Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine learning for high-throughput field phenotyping and image processing provides insight into the association of above and below-ground traits in cassava (*Manihot esculenta* Crantz)

Michael Gomez Selvaraj (m.selvaraj@cgiar.org)
Centro Internacional de Agricultura Tropical https://orcid.org/0000-0003-2394-0399
Manuel Valderrama
Centro Internacional de Agricultura Tropical
Diego Guzman
Centro Internacional de Agricultura Tropical
Milton Valencia
Centro Internacional de Agricultura Tropical
Henry Ruiz
Texas A&M University College Station
Animesh Acharjee
University Hospitals Birmingham NHS Foundation Trust

Research

Keywords: Automated aerial image processing, above-ground biomass, cassava, machine learning, multispectral UAV imagery, root yield prediction

Posted Date: June 3rd, 2020

DOI: https://doi.org/10.21203/rs.2.24148/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Plant Methods on June 14th, 2020. See the published version at https://doi.org/10.1186/s13007-020-00625-1.

Abstract

Background: Rapid non-destructive measurements to predict cassava root yield over the full growing season through large numbers of germplasm and multiple environments is a huge challenge in Cassava breeding programs. As opposed to waiting until the harvest season, multispectral imagery using unmanned aerial vehicles (UAV) are capable of measuring the canopy metrics and vegetation indices (VIs) traits at different time points of the growth cycle. This resourceful time series aerial image processing with appropriate analytical framework is very important for the automatic extraction of phenotypic features from the image data. Many studies have demonstrated the usefulness of advanced remote sensing technologies coupled with machine learning (ML) approaches for accurate prediction of valuable crop traits. Until now, Cassava has received little to no attention in aerial image-based phenotyping and ML model testing.

Results : To accelerate image processing, an automated image-analysis framework called CIAT Pheno-i was developed to extract plot level vegetation indices/canopy metrics. Multiple linear regression models were constructed at different key growth stages of cassava, using ground-truth data and vegetation indices obtained from a multispectral sensor. Henceforth, the spectral indices/features were combined to develop models and predict cassava root yield using different Machine learning techniques. Our results showed that (1) Developed CIAT pheno-i image analysis framework was found to be easier and more rapid than manual methods. (2) The correlation analysis of four phenological stages of cassava revealed that elongation (EL) and late bulking (LBK) were the most useful stages to estimate above-ground biomass (AGB), below-ground biomass (BGB) and canopy height (CH). (3) The multi-temporal analysis revealed that cumulative image feature information of EL+early bulky (EBK) stages showed a higher significant correlation (r = 0.77) for Green Normalized Difference Vegetation indices (GNDVI) with BGB than individual time points. Canopy height measured on the ground correlated well with UAV (CHuav)-based measurements (r = 0.92) at late bulking (LBK) stage. Among different image features, normalized difference red edge index (NDRE) data were found to be consistently highly correlated (r = 0.65 to 0.84) with AGB at LBK stage. (4) Among the four ML algorithms used in this study, k-Nearest Neighbours (kNN), Random Forest (RF) and Support Vector Machine (SVM) showed the best performance for root yield prediction with the highest accuracy of R 2 = 0.67, 0.66 and 0.64, respectively.

Conclusion : UAV platforms, time series image acquisition, automated image analytical framework (CIAT Pheno-i), and key vegetation indices (VIs) to estimate phenotyping traits and root yield described in this work have great potential for use as a selection tool in the modern cassava breeding programs around the world to accelerate germplasm and varietal selection. The image analysis software (CIAT Pheno-i) developed from this study can be widely applicable to any other crop to extract phenotypic information rapidly.

