applied sciences

MDPI

Article

Mapping Cassava Production in Uganda

Renata Retkute * and Christopher A. Gilligan

Epidemiology and Modelling Group, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK; cag1@cam.ac.uk

* Correspondence: rr614@cam.ac.uk

Abstract

Cassava is a critical staple crop for food security and rural livelihoods in Sub-Saharan Africa, yet high-resolution maps of its distribution remain scarce, particularly for smallholder systems. In this study, we generated a 10 m resolution cassava presence map for Uganda (CM24) by fine-tuning a Random Forest classifier on TESSERA foundation model embeddings derived from Sentinel-1 and Sentinel-2 time series. Using field survey data from the Copernicus4GEOGLAM campaign for training and validation, the model achieved excellent discriminative ability (validation AUC = 0.9532, test AUC = 0.9524). Visual validation against high-resolution satellite imagery confirmed good spatial agreement, capturing both large contiguous fields and small fragmented plots. Comparison with two existing global products (CassavaMap and SPAM2020) and two seasons of national survey data conducted by the Uganda Bureau of Statistics showed that CM24 produced national harvested area estimates that fell between the two survey totals, whereas CassavaMap and SPAM2020 systematically overestimated harvested area by factors of two to three. Our results demonstrate that foundation-model embeddings offer a robust and scalable approach for mapping cassava in heterogeneous smallholder landscapes. The resulting CM24 map provides a spatially explicit tool to support disease surveillance, agricultural monitoring, and food security planning in Uganda and beyond.

Keywords: cassava mapping; smallholder agriculture; TESSERA; foundation model; Uganda

1. Introduction

Cassava (*Manihot esculenta* Crantz) is a cornerstone of food security and rural livelihoods in Sub-Saharan Africa, where over half of the world's crop is produced and an estimated half a billion people consume cassava daily as the second most important source of dietary carbohydrates after maize [1]. Its exceptional tolerance to drought and ability to grow on marginal soils make cassava a climate-resilient staple that ensures baseline food availability even under erratic rainfall and rising temperatures [2]. Beyond subsistence, cassava generates socio-economic benefits for millions of smallholder farmers by diversifying income, creating employment, and reducing vulnerability to economic and environmental shocks [3]. Increasingly recognised as a strategic asset for climate adaptation and industrial development, cassava products—starch, ethanol, and animal feed—play an expanding role in economic growth and poverty reduction across the continent [4].

Despite this critical importance, accurate spatial information on the distribution of the crop throughout cassava growing regions, cultivation practices, and health status remains strikingly sparse. The crop is highly vulnerable to cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), which are spread rapidly through infected cuttings

Check for updates

Academic Editor: Rui Sun

Received: 28 May 2026
Revised: 17 June 2026
Accepted: 18 June 2026
Published: 25 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

and by whitefly vectors. High-resolution maps are needed to support early detection, targeted surveillance, and effective intervention [5]. Cassava's extended growing cycle (6–24 months) and cultivation in marginal, remote areas make ground-based monitoring logistically challenging and expensive, whereas remote sensing offers cost-effective alternatives for estimating planted area, yield, and production trends. The growing role of cassava as an industrial feedstock demands reliable supply forecasts that depend on spatially explicit cropping calendars and yield variability assessments [6]. Climate change is projected to alter suitable growing zones across Africa, so spatially explicit models of current distribution serve as baselines for adaptation planning, varietal selection, and targeting of extension services [7]. Thus, mapping cassava using satellite imagery and field surveys is essential for improving agricultural management, disease control, and food security planning across Sub-Saharan Africa.

Recent advances in remote sensing and machine learning have opened new pathways for scalable crop mapping, yet cassava mapping remains challenging due to intercropping practices, small fragmented fields, and physiological overlap with analogous crops [8]. Current approaches to produce cassava production maps rely on profiles of vegetation indices derived from remote sensing time series [9,10]. Remote sensing-derived phenological metrics have also been used to predict cassava mosaic disease outbreaks for areas where the presence of cassava is known in situ [11]. Nevertheless, these phenology-based strategies remain vulnerable to persistent cloud cover during critical growth phases, mixed-pixel effects in fragmented smallholder landscapes, spectral confusion between cassava and similar crops (e.g., sugarcane [10]), and an inability to capture intercropping patterns—limitations that constrain both accuracy and scalability in tropical agricultural regions.

A particularly promising development to overcome these limitations is the emergence of geospatial foundation models (GFMs) for Earth observation [12]. These pre-trained, task-agnostic representations, fine-tuned for downstream tasks with minimal labeled data, have demonstrated strong performance in applications such as crop classification [13], crop yield and tillage monitoring [14], flood mapping and burn scar detection [15]. One such model is TESSERA (Temporal Embeddings of Surface Spectra for Earth Representation and Analysis), which generates dense, precomputed representations from multi-modal satellite time series [16]. These representations condense a full year of Sentinel-1 SAR and Sentinel-2 optical observations into compact 128-dimensional latent embeddings at 10 m spatial resolution, offering a rich feature space that can potentially capture subtle phenological and structural differences between cassava and other crops without requiring explicit handcrafted indices or region-specific calibration. Such a foundation model is particularly well-suited for mapping smallholder cassava systems, where fields are small, fragmented, and often intercropped.

Cassava production in Sub-Saharan Africa is predominantly carried out by smallholder farmers, who lack timely, high-resolution agricultural information. In Uganda, cassava serves as both a staple food and a strategic reserve crop for food security. According to the 2024 National Population and Housing Census, about 1.8 million agricultural households grow cassava nationwide [17]. Despite this importance, spatially explicit data on cassava distribution at field scales remain limited. In this study, we generate high-resolution (10 m) cassava distribution maps for Uganda (Cassava Map CM24) by fine-tuning a random forest classifier on TESSERA embeddings trained on field survey data. We compare our mapped cassava presence (CM24) with Annual Agricultural Survey Data for 2019 and two existing products (CassavaMap and SPAM2020). By providing a fine-scale, spatially explicit tool, our work supports agricultural monitoring, targeted interventions, and food security planning for smallholder cassava producers in Uganda and beyond.

2. Materials and Methods

2.1. Study Area

The study covers the entire territory of Uganda, which lies approximately between longitude 29.5° E and 35.0° E, and latitude 1.5° S to 4.5° N. Terrain varies from the Lake Victoria Basin in the south (elevations 1000 m) to the Rwenzori Mountains (>5000 m) along the western border, with the Central Plateau averaging 1000–1500 m [18] (Figure 1A). Annual mean temperatures range from 5 °C in the highlands to 30 °C in the lowlands (Figure 1C), while annual precipitation varies from 2200 mm around Lake Victoria to 500 mm in the semi-arid northeast (Figure 1D). The country has distinct rainfall patterns, with the south experiencing a bimodal rainfall pattern, with the primary rainy seasons occurring from March to May (MAM) and September to November (SON) [19]. In contrast, the drier northern region follows a unimodal rainfall pattern, characterised by a single, prolonged rainy season from April to October, followed by an extended dry season between November and March [20]. Agro-ecological zones were derived from the HarvestChoice AEZ16 classification [21]. Six classes are present in Uganda: Subtropic − warm/humid (57.97%), Subtropic − cool/arid (0.03%), Subtropic − cool/semiarid (0.03%), Subtropic − cool/humid (14.50%), Tropic − warm/arid (0.03%), and Tropic − warm/semiarid (0.03%) (Figure 1B).

