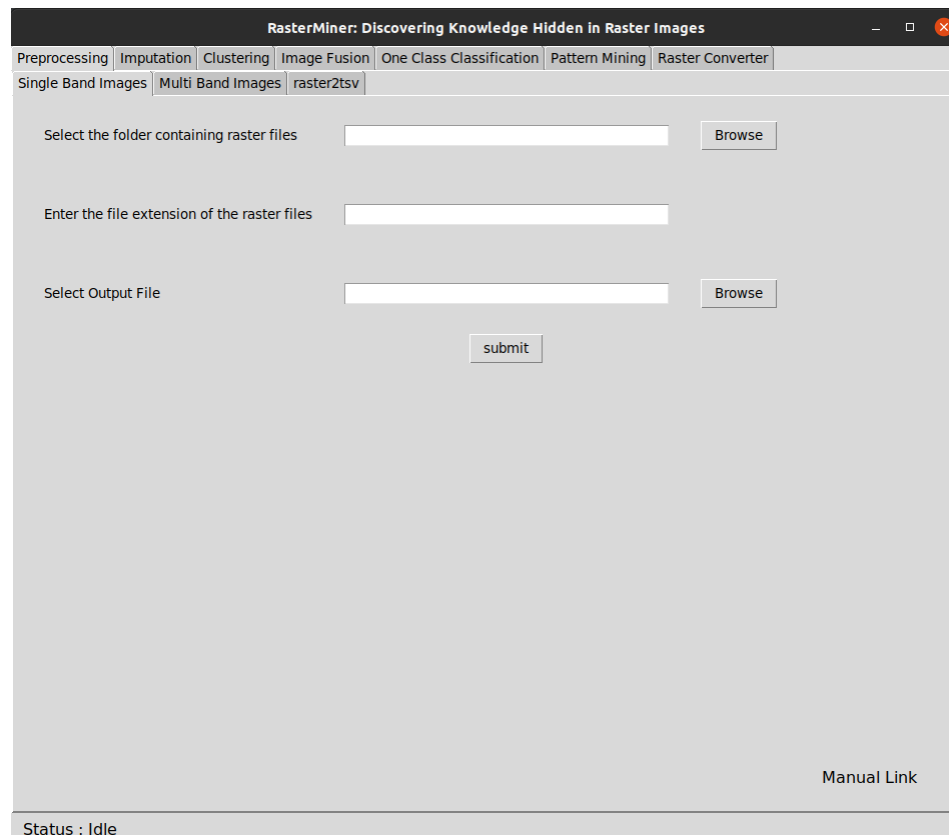
- Data Preprocessing –Data conversion.
- Imputation – Handling corrupted or missing data.
- Clustering – cluster similar character pixels.
- Image Fusion - Predict image with better spectral and spatial resolution.
- One Class Classification – classify specific target class.
- Pattern Mining
- Raster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

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

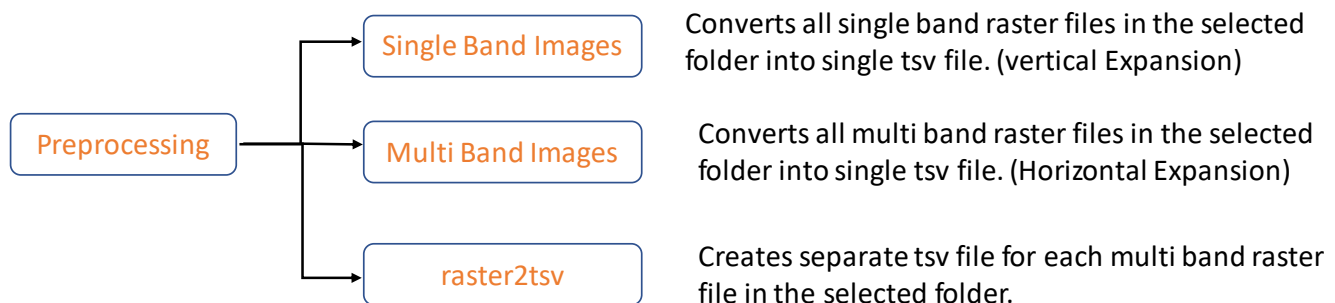
Software language : Python

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.



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.