Background

Cassava (*Manihot esculenta* Crantz), commonly referred as manioc (French), yuca (Spanish), and different names in local regions, is a tropical root crop native to South America [1], and relied by more than 800 million people as a staple food source [2]. Its versatile nature, it is often referred to as the “drought, war and famine crop of the developing world” [3], places it among the most adaptive crops during climate change. Early vigor, rapid root bulking, higher root yield, resistance to major pest and diseases, waxy cassava are the most important targeted traits in cassava breeding programs around the world [4]. Conventional breeding continues to be the main method for cassava varietal development worldwide and had a strong impact on addressing the constraints of cassava growers [5]. Traditional methods of selecting breeding/germplasm lines are labor intensive and destructive to nature, limiting the quantitative and repeated assessments in long-term research [6,7]. Therefore, establishing a non-destructive and real time monitoring tool to measure above and below-ground cassava traits are very necessary [8]. Exploring non–destructive selection methods has always been a priority in cassava breeding programs. Therefore, efforts have been taken to reduce the cassava selection cycle and develop non-destructive, low-cost phenotyping methods that precisely measure the root characteristics in the field [8–12]. Though good progress in digital phenotyping has been made, so far, no studies have been devoted to the development of non-invasive high-throughput field phenotyping (HTFP) tools and machine learning models that estimate cassava canopy traits and root yield prediction through aerial imaging. In cassava breeding programs, the establishment of non-destructive phenotyping tools, root yield prediction models can allow the early selection of elite genotypes, allowing the optimization of resources and time [13]. Digital and rapid phenotyping approaches are increasingly considered important tools for rapid advancement of genetic gain in breeding programs [14].

UAV are being used to measure with high spatial and temporal resolution capable of generating useful information for plant breeding tasks [15–17]. In the era of digital revolution, aerial image phenotyping [18–20] and ML models could predict crop yield performance [21–27] in a non-invasive means with a greater accuracy [28–31]. Efficient selection of desired phenotypes through HTP across large field populations could be achieved through incorporating ML methodologies such as, automated identification, classification, quantification and prediction [20]. To be constructive to breeding programs, phenotyping methods must be robust, automated, sensitive, and amenable to plot sizes. The ability to get more rapid growth responses of genetically different plants in the field and transmit these responses to individual genes, novel technologies such as proximal sensing, robotics, integrated computational algorithms and robust automated aerial image analytical frameworks are urgently needed [7].

Even though, UAV and sensor technologies (hardware) shows greater progress with more automation and integration, processing the massive amount of generated image data such as data management, image analysis, and result visualization of large-scale phenotypic data sets [32] from aerial phenotyping systems requires robust analytical framework for data interpretation [33]. Few commercial software are available that systematize image calibration and correction, obtaining good field maps of the studied variable. But these platforms are often developed and delivered by specific enterprises where the original hardware and software are patent protected and henceforth cannot be adapted or modified to meet particular research needs [34]. Moreover, new developments target real-time processing on-board in aerial imaging platforms, providing direct vegetation indices (VIs) maps to make rapid decisions [35]. Despite these improvements, there are middle steps that require some level of manual interface, which slow the progress, such as the recognition of coded GCP, calibration panel recognition and correction, defining region of interest, extracting plot-level data [32], batch and multi-threading processing.

In this paper, we are describing a robust feature extraction platform for aerial image processing called CIAT Pheno-i, with which we validated the developed framework using cassava time series aerial images collected from two consecutive field trials (2016-2018). Since no studies have been reported on UAV based cassava high-throughput phenotyping and root yield prediction, the specific objectives of this study is (1) to develop simple and rapid aerial image analysis framework (CIAT Pheno-i) for retrieving cassava canopy variables and VIs from multispectral (MS) time series images; 2) to find promising image based canopy metrics and VIs to estimate above and below-ground biomass of cassava over different phenological stages; and (3) to develop robust ML models to predict cassava root yield using image features.