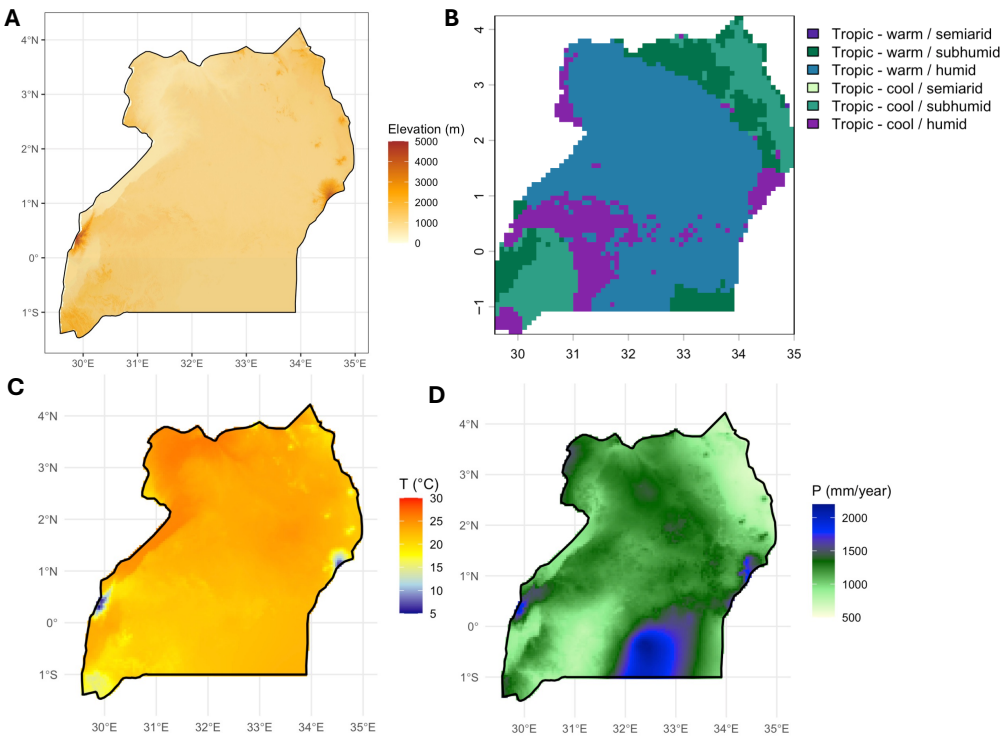


Figure 1. Physical and agricultural context of Uganda. (**A**) Elevation above sea level (m) derived from the Shuttle Radar Topography Mission (SRTM) digital elevation model [18]. (**B**) Agro-ecological zones based on the HarvestChoice AEZ16 classification [21]. (**C**) Annual mean temperature (°C) from WorldClim v2.1. (**D**) Annual precipitation (mm) from WorldClim v2.1.

2.2. Ground Reference Data

We used the publicly available land cover and crop type data from the European Commission's Joint Research Centre for dataset for Uganda [22]. We call this data set Cop4Geoglam. The data were collected during a field campaign in northern and north-eastern Uganda during the 2022 short rains season. Sampling was based on 500 × 500 m patches (338 in total); within each patch, fields were delineated, and for each

sampled field, geolocation, visit time, and field characteristics were recorded, with cropland fields further assessed for crop type.

2.3. Satellite Imagery

This study utilized TESSERA (Temporal Embeddings of Surface Spectra for Earth Representation and Analysis), which uses self-supervised learning to produce 128-dimensional latent embeddings at 10 m spatial resolution, distilling a full year of Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 optical observations into compact feature vectors for each 10 m pixel globally [16]. We obtained pre-computed TESSERA embeddings for Uganda for the 2024 calendar year using the GeoTessera Python library (version 1.0) [GeoTessera, 2025; available at: https://github.com/ucam-eo/tessera (accessed on 1 May 2024)]. The global embedding archive is arranged into 5° × 5° tiles (approximately 550 km × 550 km at the equator) [GeoTessera, 2025], each stored as a cloud-optimised GeoTIFF containing all 128 embedding channels at 10 m pixel resolution. A total of 3780 tiles covering Uganda were downloaded from Amazon Web Services (AWS) public storage [GeoTessera, 2025] using a bounding box of 29.5°CE to 35.0°CE and 1.5°CS to 4.5°CN. For each 10 m pixel, the 128-dimensional feature vector was extracted and used directly as input to a pixel-wise supervised classifier. The trained classifier was then applied to the same embedding tiles to produce binary maps for cassava presence/absence at 10 m resolution for the entire study area.

2.4. Model Training and Evaluation

Using built-in R caret functions [23], the Random Forest (RF) model was trained via repeated 5-fold cross-validation with three repeats. To address class imbalance, up-sampling of the minority class was applied, and class probabilities were requested for subsequent threshold optimization—all handled natively by caret. The model was configured with 100 trees; all other hyperparameters retained their default values as implemented in the caret package.

Model performance was assessed using accuracy, precision, recall, and F1 score [24]. These metrics were derived from the confusion matrix as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where,

- True Positive, TP, is the number of samples labelled as positive by the model that are actually positive;
- False Positive, FP, is the number of samples labelled as positive by the model that are actually negative;
- True Negative, TN, is the number of samples labelled as negative by the model that are actually negative;
- False Negative, FN, is the number of samples labelled as negative by the model that are actually positive.

2.5. Annual Agricultural Survey Data

The Uganda Bureau of Statistics (UBOS) established the Annual Agricultural Survey (AAS) in 2017 to address the growing demand for agricultural statistics [25]. We used results from the AAS 2019—the third survey implemented by UBOS in collaboration with the Ministry of Agriculture, Animal Industries and Fisheries [26]. We extracted data from the annexes on estimated harvested area of cassava by sub-region during first and second seasons. Figure 2 shows cassava harvested area (ha) per sub-region in Uganda between two survey seasons. The total area harvested was defined as the total area planted calculated on those observations whose production is available (not missing) and higher than zero. The AAS collects data for the timespan of an agricultural year (from January to December 2019). In Uganda, the agricultural year comprises two seasons: the first season from January to June and the second season from July to December. For each season, agricultural households are interviewed twice: during post-planting and post-harvesting visits. The assignment of districts to sub-regions follows the classification of the Uganda Annual Agricultural Survey 2019, with list provided in Appendix A. The main cassava production areas are concentrated in the northern sub-regions of Lango, Teso, and West Nile.