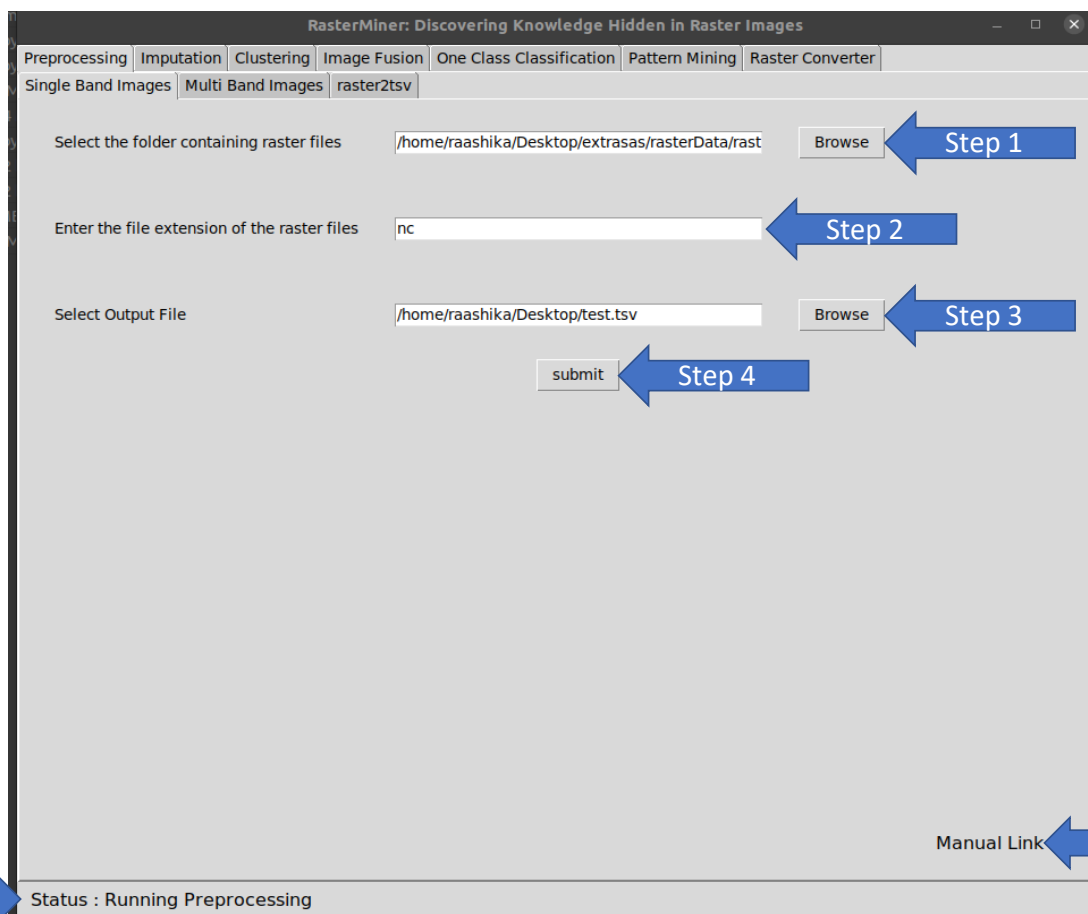
Step 2 : Enter the file extension of the raster files.

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 : ".lbi"

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.

The screenshot displays the RasterMiner web application interface. The title bar reads "RasterMiner: Discovering Knowledge Hidden in Raster Images". The navigation menu includes "Preprocessing", "Imputation", "Clustering", "Image Fusion", "One Class Classification", "Pattern Mining", and "Raster Converter". The "Preprocessing" tab is active, and the "Multi Band Images" sub-tab is selected. The interface contains the following fields and buttons:

- Select the folder containing raster files:** A text input field with the path "/home/raashika/Desktop/extrasas/rasterData" and a "Browse" button. A blue arrow labeled "Step 1" points to the "Browse" button.
- Enter the file extension of the raster files:** A text input field with the value ".lbi". A blue arrow labeled "Step 2" points to the input field.
- Select Output File:** A text input field with the path "/home/raashika/Desktop/MIData.tsv" and a "Browse" button. A blue arrow labeled "Step 3" points to the "Browse" button.
- Initial Band Number:** A text input field with the value "1". A blue arrow labeled "Step 4" points to the input field.
- Final Band Number:** A text input field with the value "9". A blue arrow labeled "Step 5" points to the input field.
- submit:** A button. A blue arrow labeled "Step 6" points to the button.

A "Manual Link" is located at the bottom right of the interface. The status bar at the bottom indicates "Status : Preprocessing Completed !".

## 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 : "lbf"

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.