Materials And Methods

Experimental site and trial conditions

To validate the performance of CIAT Pheno-i, two field trials, trial one was planted on December 2016 and harvested in November 2017; trial two was planted in December 2017 and harvested in December, 2018, these trials were conducted at the International Center for Tropical Agriculture (CIAT) headquarters Valle del Cauca, Cali, Colombia at 970.67 m.a.s.l (3°30'29.21"N -76°20'53.98"W) (**Figure 1a**). Climate and experimental conditions were characterized for both trials (**Table 1**). For both trials, we selected four contrasting genotypes GM3893-65, CM523-7, MPER-183, and HMC-1, representing three types of canopy architecture; cylindrical, open and compact [36] and morphological and agronomic growth descriptors are listed in **Table S1** and **Table S2**. The trial one was established in 0.8 hectares under a split-plot design with three replications and a total of 135 plots (3.0, long x 9.6 wide) with staggered planting (Nine planting dates from December 2016 to August 2017) (**Table 1**). Cuttings were planted of 1.5 m between hills and 2.4 m between rows and water management was applied by the surface irrigation system from planting to seven months, using approximately 4000 m^3 per hectare. The second trial was planted in 0.6 hectares with four replications per genotype and plot size of 9.6 m long and 9.6 m wide (**Table 1**). Cuttings were planted of 1.2 m between hills and 2.4 m between rows and water management was applied with an efficient drip irrigation system from planting to seven months, using approximately 900 m^3 per hectare. In both trials, stem cuttings between 20 to 25 cm were planted vertically into the soil, leaving exposed three buds. Weeds were controlled by hand weeding, brush-cutter, and applying herbicides in late cassava stages. Standard agronomic, insects and diseases management practices were followed. A recommended dose of diammonium phosphate (DAP) and potassium chloride (KCL) were applied at the rate of 35.89 and 179 $Kg.ha^{-1}$, respectively.

Ground-truth measurements

Cassava agronomic traits such as leaf area index (LAI), canopy height (CH), above-ground biomass (AGB) and below-ground biomass (BGB) were acquired as ground-truth measurements. Five plants per plot were measured using LICOR LAI-2200C Plant Canopy Analyzer [37] during trial two. CH was sampled from soil level to the upper canopy at all four important phenological stages: elongation (EL), early bulking (EBK), late bulking (LBK) and dry matter accumulation (DMA) in both the trials. Each phenological stage is defined in **Table 1**. CH of 21 and five plants per plot were measured in trial one and two, respectively. The AGB and BGB were measured at the harvest time in both the trials using a conventional scale with the accuracy of 1g. For AGB, three and five plants per plot were sampled in trial one and two, respectively. For BGB, 15 and 45 plants per plot were sampled in trial one and two, respectively.

UAV platform and images acquisition

In this study, aerial multispectral (MS) time-series images were obtained using a MS camera (MicaSense RedEdge 3) mounted on a commercial UAV DJI S1000 octocopter (**Figure 1b**). The MS camera has five spectral bands - Blue, Green, Red, near-infrared (NIR), and Red Edge (RE) with the wavelengths of 455 - 495 nm, 540 - 580 nm, 658 - 678 nm, 800 - 880 nm and 707 - 727 nm, respectively (**Figure 1c**). The camera was attached to UAV by one plate with a shock absorption rubber/spring damping suspension system to protect against any vibration and to ensure better quality of the images. Six automated PhotoScan coded target detection (concentric rings) as ground control points (GCPs) were printed on a 50 × 50 cm plastic sheet (**Figure 1e**) and evenly distributed within the field trial (**Figure 1f**). These GCPs were georeferenced using the highly accurate RTK-GPS (Real-Time Kinematic Global Positioning System, South, Galaxy G1, China) with a horizontal accuracy of 0.25m and a vertical accuracy of 0.5m, which was used for geometric corrections (**Figure 1g**). These GCPs were maintained until all the UAV images were acquired. The automatic fly mission was performed using DJI Ground Station Pro Application (DJI GS Pro, China). Before each image acquisition, one image was taken to the MicaSense reflectance panel for radiometric calibration (**Figure 1h**). Each image acquisition was taken between 10:00 to 14:00 UTC-05:00. In order to achieve overlapping of 75% vertically and horizontally, we triggered the camera using the UAV DJI A3 flight controller and Arduino Nano as an interface configured by DJI GS app (**Figure 1d**). The altitude for image acquisition was between 30 and 40 m above ground level (from 2.7 to 5.4 cm per pixel). DJI S1000, batteries, and multi-sensors weights 3kg. DJI S1000 UAV includes a Global Navigation Satellite System (GNSS), an inertial measurement unit (IMU), barometer and compass; all these components aid in position accuracy and vertical stability of the UAV during image acquisitions. The time series UAV images captured at different phenological stages at trial one and two are listed in **Table 1** and these acquired time series images used to create the orthomosaic employed structure from motion (SfM) were listed in (**Table S1 and Table S2**).

