**OVERVIEW**

The mwBTfreddy dataset is a dataset that was developed by [Kuyesera AI Lab](https://kailab.tech/) to aid in evaluating flash flood damage in urban areas of Malawi, with a particular focus on the aftermath of Cyclone Freddy in 2023. This dataset includes paired satellite images taken before and after the disaster and accompanying these images are JSON files containing annotated building data, detailing geographic coordinates and categorizing damage levels as no damage, minor, major, or destroyed.

This repository comprises supporting files (KML files and scripts) used in the dataset creation process. The dataset is distributed as a compressed .zip file containing the images and corresponding JSON files, and can be accessed from [Zenodo](https://zenodo.org/records/14190390). The full datasheet is available on [arXiv](https://arxiv.org/abs/2505.01242).

**SCRIPTS**

The table below provides an overview of the scripts used in the data creation process, detailing the description, inputs, and outputs of each script.

**Table 1. Dataset Preparation Scripts**

| **No** | **Name** | **Description** | **Input** | **Output** |
| --- | --- | --- | --- | --- |
| 1 | generating\_gridlines.py | Generates grid lines overlapping each focus area (Chirimba, Chilobwe, and Ndirande) to systematically divide the imagery into contiguous, non-overlapping tiles. | Coordinates for each focus area (see Table II below). | KML files for each focus area. |
| 2 | cropping\_images.py | Crops and resizes satellite images using polygon coordinates from georeferenced .points files. The script removes irrelevant content while preserving geolocation metadata. | Georeferenced .points files. | Cropped and resized .tiff images. |
| 3 | renaming\_tiff\_images.py | Renames cropped images using the xBD naming convention and creates a record of the renaming process. | Cropped .tiff images. | Renamed .tiff images. |
| 4 | converting\_geocoordinates\_to\_csv.py | Extracts building coordinates from labeled polygons and generates a CSV file for each image. | QGIS project files containing building labels. | CSV files with labeled building coordinates. |
| 5 | csv\_to\_json.py | Converts CSV files with polygon data (map and pixel coordinates) into JSON format. | CSV files containing polygon data. | JSON files formatted for use with the dataset. |

**Table II. Image Counts and Coordinates of the Selected Areas**

| AREA | COORDINATES | NO. IMAGES |
| --- | --- | --- |
| CHIRIMBA | 35.06135646185926,-15.7455674468705  35.0630140330653,-15.74401809140623  35.0606258243407,-15.74217357016835  35.05878123267473,-15.74352129840774 | 6 |
| NDIRANDE | 35.04990122048631,-15.76751173684101  35.04753585118715,-15.76823081389567  35.05147913477497,-15.7744571861299  35.05471668399898,-15.77301319505322 | 20 |
| CHILOBWE | 35.00004428168152, -15.82503923829606  35.04009739154355, -15.82507457317074  35.04012666318832, -15.85010734237624  35.00015291121928, -15.85003454786558 | 1000 |
| TOTAL |  | 1026 |

**NOTES**

* Users are encouraged to read the [datasheet](https://arxiv.org/abs/2505.01242) for complete documentation on data creation, and usage guidelines.
* Install all required dependencies from the requirements file before running the scripts.