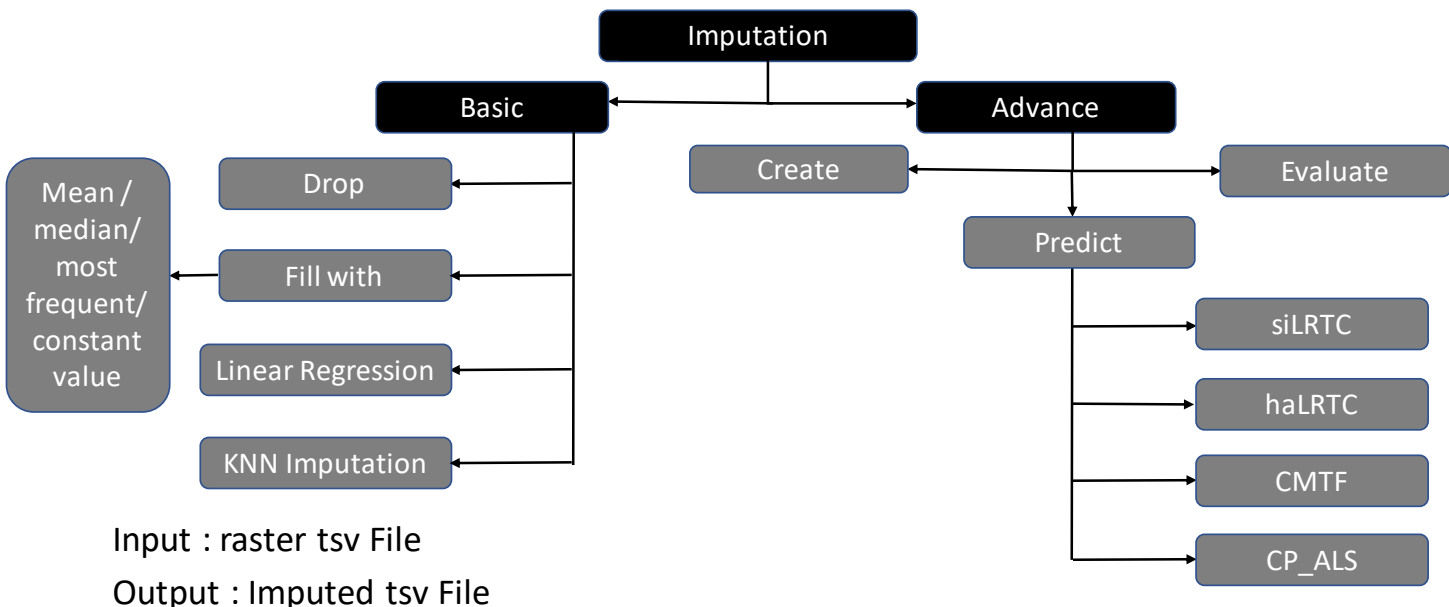
NOTE : The output file is stored in the input with input file names.

The screenshot displays the RasterMiner web application interface. At the top, there is a navigation bar with tabs for Preprocessing, Imputation, Clustering, Image Fusion, One Class Classification, Pattern Mining, and Raster Converter. Below this, there are sub-tabs for Single Band Images, Multi Band Images, and raster2tsv. The main form area contains the following fields and buttons:

- Select the folder containing raster files:** A text input field with the path `/home/raashika/Desktop/extrasas/imageFusionTe` and a **Browse** button. A blue arrow labeled **Step 1** points to the **Browse** button.
- Enter the file extension of the raster files:** A text input field with the value `tif`. A blue arrow labeled **Step 2** points to this field.
- Initial Band Number:** A text input field with the value `1`. A blue arrow labeled **Step 3** points to this field.
- Final Band Number:** A text input field with the value `4`. A blue arrow labeled **Step 4** points to this field.
- submit** button: A button labeled **submit**. A blue arrow labeled **Step 5** points to this button.

At the bottom right of the form area, there is a **Manual Link** text. At the very bottom of the page, a status bar indicates **Status : Completed**.