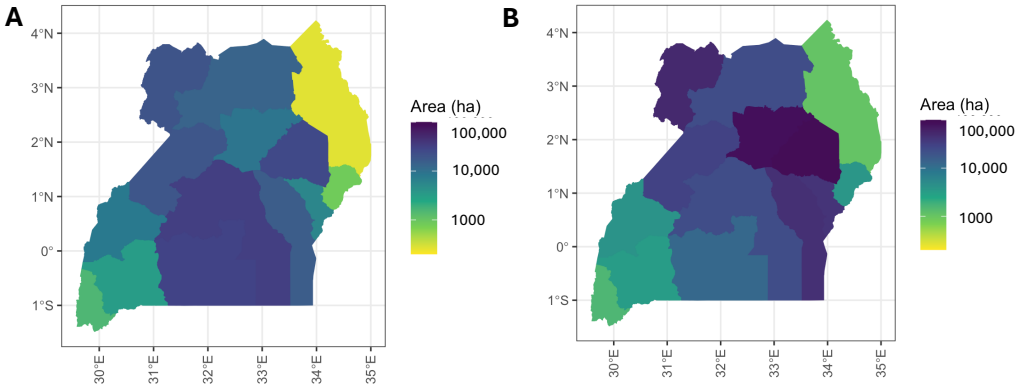


Figure 2. Cassava harvested area (ha) per sub-region in Uganda between two survey seasons: (**A**) AAS 2019 Season 1; and (**B**) AAS 2019 Season 2.

2.6. Existing Cassava Production Maps

Few cassava production maps at country, regional or continental scales have been derived from empirical approaches using demographic and economic data rather than remote sensing. CassavaMap provides a 1 km disaggregation of cassava production and harvested area using administrative unit data for 32 countries within Uganda; the estimates are standardized to 2014 FAO reported levels and downscaled based on the distribution of rural population in 2014 [27].

Similarly, the International Food Policy Research Institute (IFPRI) Spatial Production Allocation Model (SPAM) disaggregates crop production and harvested area at 10 km resolution using a cross-entropy approach that allocates sub-national statistics to pixels based on agricultural land suitability, population density, and irrigation extent [28]. For the current study, we used Global Spatially-Disaggregated Crop Production Statistics Data for 2020 Version 2.0 Release 2 [28]. This dataset compiles sub-national crop production statistics for 42 major crops, harmonized with FAOSTAT country totals and supplemented by multiple national and sub-national sources.

3. Results

3.1. Characteristics of Ground Reference Data

Of the 23,349 field polygons delineated in the Copernicus dataset, we retained only those with valid land cover labels, excluding fields marked “NA”, i.e., fields that were

not labelled (Figure 3A,B). The resulting pixel-level land cover and crop-type labels comprised 32.2% of the total 839,552 set. Cassava presence pixels accounted for 14.0% of this labelled set. The remaining non-cassava pixels were distributed across the following classes: built-up area (1.7%), forest (2.2%), natural grassland (31.6%), natural shrubs (8.3%), water (0.1%), and other crop types (39.7%).

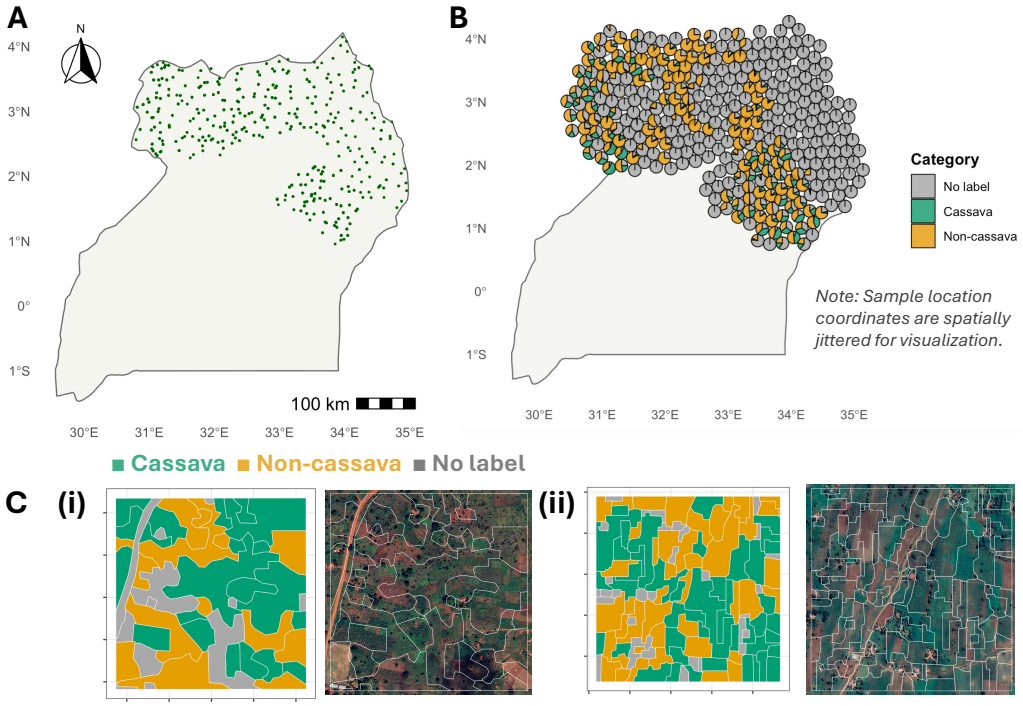


Figure 3. Ground data distribution and examples. (**A**) Map of Uganda showing the spatial distribution of ground truth samples (green dots). (**B**) Distribution and composition of ground truth samples. Sample coordinates have been jittered (max 0.1° radius) to prevent overlap. Pie charts indicate the pixel-level proportion of not labelled (dark gray), cassava (green), and non-cassava (orange) within each sample. The size of each pie chart does not encode any data. (**C**) Two illustrative samples selected for their high proportion of cassava pixels: (**i**) patch ID = 124098; and (**ii**) patch ID = 525062. For each sample, the left column shows delineated field polygons coloured according to the categories in A, and the right column shows the corresponding high-resolution satellite imagery obtained from Google Earth Engine (Image ©2025 CNES/Airbus).

Two patches of delineated fields with the highest cassava pixel proportions are presented (Figure 3C). Patch ID 124098 (West Nile region; 31.474° E, 3.250° N), has a cassava proportion of 59%, comprising 35 cassava, 35 non-cassava, and 50 NA fields, while patch ID 525062 (Eastern region; 33.844° E, 1.243° N) has a cassava proportion of 57%, comprising 86 cassava, 86 non-cassava, and 116 NA fields. Despite the larger number of fields in the second example, field polygons are physically smaller. Overall, field boundaries align well with the satellite imagery. The occasional slight misalignment may be explained by shifting planting decisions and field utilization in response to weather and market signals.

3.2. Performance of the Classifier

The labelled ground reference dataset, comprising all field data with valid labels aggregated into a single set, was used to train, test, and validate a Random Forest classifier for pixel-level cassava detection. The model was trained using 80% of the data (training set), tested using 10% of the data (testing set), and final performance was assessed on the remaining 10% of the data (validation set).