Image data processing

Generation of orthomosaic and digital elevation models

To ensure the reflectance quality of the orthomosaic, we followed the steps suggested by Agisoft and MicaSense RedEdge cameras (Agisoft, https://bit.ly/32swtn2). These steps include the usage of the MicaSense downwelling light sensor to fix any illumination issues caused by the weather

conditions and MicaSense reflectance calibration panel. The acquired images were processed through Agisoft MetaShape Pro software (Version 1.2.2, Agisoft LLC, http://www.agisoft.com) and its Python API (Application Program Interface) generates and exports a five-band orthomosaic and digital elevation models (DEM) automatically as GeoTIFF format. Our processing workflow includes following nine main steps (1) Uploading UAV images, (2) calibration, (3) GCPs detection and geo-tagging, (4) photo alignment, (5) camera optimization (6) build dense point cloud, (7) build DEM, (8) build orthomosaic (9) export DEM and orthomosaic (**Figure S1**). In step three, coded GCPs are automatically detected through Agisoft Metashape API (**Figure 1e**).

Comparison of manual and automatic orthomosaic and DEM generation

In order to evaluate the efficiency of the Agisoft Metashape Python API, we generated orthomosaic and DEM using manual (M1-M8) and auto mode (A1-A8) from MS and RGB datasets. All data sets (MS and RGB) were processed using the image processing workflow listed in **Figure S1**.

CIAT Pheno-i image analysis framework

The CIAT Pheno-i is a web-based application (http://pheno-i.ciat.cgiar.org/), designed to extract UAV derived vegetation indices (VIs) and canopy metrics such as canopy height (CHuav), canopy cover (CCuav) and canopy volume (CVuav) rapidly. Canopy height defines the 95th percentile pixel height of the canopy point cloud. Canopy cover is the pixel surface area covered by the canopy. Canopy volume, is the total volume under observed canopy values, which is derived as follows where is the pixel associated to the plot. CIAT Pheno-i admits MS orthomosaics and DEM as an input and visualizes them as VIs maps (**Figure S2**). Users have the privilege to select their Region of Interest (ROI) using shapefiles and perform radiometric calibration, if necessary. Currently, eight VIs (**Table 2**) [17,38–43] and three canopy metrics could be rapidly generated through Pheno-i and users can visualize real-time data captured over multiple timing points during the crop development.

CIAT Pheno-i software architecture

CIAT Pheno-i back-end:

On the top of a PostgreSQL database model, two main components constitute the Pheno-i back-end: A Python library, where the core algorithms in the pipeline were implemented (**Figure 2 and Figure 3**), and a REST (REpresentational State Transfer) API that allows the data processing through HTTP protocol. Most of the functions in the library were optimized using Numba, a python package that translates Python functions to optimized machine code, which could be executed in a parallel way on the CPU or the GPU. In addition to this, geo-spatial data manipulation, machine learning algorithms, GDAL, and Scikit-Learn were also employed. The following steps described below were coded in the CIAT Pheno-i python library.

Step 1. Radiometric calibration: Using orthomosaics **(Figure 3a),** Pheno-i back-end implements Empirical Line Calibration (ELC) process **(Figure 3b)** using ground targets, allowing the user to calibrate orthomosaics after the flight. Before implanting the ELC process, the pixel digital numbers should range from 0 to 65535 corresponding to a 16-bit standard GeoTIFF format, after applying ELC the pixel values were converted to reflectance values between 0 and 1.

Step 2. Crop masking and Vegetation indices calculation: To segment cassava canopy, the green minus red (GMR) processing was used [44]. The binarization of GMR was determined by the Otsu method to perform clustering-based image thresholding [45], which implies the reduction of a gray level image to two-pixel values (0 and 1) and this binary image was used to select and discard the pixels associated with the soil **(Figure 3c)**. Using five camera channels (B: blue, G: green, R: red, RE: red-edge, and NIR: near-infrared), eight normalized vegetation indices (VIs) were intended **(Table 2**.)

Step 3. Plot-level data extraction: Using the calibrated version of the orthomosaic, the boundaries of each plot ids are defined using an ESRI Shapefile format polygon. Then, shapefile was further used to select and extract the pixel values to compute statistics such as mean, variance, median, standard deviation, sum, minimum and maximum **(Figure 3e)**.