**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.



### 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

**RasterMiner: Discovering Knowledge Hidden in Raster Images**

Preprocessing | Imputation | Clustering | Image Fusion | One Class Classification | Pattern Mining | Raster Converter

Basic | Advanced

Select Input File:   **Step 1**

Select Output File:   **Step 2**

Select Imputation Technique:    **Step 3**

**Step 4**

Manual Link

Status : Idle

## 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

**RasterMiner: Discovering Knowledge Hidden in Raster Images**

Preprocessing | Imputation | Clustering | Image Fusion | One Class Classification | Pattern Mining | Raster Converter

Basic | Advanced

Create | Predict | Evaluate

Select Input File:   **Step 1**

Select Output File:   **Step 2**

% of Missing Pixels:  **Step 3**

**Step 4**

## 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.

The screenshot shows the 'Predict' tab of the RasterMiner application. The interface includes a navigation bar with tabs for Preprocessing, Imputation, Clustering, Image Fusion, One Class Classification, Pattern Mining, and Raster Converter. Below this are sub-tabs for Basic, Advanced, and Create. The 'Predict' sub-tab is active. The form contains the following fields and controls:

- Select Input File:** A text box with the path '/home/raashika/Desktop/MIDData.tsv' and a 'Browse' button. An arrow labeled 'Step 1' points to the 'Browse' button.
- Select Output File:** A text box with the path '/home/raashika/Desktop/ImputedMIDData.tsv' and a 'Browse' button. An arrow labeled 'Step 2' points to the 'Browse' button.
- Select algorithm for prediction:** A dropdown menu with 'CMTF' selected. An arrow labeled 'Step 3' points to the dropdown.
- Estimated Tensor Rank:** A text box with the value '13'. An arrow labeled 'Step 4' points to the text box.
- Speed of Algorithm:** A dropdown menu with 'slow (more iterations)' selected. An arrow labeled 'Step 5' points to the dropdown.
- Tolerance:** A text box with the value '0.0001'. An arrow labeled 'Step 6' points to the text box.
- submit:** A button at the bottom right. An arrow labeled 'Step 7' points to the button.

**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.

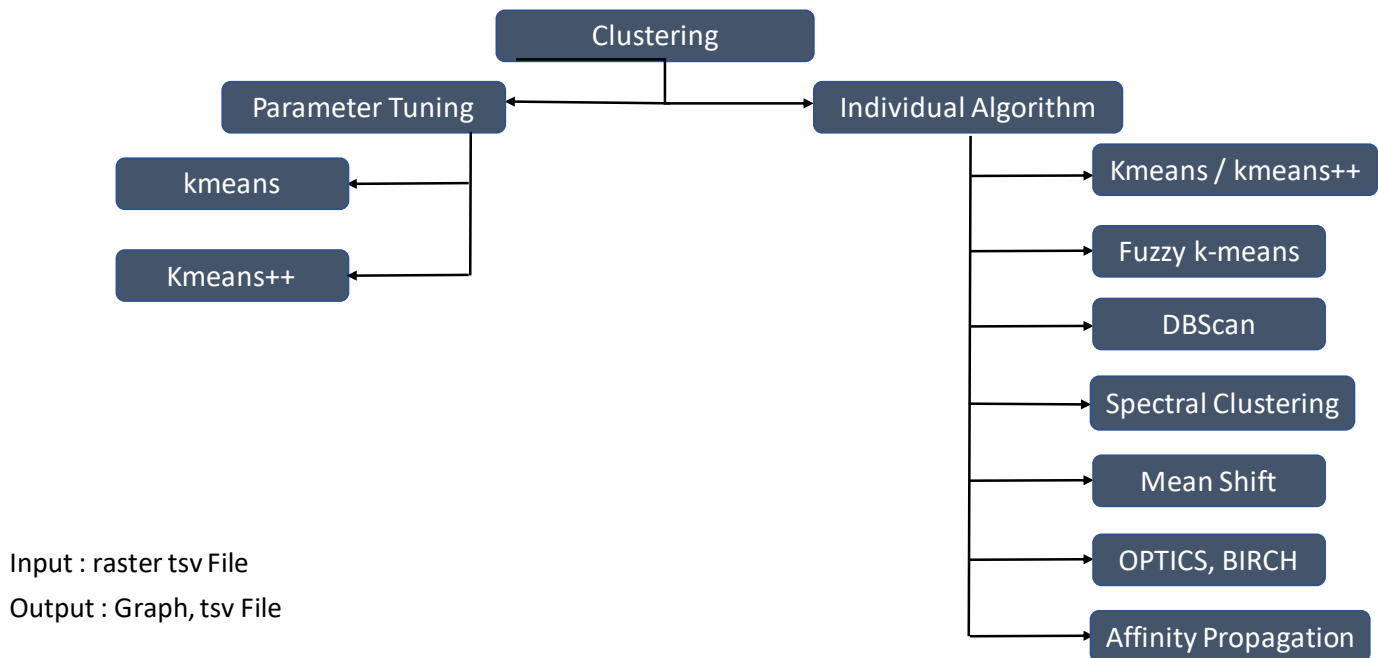
Step 3 : Click on submit button to evaluate.  
RSE Error is displayed.