To identify the most suitable classifier for pixel-wise cassava detection, three algorithms were compared: Random Forest (RF), gradient-boosted decision tree algorithm (XGBoost), and Support Vector Machine (SVM). All models were trained on the TESSERA 128-dimensional embeddings using repeated 5-fold cross-validation with three repeats and up-sampling to address class imbalance. This procedure was implemented within the caret package in R, which provides a unified interface for model training and resampling. In practical terms, the training data were randomly split into five equal parts (folds). The model was trained four times on four of the parts and tested on the remaining fifth part, rotating which part was used for testing until each fold had served as the test set once. This process was repeated three times with different random splits to obtain more reliable performance estimates. Because cassava fields (the minority class) were much less frequent than non-cassava areas (the majority class), we applied up-sampling: randomly duplicating cassava samples during training until both classes were equally represented. This prevents the model from simply learning to predict "non-cassava" for every pixel and helps it correctly identify the rarer cassava class. The entire repeated cross-validation and up-sampling workflow was handled natively by caret's trainControl function, with the number of folds and repeats specified via the number and repeats arguments, and up-sampling enabled using the sampling = "up" option.

A full grid search was performed over the hyperparameter space of each model. For RF, the number of trees (100, 300, 500) and the number of variables randomly sampled at each split (`mtry` = 2, 65, 128) were tuned. For XGBoost, the learning rate (0.01, 0.1, 0.3), maximum tree depth (3, 6, 9), and number of boosting rounds (50, 100, 200) were optimised. For SVM, a radial basis function kernel was used, and the cost parameter 'C' (1, 10, 100) and kernel coefficient `gamma` (0.001, 0.01, 0.1) were varied. Performance was evaluated on the independent validation set using accuracy, precision, recall, and F1 score. RF achieved the highest accuracy (0.8605 ± 0.0012) and F1 (0.9244 $\pm$ 0.0008), closely followed by XGBoost (accuracy 0.8603 $\pm$ 0.0003, F1 0.9249 $\pm$ 0.0002). SVM yielded slightly lower performance (accuracy 0.8534 $\pm$ 0.0022, F1 0.9191 $\pm$ 0.0015). Based on these results and its computational efficiency for large-area inference, Random Forest was chosen as the final classifier for generating the CM24 cassava map.

After training the RF model, the performance was optimized by selecting the classification probability threshold based on the testing data. Figure 4A shows the performance metrics of the Random Forest classifier as a function of this threshold (ranging from 0 to 1). Accuracy rose from 0.323 at a threshold of 0.01 to a peak of 0.918 near thresholds of 0.42–0.44, then gradually declined to approximately 0.860 at higher thresholds up to 0.99. Precision remained at or near 1.0 for thresholds up to 0.03, after which it declined gradually to about 0.86, indicating that the model maintained very high positive predictive value across most threshold values. The weighted F1 score increased from 0.35 at a threshold of 0.01 to a peak of approximately 0.953 near thresholds of 0.43–0.44, then gradually declined to around 0.925 for thresholds above 0.90. Based on these results, the optimal classification threshold was set to 0.43 for the rest of the analysis. The random forest classifier demonstrated excellent discriminative ability in both the test and validation sets. The validation AUC was 0.9532 and the test AUC was 0.9524, indicating nearly identical and highly accurate performance with minimal overfitting (Figure 4B).

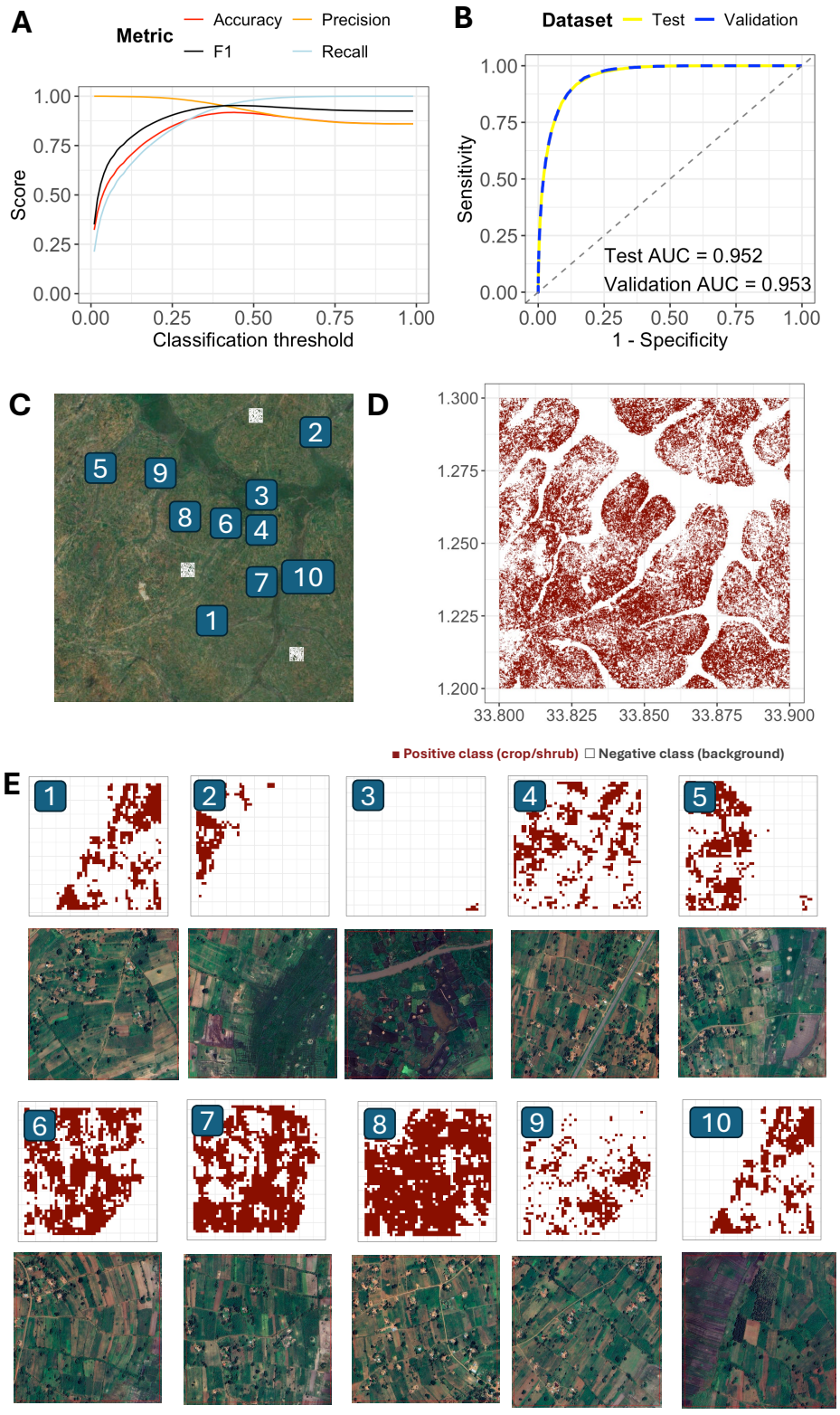


Figure 4. Performance of the RF classifier. (**A**) Values of metrics of the Random Forest classifier across classification probability thresholds for testing data. (**B**) ROC curves with AUC for testing and validation sets. (**C**) Location of Cop4Geoglam data (white polygons), and randomly selected 10 samples for visual inspection. (**D**) CM24 cassava presence (red pixels) over the same region, 10-m resolution. (**E**) CM24 cassava presence (red pixels) and corresponding high-resolution satellite imagery for 10 sites shown in B. Numbers correspond to locations of sites. High-resolution satellite imagery was obtained from Google Earth Engine (Image ©2026 CNES/Airbus).