CIAT Pheno-i web: A single page app (SPA) was developed using React.js and Redux. This web application can be executed using any modern web browser (IE 11, Edge >=14, Firefox >= 52, Chrome >= 49, Safari >= 10). The user interface follows the Material-UI v4.7.0 (https://material-ui.com/) guide design, LeafletJS v1.6.0 API (https://leafletjs.com/) was used to draw the geo-referenced orthomosaics and polygons in an OpenStreetMap (https://www.openstreetmap.org/). **Figure S3** and **Figure S4** shows the overall architecture and database schema implemented.

CIAT Pheno-i back-end performance

To evaluate the CIAT Pheno-i back-end performance, a single and multi-thread analysis were performed under server and workstation platforms over 50 different datasets. Hardware and software specifications are listed in **Table S3**.

Statistical analysis

To investigate the relationship between agronomic traits, VIs and canopy metrics, we conducted Pearson correlations, where the traits were calculated using a pearsonr function from Python SciPy (https://www.scipy.org/) package. Pearson's correlation coefficients and a *P-value* less than 0.05 was considered significant.

Dataset preparation

In order to validate Pheno-i analysis, a Comma Separated Values (CSV) file with 693 characteristics was generated. Four machine-learning algorithms such as Support Vector Machine (SVM), k-Nearest Neighbours (kNN), Random Forest (RF) and Artificial Neural Networks (ANN) were evaluated to predict cassava root yield. For the pre-processing, data scaling between -1 and 1 and a Box Cox transformation were performed to achieve a normal distribution [46]. Principal component analysis (PCA) and principal component regression (PCR) [47] was applied to compare performance and reduce the model complexity providing a lower-dimensional representation of predictor variables and to avoid multi-collinearity between predictors [48–50]. To analyze the data at different growth stages, a multi-temporal VIs technique was applied [27]; this procedure increases the predictor variables from 77 to 693 per timing point accumulating the VIs value per phenological stage.

Machine learning (ML) model development

[Please see the supplementary files to view this section.]

Assessing the quality of the model

Based on the experimental field design, a total of 609 samples were used to develop the models, three data repetitions (454 samples) were used to train and, one last repetition to test (155 samples). Regression model hyper-parameters were tuned using grid-search with ten-fold cross-validation to reduce variability and over-fitting while modeling; methods provided with scikit-learn Python package. To assess the accuracy and performance between models root median square error (RMSE), relative root mean square error (RRMSE) and the coefficient of determination (R^2) were used. Ten-fold cross-validation was performed over SVM, RF and kNN to get optimal hyper-parameters that minimize the error and stochastic gradient descendant over ANN model to reduce the training error.

Results And Discussion

CIAT Pheno–i: an automated image analysis framework for HTFP

The increased use of UAVs in field phenotyping considerably decreased the hardware costs, however, image processing is the major challenge to the crop phenotyping scientists around the world [56]. As mentioned in the introduction, midway steps to extract information from the plot level field experiments need full automation and integration. Therefore, a need for accurate, robust, and automated analysis framework building orthomosaics and extract phenotyping information corresponding to each image of micro-plots (breeding) or large scale (precision agriculture) field experiments is necessary. Here, we are describing the Pheno-i image analysis software (**Additional File 1**) developed by CIAT phenomics platform (https://phenomics.ciat.cgiar.org/) and the automated orthomosaic generation pipeline. The primary criterion for any image analysis software should be cost effective, easy-to-use and rapid generation of actionable data from time-series images irrespective of experimental plot sizes. Making use of Agisoft Metashape Python API, the orthomosaic and DEM generation process was automated (Figure S1), achieving a reduction in time of ~30%, saving ~1.1h for RGB imagery and ~0.33h for MS imagery (**Figure S5**), compared to our manual processing method. CIAT Pheno-i back-end image analysis software design brings a significant improvement over any regular single thread Python implementation reducing the processing time of MS imagery processing up to 5x (**Figure 4**). Aforesaid processing time was calculated using two different CPU architectures as seen in **Table S3**. Our CIAT Pheno-i front-end software design comes with the advantage for the user to create, upload, calibrate, visualize, and analyze orthomosaics in a map-based canvas, giving a privilege to a non-programmer to analyze his own data through the internet. The image analysis report comes in CSV format with a timestamp and a reference to a quantified plot level data, in which the data can be used either to develop plant models or just to monitor the crop health status. We offered CIAT Pheno-i as a simple and easy to use solution to extract plot/plant-level information.

