

Dataset Documentation

Dataset Name: PlantVillage Radiant Earth Foundation Kenya Crop Type Training Dataset

Location and boundaries

Overall Location Method

- Ground collection only
- Ground collection with boundary drawn using imagery
- Ground collection with spatial buffer added
- Boundary drawn from imagery
- Other _____
- Unknown

GeoLocation Device

- Industrial grade GPS (List model) _____
- Retail grade GPS
- Mobile Phone GPS
- N/A
- Unknown

Ground Boundary Method (Details explained in Appendix A)

- Live/Continuous point capture of walk-around
- Manual point capture of walk-around
- Manual point capture of polygon boundaries (not whole field)
- Manual point capture for later image annotation
- Manual point capture for spatial buffer within field
- Manual point capture while looking at but not in field, with heading recorded
- Other _____
- Unknown

Imagery used (Skip if no imagery used)

Sensor: Unknown

Date(s):

List scenes used in Appendix B

Imagery Annotation methods

- Boundaries drawn based on a single ground point captured
- Boundaries drawn/edited based on multiple ground points captured
- Buffer validated from ground point captured
- Boundary drawn without ground reference data (Include description of methods in Appendix C)
- Pixels annotated without ground reference data (Include description of methods in Appendix C)
- Unknown

Boundary inclusion

- Captured polygon includes the entire field/area
- Captured polygon includes only a sample of the field/area

Classification

Classification Type

- Land cover
- Crop type
- Other _____

Classes/fields used

Describe in Appendix D

Ground Referenced Classification

- Observation (Describe methods of determination in Appendix E)
- Survey/interview with land holder (Describe methods in Appendix E)
- Other (Describe methods in Appendix E)

Image Referenced Classification

Describe methods used in Appendix C

Data Properties

Property name	Property Description	Parameters/Allowed responses (optional)
Field_id	A unique integer number to identify pixels corresponding to one field/farm	
Label	Crop Type	1, 2, 3, 4, 5, 6, 7

Appendix A: Describe the method of geographic ground data collection

PlantVillage app is used to collect multiple points around the field and collectors have access to basemap imagery in the app during data collection. They use the basemap as a guide in collecting and verifying the points.

Post ground data collection, Radiant Earth Foundation conducted a quality control of the polygons using Sentinel-2 imagery of the growing season as well as Google basemap imagery. Three actions were taken on the data 1) several polygons that had overlapping areas with different crop labels were removed, 2) invalid polygons where multiple points were collected in corner of the field (within a distance of less than 0.5m) and the overall shape was not convex, were corrected, and 3) fields that were too small to have a corresponding Sentinel-2 pixels were removed.

Appendix B: List imagery scenes used for annotation (ideally also included in metadata)

Appendix C: Describe how boundaries and classes were determined without ground reference data

Appendix D: List all top-level classes or the classification guidance used

Crops were identified using FAO list, and include:

1) Maize, 2) Cassava, 3) Common Bean, 4) Maize & Common Bean (intercropping), 5) Maize & Cassava (intercropping), 6) Maize & Soybean (intercropping) and 7) Cassava & Common Bean (intercropping)

Appendix E: Describe methods for determining classes based on direct/ground observation

Farmers were asked when available, otherwise observations of the field collector.

Appendix F: Detailed description of the dataset properties