To further assess the spatial reliability of the predictions, a visual validation was performed using high-resolution satellite imagery from Google Earth Pro (Image ©2026 CNES/Airbus). A single 10 km × 10 km tile centred at longitude 33.85, latitude 1.25 was selected. Ten 500 m × 500 m squares were randomly sampled to represent a range of cassava fractional cover, ensuring that these squares did not overlap with the Cop4Geoglam data (Figure 4C). The CM24 cassava presence map at 10 m resolution across the region showed that cassava covered approximately 36% of the total area (Figure 4D). For each square, the model's predicted cassava presence (red pixels) was overlaid on the corresponding high-resolution imagery (Figure 4E). In all ten cases, the spatial distribution of predicted cassava pixels showed good visual agreement with the actual field patterns visible in the satellite data, confirming that the model captures both large cassava blocks and small, fragmented plots.

3.3. Comparing Mapped Areas with Annual Agricultural and Existing Cassava Maps

To evaluate the accuracy of our cassava distribution map (CM24), we compared it against two existing products (CassavaMap and SPAM2020) and estimated production areas from two agricultural survey seasons (AA19S1 and AA19S2) from the 2019 AAS in Uganda. To enable comparison, we aggregated CM24 to 1 km (Figure 5A) and 10 km (Figure 5D) grid cells. For analysis, we extracted data on harvested area for Uganda from global maps in CassavaMap (Figure 5B) and SPAM2020 (Figure 5E). The pixel-wise comparison showed differences between CM24 and each of the other available product (Figure 5C,F). The difference between CM24 and CassavaMap ranged from −50.00 to 56.63 ha per 1 km grid cell (median = −0.38 ha, mean = −2.14 ha), indicating a slight negative bias overall towards CM24. The difference between CM24 and SPAM2020 at 10 km resolution was substantially larger in magnitude, ranging from −3850 to 3601 ha per 10 km cell (median = −348 ha, mean = −428 ha), reflecting systematic lower estimates by CM24 relative to SPAM2020 across most of Uganda. One notable outlier was a 10 km grid cell in Zombo district (longitude 30.92351, latitude 2.3923995: West Nile subregion), where SPAM2020 reported a high harvested area of 8287.3 ha—a value that strongly deviates from surrounding patterns and likely contributes to the large negative differences observed in that region.

Survey values differed notably between seasons, with Season 2 generally showing larger harvested areas across most subregions (Figure 5G,H). For example, in Lango (84,678 ha vs. 8856 ha) and Busoga (45,518 ha vs. 16,146 ha). At the national level, total harvested area from the two survey seasons differed considerably: 184,814 ha (Season 1) versus 355,878 ha (Season 2) [26]. CM24 estimated a total of 361,160 ha (calculated from data across both seasons), which falls between the two AAS survey estimates and is close to the AAS report for Season 2. In contrast, CassavaMap (729,253 ha) and SPAM2020 (804,638 ha) far exceeded both survey totals, indicating systematic over-prediction. Among the three map products, CM24 frequently aligned more closely with at least one of the survey seasons, whereas CassavaMap and SPAM2020 consistently estimated higher harvested area in most subregions. For instance, in West Nile, CM24 (55,245 ha) approximated Season 2 (52,737 ha), while CassavaMap (98,605 ha) and SPAM2020 (101,816 ha) were substantially higher. Agreement between each map-based harvested area estimate and the survey-based estimates was quantified using the root-mean-square error (RMSE), which measures the average deviation between map and survey values across subregions; lower RMSE indicates closer agreement. RMSE against the two survey seasons also favoured CM24: for Season 1, RMSE values were 23,295 ha (CM24), 49,732 ha (CassavaMap), and 60,129 ha (SPAM2020); for Season 2, RMSE values were 17,831 ha (CM24), 33,994 ha (CassavaMap), and 38,039 ha (SPAM2020). Overall, CM24 consistently produced harvested area estimates

that were closer to the available survey data than either CassavaMap or SPAM2020 while also delivering improved spatial granularity.