We validated the developed platform using proof-of-concept experiments with cassava genotypes over the two seasonal field trials to demonstrate the end-to-end application. The results obtained from the platform are described below.

High-throughput field phenomics for aerial imaging of cassava

UAV offers very attractive alternatives such as, convenient operation, high spatial and temporal resolutions with reasonable spatial coverage [57–59], makes it possible to document the within-microplot variability in phenotyping field experiments [60,61]. UAV, a current and an invaluable tool for crop monitoring at large scale (e.g., [27,59,62–65], has been proved to be useful for estimating canopy height and biomass in crops including rice [65], wheat [66] maize [30], sorghum [67] and peas [17]. However, in cassava, the UAV based high-throughput phenotyping methods need to be standardized for feasibility and accuracy in estimating various phenotyping parameters such as, biotic and abiotic stresses. So far, most studies have attempted to correlate morpho-physiological data with the productive potential (root yield) of the genotypes at the end of the crop cycle [68]. Subsequently, these pre-breeding field experiments go through long

selection cycles, leading to high maintenance costs. The correlation analysis between important breeding traits at different phenological stages and UAV image derived VIs are discussed below.

Relationship between UAV images derived features and canopy height

Canopy height (CH) is a key factor in cassava root yield, dry matter, leaf area, and plant architecture [69]. Collecting CH within cassava field breeding programs are labor intensive and prone to assessment error. In this study, orthomosaics and DEMs were generated using Methashape Agisoft API. Canopy metrics (CHuav, CCuav and CVuav) and VIs derived from high-resolution MS images (2.7 cm x pixel) were extracted through our CIAT Pheno-i web-based application. The pearson's correlation analysis between UAV features (VIs, CHuav, CCuav and CVuav) and canopy height (CH) at EL and LBK stage showed that the UAV feature are positively correlated (**Figure 5c and Figure 7a**), except during the trial two, where most of the VIs showed low and negative correlations at DMA stage **(Figure 7a)**. This low or poor correlation is mainly due to the saturation of VIs at later stages of growth and crop lodging. Significant correlation was found at EL stage between manually estimated CH and CHuav (**Figure 8a**). However, the best relationship was reached at the late bulking stage for both the trials with *r* values 0.89 and 0.92, respectively (**Figure 5c, Figure 7a, and Figure 8b**). Similar results were found in cotton using DEMs from MS cameras [70]. In trial one, among the VIs, NDRE index showed significant relationship with CH manually with an *r* value of 0.83 at LBK stage. **(Figure 5c)**. The CH data collected by the UAV were credible and the correlation with ground-truth measurement was very high. Therefore, UAV based CH measurements in cassava has great potential for use in studies of physiological and genetic mapping experiments.

Relationship between UAV metrics and canopy structure related traits

Time series measurements of canopy related traits are very useful to develop crop growth curves. Estimating AGB traits such as canopy volume is laborious, destructive and time-consuming and therefore needs an easier and convenient method [71]. In cassava, AGB can provide valuable insights into understanding the carbon assimilation mechanism and storage root development. In this paper, canopy metrics such as CCuav and CVuav across the phenological stages showed positive significant relationship with AGB. During the trial one and two, significant correlation (r = 0.80 and r = 0.54, respectively) was found between CCuav and AGB at LBK stage (**Figure 5b and Figure 6b**). A similar relationship was previously reported between dry leaf biomass and UAV derived green CC [72]. Also, at LBK stage a similar relationship (r = 0.70) was found between CVuav and BGB during the trial one (**Figure 5a**). High-throughput canopy metrics tools developed from this study could provide quantitative data for novel traits that define canopy structure. Recurrent measurement offers time-series data from which we can estimate growth rates and dynamics. Such non-invasive measurements are very useful to understand genotype specific responses to environmental stresses during the growth period. Cassava canopy structure parameter data can also contribute to the development of root yield prediction models and could help cassava breeders in the selection procedure by providing early hints on the performance of novel lines.