The screenshot shows the 'Evaluate' tab of the RasterMiner application. The interface includes the same navigation bar as the Predict tab. Below this are sub-tabs for Basic, Advanced, and Create. The 'Evaluate' sub-tab is active. The form contains the following fields and controls:

- Select Ground Truth File:** A text box with the path '/home/raashika/Desktop/MIDData.tsv' and a 'Browse' button. An arrow labeled 'Step 1' points to the 'Browse' button.
- Select Predicted File:** A text box with the path '/home/raashika/Desktop/ImputedMIDData.tsv' and a 'Browse' button. An arrow labeled 'Step 2' points to the 'Browse' button.
- Error :** A text box for displaying the RSE Error.
- submit:** A button at the bottom right. An arrow labeled 'Step 3' points to the button.



**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.



### 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.

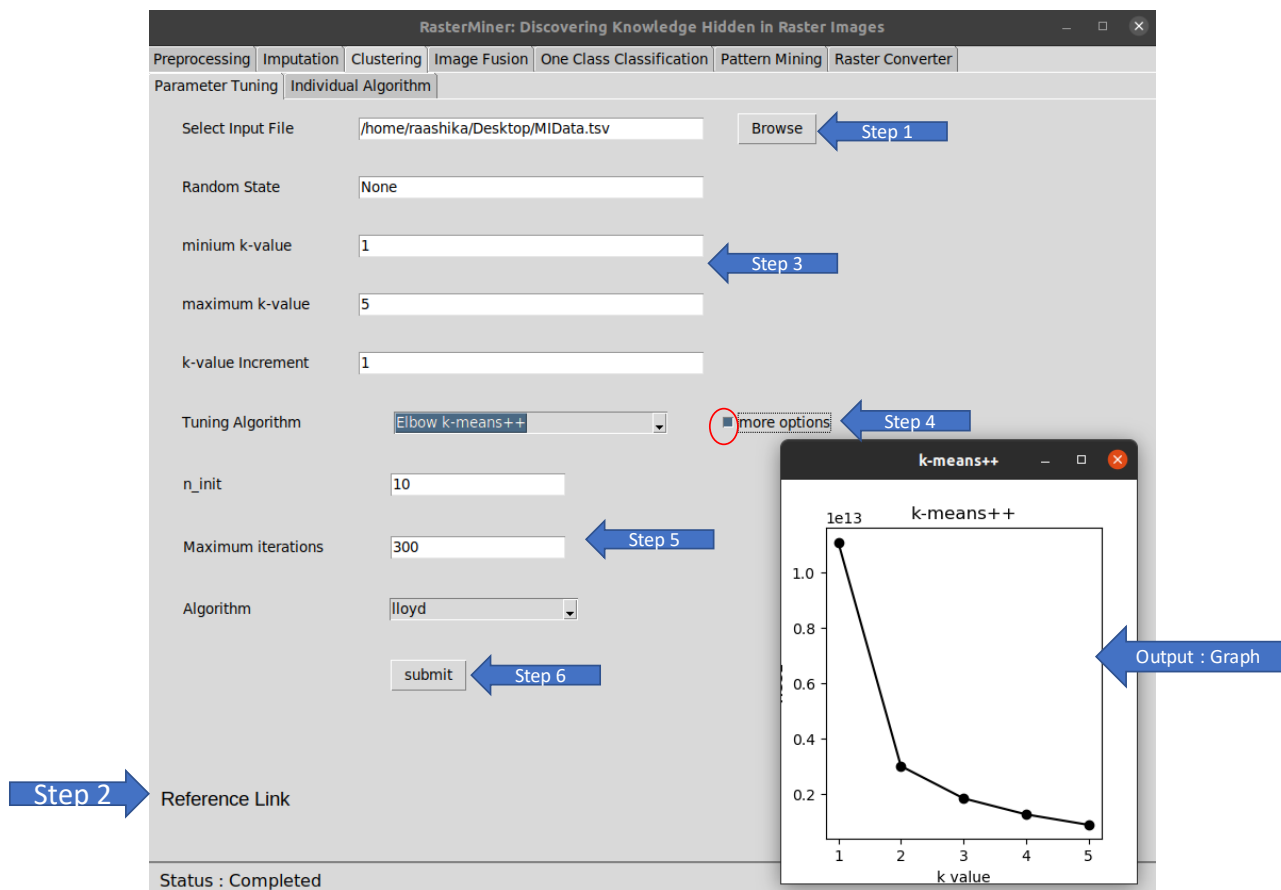
Step 3 : Enter minimum and maximum k-values and increment.