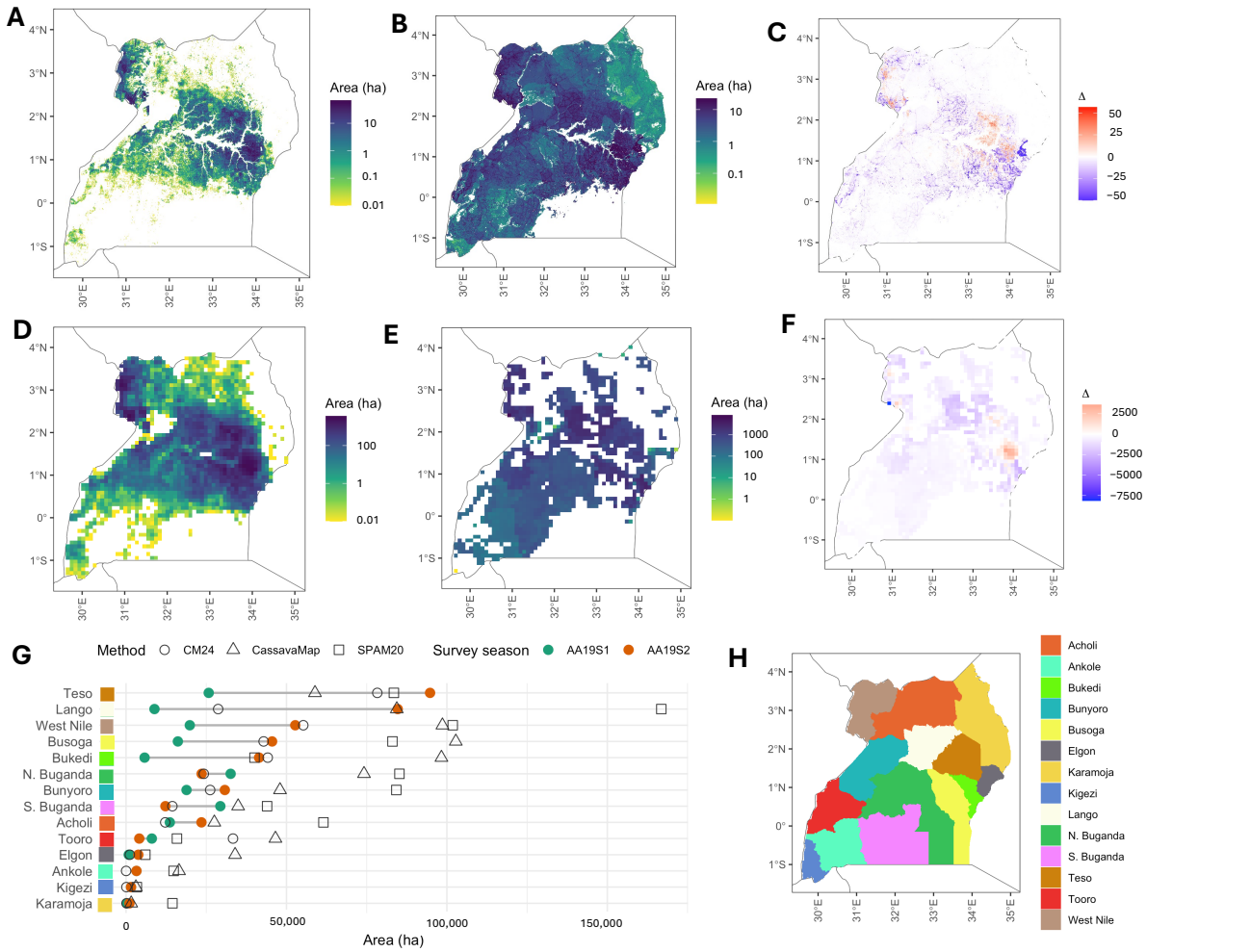


Figure 5. (**A**) CM24 cassava map aggregated to ha per 1 km grid cell. The colour scale is log-transformed. (**B**) CassavaMap harvested area in Uganda [27]. The colour scale is log-transformed. (**C**) Difference between CM24 and CassavaMap in cassava production per 1 km grid cell. (**D**) CM24 cassava map aggregated to ha per 10 km grid cell. The colour scale is log-transformed. (**E**) SPAM2020 cassava harvested area in Uganda [28]. The colour scale is log-transformed. (**F**) Difference between CM24 and SPAM2020 in cassava production per 10 km grid cell. (**G**) Comparison of cassava harvested area (ha) per subregion in Uganda between two survey seasons (AA19S1 and AA19S2) and three map-based products: CM24 (10 m map), CassavaMap, and SPAM2020. Subregions are ordered by the maximum surveyed harvested area. Gray lines connect the two survey seasons for each subregion. (**H**) Map of the sub-regions in Uganda (adapted from [26]).

4. Discussion

4.1. Technical Approach and Methodological Contributions

High-resolution crop distribution maps are essential for disease surveillance, yield forecasting, and climate adaptation, yet they remain conspicuously absent for smallholder cassava systems in Sub-Saharan Africa. By fine-tuning a Random Forest classifier on TESSERA foundation model embeddings, we generated a 10 m resolution cassava presence map for Uganda (CM24) suitable for use in heterogeneous agricultural landscapes. The visual validation against high-resolution satellite imagery confirmed that the model's spatial predictions align well with ground patterns, providing confidence for operational use.

Recent continental-scale mapping of maize and cropland in Africa [29,30] has shown that dense Sentinel-1 and Sentinel-2 time series can achieve accuracies comparable to ours. For soybean, Wang et al. [31] demonstrated 10 m mapping using phenological metrics. Cassava presents unique challenges: a flexible growing cycle of 6–24 months, frequent intercropping, and spectral similarity with other broadleaf crops. The TESSERA embeddings, which condense a full year of multi-modal observations into a compact feature space, appear to overcome these challenges without requiring manually defined phenological windows. This suggests that self-supervised representations could serve as a generalisable backbone for mapping other root crops or minor cereals in tropical regions.

4.2. Comparison with Existing Products

Existing global cassava maps (CassavaMap, SPAM2020) rely on coarse demographic and agricultural suitability proxies rather than direct remote sensing signals. Their systematic over-prediction of harvested area compared with CM24 likely reflects the difficulty of downscaling administrative statistics to pixel-related scales without a strong crop-specific phenological signal. CM24, by contrast, produces national totals that fall between the two survey seasons and tracks subregional patterns more closely, as reflected in its lower RMSE values. The extreme outlier in SPAM2020 (Zombo district) illustrates a broader vulnerability of cross-entropy allocation methods when input data contain local anomalies. Our findings suggest that foundation-model embeddings offer a viable path toward crop-specific maps that respect actual satellite-observed patterns, even in data-sparse smallholder contexts.

The comparison between CM24 and the AAS 2019 survey data should be interpreted with caution, as cassava area may have fluctuated over the five-year interval. However, the systematic over-prediction by CassavaMap and SPAM2020 relative to both CM24 and the survey data—and the closer agreement of CM24 with the survey benchmarks—suggests that the temporal mismatch does not fully explain the discrepancies between products. Nevertheless, direct validation against contemporaneous 2024 field data, were these to become available, would provide a more robust assessment of CM24 accuracy, or more generally for any contemporaneous field and remotely sensed data to assess the accuracy of the approach.

4.3. Implications for Disease Surveillance and Management

The spatial arrangement of host plants is a key driver of the spread of CMD and CBSD. Several epidemiological models have been developed for CBSD [32,33], but their predictive utility depends critically on accurate, high-resolution host distribution data. A detailed cassava presence map is therefore essential to parameterise disease spread models and to inform evidence-based surveillance and control policies at national or regional scales [34–38]. Coupling crop distribution maps with mechanistic epidemic models enables a transition from reactive to proactive plant health management, allowing targeted interventions before outbreaks escalate.

4.4. Limitations and Future Directions

Several limitations should be acknowledged. The ground reference data were collected during a single season and concentrated in northern and north-eastern Uganda. This spatial and temporal bias may reduce performance in southern Uganda, where bimodal rainfall and different cassava varieties may alter phenological profiles. The near-identical test and validation AUC values (0.9524 and 0.9532) suggest minimal overfitting to the training domain, though performance in unsampled agro-ecological zones remains untested. The magnitude of this geographical bias error could be quantified in future work by collecting independent validation samples from southern Uganda and comparing regional accuracy metrics against those obtained in the north-east.

While the TESSERA embeddings are pre-trained globally, expanding the training set to include multiple years and regions would improve generalisability, particularly for southern Uganda where bimodal rainfall patterns may alter cassava phenology. Our pixel wise classifier does not explicitly model field boundaries; this was a deliberate choice to focus on evaluating the raw feature space, and delineating individual fields was outside the scope of this work. Consequently, the resulting CM24 map exhibits some sporadic pixel scale misclassification noise and does not consistently conform to actual farmland boundary contours. Future research should integrate object oriented image segmentation algorithms or convolutional architectures that learn spatial context, and should combine land parcel boundary constraints e.g., from very high resolution imagery [30,39] or cadastral data [40], to optimize classification outputs. While visual validation against high-resolution satellite imagery provided qualitative confidence in spatial predictions, this approach is inherently subjective and does not provide quantitative error estimates. We, therefore, treated visual validation as a supplementary check rather than a primary accuracy metric. The survey data (AAS 2019) themselves carry sampling and reporting uncertainties, and direct plot-level validation of CM24 remains a priority.

4.5. Conclusions

Despite these limitations, CM24 provides a substantial step forward. Its 10 m resolution and computationally moderate requirements make it suitable for routine use by agricultural monitoring agencies and plant health services. Future work should extend the time series to capture inter-annual crop rotation, integrate with yield models, and scale the approach to neighbouring countries facing similar smallholder cassava challenges. The present study delivers a high-quality cassava map for Uganda, and the methodology is well positioned for systematic extension across Sub-Saharan Africa. The TESSERA foundation model used in CM24 was explicitly designed for label-efficient transfer and has already demonstrated strong performance in temporal transfer tasks [41], resonating with established methods for the geographical transfer of crop type classification models [42]. Building on this foundation, we propose a three-stage transfer strategy for future regional roll-out. In the first stage, fine-tuning on modest locally collected sample sets would adapt the model to country-specific rainfall regimes, planting calendars, and field sizes with minimal additional fieldwork. A second stage would use pseudo-label generation from the pre-trained model to scale training data cost-effectively, avoiding extensive new field surveys. A third tier would integrate ancillary geospatial layers (digital elevation models, soil maps, and weather data) as domain adaptation features, enhancing robustness agro-ecologically distinct regions.

Author Contributions: Conceptualization, R.R. and C.A.G.; methodology, R.R.; software, R.R.; validation, R.R.; formal analysis, R.R.; investigation, R.R.; resources, C.A.G.; data curation, R.R.; writing—original draft preparation, R.R.; writing—review and editing, R.R. and C.A.G.; visualization, R.R.; project administration, R.R.; funding acquisition, C.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Gates Foundation grant INV070408, which we gratefully acknowledge.

Data Availability Statement: https://doi.org/10.5281/zenodo.20787601 (accessed on 22 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Sub-Regions and Their Districts (Based on AAS 2019)

Table A1. Sub-regions and districts of Uganda.

Sub-Region	Districts
South Buganda	Kalangala, Masaka, Mpigi, Rakai, Ssembabule, Wakiso, Lyantonde, Bukomansimbi, Butambala, Gomba, Kalungu, Lwengo
North Buganda	Kiboga, Luwero, Mubende, Mukono, Nakasongola, Kayunga, Mityana, Nakaseke, Buikwe, Buvuma, Kyankwanzi
West Nile	Adjumani, Arua, Moyo, Nebbi, Yumbe, Koboko, Maracha, Zombo
Lango	Apac, Lira, Amolatar, Dokolo, Oyam, Alebtong, Kole, Otuke
Acholi	Gulu, Kitgum, Pader, Amuru, Agago, Lamwo, Nwoya, Omoro
Kigezi	Kabale, Kisoro, Rukungiri, Kanungu, Mitooma, Rubanda
Bunyoro	Hoima, Kibaale, Masindi, Buliisa, Kiryandongo, Kagadi, Kakumiro
Tooro	Bundibugyo, Kabarole, Kasese, Kamwenge, Kyenjojo, Kyegegwa
Busoga	Bugiri, Iganga, Jinja, Kamuli, Mayuge, Kaliro, Buyende, Luuka, Namayingo
Teso	Katakwi, Kumi, Soroti, Kaberamaido, Amuria, Bukedea, Serere
Bukedi	Busia, Pallisa, Tororo, Budaka, Butaleja, Kibuku
Elgon	Kapchorwa, Mbale, Sironko, Bududa, Bukwo, Manafwa, Bulambuli, Kween
Karamoja	Kotido, Moroto, Nakapiripirit, Abim, Kaabong, Amudat, Napak
Ankole	Bushenyi, Mbarara, Ntungamo, Ibanda, Isingiro, Kiruhura, Buhweju, Sheema

References

1. Gleadow, R.; Maher, K.; Cliff, J. Cassava. *Curr. Biol.* **2023**, *33*, R384–R386. [CrossRef] [PubMed]
2. Bacsi, Z.; Jarso, D.D. Cassava Response to Weather Variability in Eastern Africa. *Agriculture* **2026**, *16*, 209. [CrossRef]
3. Borku, A.W.; Tora, T.T.; Masha, M. Cassava in focus: A comprehensive literature review, its production, processing landscape, and multi-dimensional benefits to society. *Food Chem. Adv.* **2025**, *7*, 100945. [CrossRef]
4. CGIAR Research Program on Roots, Tubers and Bananas. RTB crop breeding in Africa shows wide-scale adoption of improved varieties. In *Crop Improvement, Adoption, and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa*; CAB International: Oxfordshire, UK, 2015.
5. Legg, J.P.; Lava Kumar, P.; Makeshkumar, T.; Tripathi, L.; Ferguson, M.; Kanju, E.; Ntawuruhunga, P.; Cuellar, W. Cassava virus diseases: Biology, epidemiology, and management. *Adv. Virus Res.* **2015**, *91*, 85–142. [CrossRef] [PubMed]
6. Adebayo, W. Cassava production in Africa: A panel analysis of the drivers and trends. *Heliyon* **2023**, *9*, e19939. [CrossRef] [PubMed]
7. Sikazwe, G.; Yocgo, R.E.E.; Landi, P.; Richardson, D.M.; Hui, C. Predicting the current and future suitable habitats of cassava and cassava brown streak disease in Africa. *East Afr. J. Sci. Technol. Innov.* **2026**, *7*, 1–18. [CrossRef]
8. Silva, D.; Ferreira, E.; Oliveira, M.; Pereira, G.; Oliveira, R. Productivity of cassava and other crops in an intercropping system. *Cienc. E Investig. Agrar.* **2016**, *43*, 159–166. [CrossRef]
9. Daraneesrisuk, J.; Ninsawat, S.; Losiri, C.; Sitthi, A. Sugarcane and Cassava Classification Using Machine Learning Approach Based on Multi-temporal Remote Sensing Data Analysis. In *Applied Geography and Geoinformatics for Sustainable Development*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 183–194. [CrossRef]
10. Wang, X.; Wang, Q.; Lai, H.; Zhang, Z.; Yun, T.; Lu, X.; Wang, G.; Lao, S.; Liao, Q.; Lu, S.; et al. A multi-sensor, phenology-based framework for mapping cassava cultivation dynamics and intercropping in highly fragmented agricultural landscapes. *ISPRS J. Photogramm. Remote Sens.* **2025**, *228*, 44–63. [CrossRef]
11. Chaiyana, A.; Khiripet, N.; Ninsawat, S.; Siriwan, W.; Shanmugam, M.S.; Virdis, S.G. Mapping and predicting cassava mosaic disease outbreaks using earth observation and meteorological data-driven approaches. *Remote Sens. Appl. Soc. Environ.* **2024**, *35*, 101231. [CrossRef]
12. Xiao, A.; Xuan, W.; Wang, J.; Huang, J.; Tao, D.; Lu, S.; Yokoya, N. Foundation models for remote sensing and earth observation: A survey. *arXiv* **2024**, arXiv:2410.16602. [CrossRef]