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.

**fuzzyK-means**

Select Input File:   ← Step 2

Select Output File:   ← Step 3

Number of cluster:  ← Step 4

fuzzy-ness parameter:

Enter Max iterations value:  ← Step 5

random state:

tol value:

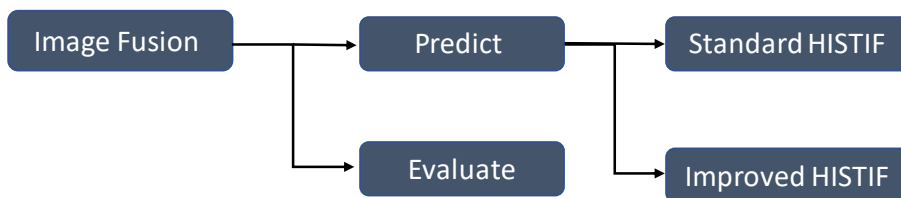
☐ more options

← Step 6

Status : Completed

Reference Link ← Step 1

**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

The screenshot shows a web application interface for image fusion prediction. At the top, there are tabs for Preprocessing, Imputation, Clustering, Image Fusion (selected), One Class Classification, Pattern Mining, and Raster Converter. Below these are sub-tabs for Predict and Evaluate. The main form contains several input fields and buttons, with blue arrows and labels indicating the steps:

- Step 1:** Select Algorithm (Standard HISTIF is selected).
- Step 2:** Select input files (fine Image at t0, coarse Image at t0, coarse Image at t1, and fine Image at t1 (For Evaluation [Optional])) using the Browse button.
- Step 3:** Select Output Folder using the Browse button.
- Step 4:** Enter input parameters (Scale Factor, Iterations, Neighbours) and click on "Use Recommended Values".
- Step 5:** Click on submit.

The parameters section includes:

- Scale Factor: 10
- Iterations: 30
- Neighbours: 0
- FWHM\_x: 5 to 25
- FWHM\_y: 5 to 25
- shift\_x: -10 to 10
- shift\_y: -10 to 10
- rotation angle: 35 to 65

At the bottom, there is a status bar that says "Status : Running Image Fusion Prediction".

## 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

Preprocessing Imputation Clustering Image Fusion One Class Classification Pattern Classification

Predict Evaluate

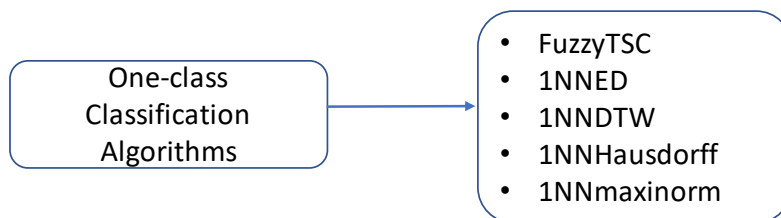
Select ground truth file: /home/raashika/Desktop/ Browse Step 1

Select predicted file: /home/raashika/Desktop/ Browse Step 2

Select output file: /home/raashika/Desktop/ Browse Step 3

submit Step 4

- **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.

The screenshot shows the 'Raster Converter' tab selected in the top navigation bar. The interface includes several radio buttons for selection: 1NNED, 1NNDTW, FuzzyTSC, 1NNHausdorff, and 1NNmaxinorm. Below these are input fields for 'input training file', 'input test file', and 'output Folder', each with a 'Browse' button. A 'select top K value' field contains the number '10'. A 'submit' button is at the bottom left. Blue arrows with labels 'Step 1' through 'Step 5' point to the FuzzyTSC radio button, the first 'Browse' button, the second 'Browse' button, the 'select top K value' field, and the 'submit' button respectively. A 'Manual Link' is at the bottom right, and the status bar at the very bottom reads 'Status : Running Recommender System'.

**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.

This screenshot provides a closer look at the 'Raster Converter' interface. It shows the 'Select Input tsv File' and 'Select Output Folder' fields, each followed by a 'Browse' button. A 'submit' button is located at the bottom. Blue arrows with labels 'Step 1', 'Step 2', and 'Step 3' point to the first 'Browse' button, the second 'Browse' button, and the 'submit' button respectively. The top navigation bar and the status bar are also visible.