13. Szwarcman, D.; Roy, S.; Fraccaro, P.; Gíslason, Þ.E.; Blumenstiel, B.; Ghosal, R.; de Oliveira, P.H.; de Sousa Almeida, J.L.; Sedona, R.; Kang, Y.; et al. Prithvi-EO-2.0: A Versatile Multitemporal Foundation Model for Earth Observation Applications. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 4400120. [CrossRef]
14. Ma, Y.; Shen, Y.; Swatantran, A.; Lobell, D. Harvesting AlphaEarth: Benchmarking the geospatial foundation model for agricultural downstream tasks. *arXiv* **2026**, arXiv:2601.00857. [CrossRef]
15. Astruc, G.; Gonthier, N.; Mallet, C.; Landrieu, L. AnySat: One earth observation model for many resolutions, scales, and modalities. *arXiv* **2024**, arXiv:2412.14123. [CrossRef]
16. Feng, Z.; Atzberger, C.; Jaffer, S.; Knezevic, J.; Sormunen, S.; Young, R.; Lisaius, M.; Immitzer, M.; Jackson, T.; Ball, J.; et al. TESSERA: Temporal embeddings of surface spectra for earth representation and analysis. *arXiv* **2025**, arXiv:2506.20380. [CrossRef]
17. Uganda Bureau of Statistics. *The National Population and Housing Census 2024–Final Report—Volume 1 (Main)*; Uganda Bureau of Statistics: Kampala, Uganda, 2024.
18. Danielson, J.J.; Gesch, D.B. Global multi-resolution terrain elevation data 2010 (GMTED2010). In *Technical Report Open-File Report 2011–1073*; U.S. Geological Survey: Reston, VA, USA, 2011. [CrossRef]
19. Ngoma, H.; Wen, W.; Ojara, M.; Ayugi, B. Assessing current and future spatiotemporal precipitation variability and trends over Uganda, East Africa, based on CHIRPS and regional climate model datasets. *Meteorol. Atmos. Phys.* **2021**, *133*, 823–843. [CrossRef]
20. Phillips, J.; McIntyre, B. ENSO and interannual rainfall variability in Uganda: Implications for agricultural management. *Int. J. Climatol.* **2000**, *20*, 171–182. [CrossRef]
21. HarvestChoice; International Food Policy Research Institute (IFPRI). *Agro-Ecological Zones for Africa South of the Sahara*; HarvestChoice: Washington, DC, USA, 2015. [CrossRef]
22. European Commission, Joint Research Centre. *Uganda AOI*; European Commission: Brussels, Belgium, 2026. [CrossRef]
23. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
24. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756. [CrossRef]
25. Ponzini, G.; Baryahirwa, S.; Brunelli, C.; Ilukor, J.; Kilic, T.; Mugabe, S.; Mupere, A.; Okello, P.; Oumo, F.; Ssennono, V. The integration of socio-economic and agricultural surveys by national statistical offices: The case of the Uganda Harmonized Integrated Survey 1. *Stat. J. IAOS* **2022**, *38*, 141–161. [CrossRef]
26. Uganda Bureau of Statistics. *Annual Agriculture Survey 2019 Report*; Uganda Bureau of Statistics: Kampala, Uganda, 2022. Available online: https://www.ubos.org/wp-content/uploads/publications/04_2022AAS2019_Report.pdf (accessed on 12 May 2026).
27. Szyniszewska, A.M. CassavaMap, a fine-resolution disaggregation of cassava production and harvested area in Africa in 2014. *Sci. Data* **2020**, *7*, 159. [CrossRef] [PubMed]
28. International Food Policy Research Institute (IFPRI). *Global Spatially-Disaggregated Crop Production Statistics Data for 2020 Version 2.0 Release 2*; International Food Policy Research Institute (IFPRI): Washington, DC, USA, 2024. [CrossRef]
29. Abdelrahim, N.; Jin, S. Continental maize mapping and distribution in Africa by integrating radar and optical imagery. *Environ. Monit. Assess.* **2025**, *197*, 1072. [CrossRef] [PubMed]
30. Rufin, P.; Hammer, P.L.; Thomas, L.F.; Lisboa, S.N.; Ribeiro, N.; Sitoe, A.; Hostert, P.; Meyfroidt, P. National-scale field delineation in Mozambique refines our understanding of cropland distribution, field size, and deforestation actors. *Environ. Res. Lett.* **2026**, *21*, 084009. [CrossRef]
31. Wang, R.; Zhang, J.; Lu, X.; Fu, Z.; Cai, G.; Liu, B.; Li, J. JM-Guided Sentinel 1/2 Fusion and Lightweight APM-UNet for High-Resolution Soybean Mapping. *Remote Sens.* **2025**, *17*, 3934. [CrossRef]
32. McQuaid, C.F.; Sseruwagi, P.; Pariyo, A.; van den Bosch, F. Cassava brown streak disease and the sustainability of a clean seed system. *Plant Pathol.* **2015**, *65*, 299–309. [CrossRef] [PubMed]
33. Godding, D.; Stutt, R.O.J.H.; Alicai, T.; Abidrabo, P.; Okao-Okuja, G.; Gilligan, C.A. Developing a predictive model for an emerging epidemic on cassava in sub-Saharan Africa. *Sci. Rep.* **2023**, *13*, 12603. [CrossRef] [PubMed]
34. Godding, D.; Stutt, R.; Savi, M.; Ahanhanzo, C.; Tiendrebeogo, F.; Doungous, O.; Godefroid, M.; Bakelana, Z.; Mavoungou, J.; Oppong, A.; et al. Predicting the cross-continental spread of the cassava brown streak disease epidemic in sub-Saharan Africa. *openRxiv* **2025**, 1–28. [CrossRef]
35. Retkute, R.; Gilligan, C.A. Developing a spatio-temporal model for banana bunchy top disease: Leveraging remote sensing and survey data. *Front. Plant Sci.* **2025**, *16*, 1521620. [CrossRef] [PubMed]
36. Retkute, R.; Zandjanakou-Tachin, M.; Omondi, B.; Agoi, U.; Vodounou, Y.; Akofodji, H.; Akpla, E.; Dossou, L.; Médénou, E.; Etchiha, A.; et al. Controlling banana bunchy top disease in Benin: Crop protection strategies with socio-economic perspectives. *Plants People Planet* **2025**, 1–17. [CrossRef]
37. Retkute, R.; Gilligan, C. Cost-effective early detection of banana bunchy top disease: Insights from spatio-temporal modelling in Benin. *openRxiv* **2026**, 1–20. [CrossRef]
38. Smith, J.W.; Stutt, R.O.J.H.; Retkute, R.; Mona, T.; Thurston, W.; Bacha, N.; Gutu, K.; Horo, J.T.; Alemayehu, Y.; Hodson, D.; et al. Evaluating a landscape-scale model to forecast wheat stem rust. *Environ. Res. Lett.* **2026**, *21*, 014034. [CrossRef]

39. Lu, R.; Zhang, Y.; Huang, Q.; Zeng, P.; Shi, Z.; Ye, S. A refined edge-aware convolutional neural networks for agricultural parcel delineation. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *133*, 104084. [CrossRef]
40. Kim, H.S.; Song, H.; Jung, J. Cadastral-to-Agricultural: A Study on the Feasibility of Using Cadastral Parcels for Agricultural Land Parcel Delineation. *Remote Sens.* **2024**, *16*, 3568. [CrossRef]
41. Lisaius, M.; Keshav, S.; Blake, A.; Atzberger, C. Embedding-based crop type classification in the groundnut basin of Senegal. *arXiv* **2026**, arXiv:2601.16900. [CrossRef]
42. Nyborg, J.; Pelletier, C.; Lefèvre, S.; Assent, I. TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS J. Photogramm. Remote Sens.* **2022**, *188*, 301–313. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.