Correlation between LAI and UAV derived features

The leaf area index (LAI) refers to the per unit area of the one-sided leaf per unit area of ground surface. The maximum LAI in cassava ranges from 4 to 8, depending on the cultivar, the atmospheric and edaphic conditions that prevails during crop growth stages [73]. Selection for higher LAI should favor high root yield, since there is an optimum relationship between root yield and LAI [68]. Positive contribution of LAI with cassava yield has also been reported by [74], and [75] also reported significant high correlation between ground cover and LAI in grass, legume and crucifer crop. Measuring LAI is a tedious [76] and time-consuming process, and an image trait complimenting LAI can be very useful. In order to establish this relationship, in trial two, LAI was measured and the correlation analysis was performed with UAV derived canopy metrics and VIs. The results of canopy metrics (CCuav and CVuav) and VIs showed highly significant and positive correlation with LAI in all the tested phenological stages, whereas, CCuav and CVuav correlated with DMA with *r* value of 0.56 (**Figure 7b**). Among the tested VIs, NDREI showed highly significant correlation with LAI at EL and DMA stage with *r* values of 0.53 and 0.63, respectively (**Figure 9a, d**); whereas, the correlation decreased slightly with the bulking stages (EBK and LBK) (**Figure 9b, c**). Additionally, highly significant correlations were found with LAI and NDVI at EL and DMA stages with *r* values of 0.55 and 0.59, respectively (**Figure 7b**). Strong correlation between NDVI and LAI using UAV images has also been reported in different crops such as rice [65], sorghum [67]**;** for NDREI in bread wheat [77]. These results indicate that NDREI could explain the green leaf area during senescence.

Relationship between UAV features and above-ground biomass

Breeding for early vigor, fast growing cassava genotypes is ideal to tackle several issues especially in early stages of crop management. Vigorous and early growth cultivars were less sensitive to lack of weed control than non-vigorous slow growth types. Above-ground biomass (AGB) estimation in cassava, is a most laborious and time-consuming method, requires a multi-step process: crop sacrifice from the field plot, oven dried before being weighed to assess the fresh and dry biomass of each sample. This multi-step destructive process is prone to error, from variability in the area within the plot sampled, to the potential loss of material while collecting and transporting [6]. In this present study, we estimated fresh canopy biomass in cassava using remote aerial imaging methods. Our results from both the trials revealed significant positive correlations between VIs (NDRE, NDVI, GNDVI, BNDVI, NDREI, NPCI and GRVI) and AGB, at three different phenological stages (EL, EBK and LBK). A further comparison between VIs and AGB at LBK stage, using NDRE values alone, also showed positive significant correlation in both the trials with *r* values of 0.84 and 0.65, respectively (**Figure 5b, Figure 6b**). Across UAV derived canopy metrics at LBK stage, we found significant correlation between CCuav and AGB above *r* = 0.54 (**Figure 5b, Figure 6b)**. Our results clearly indicate that EBK is one of the key phenological stages to predict AGB through remote sensing in cassava. Combining VIs at three phenological stages (EL, EBK and LBK), the trial two showed good AGB relationship with NDRE, NDVI, GNDVI, BNDVI, NDREI, NPCI and GRVI with *r* values of 0.71, 0.62, 0.66, 0.59, 0.64, 0.55, and 0.66, respectively (**Figure 6b, Figure 10a**).

Relationship between UAV derived VIs and below-ground biomass

Measuring root biomass through non-destructive methods over different cassava varieties will help cassava breeders in the efficient selection of cultivars with favorable rooting architectures e.g. root area and harvesting [78]. Thereby, the impact of agronomic research through unique agricultural practices on root bulking can be assessed. Destructive root sampling in cassava requires sampling large populations and trials that are laborious and expensive [8]. Rapid and non-destructive process of estimating below-ground biomass (BGB) across different environments would reduce time, cost and sample size requirements in phenotypic data collection. In this study, we determine the capability of MS aerial imaging to estimate BGB. In both trials, except at DMA stage, all the tested VIs showed positive and significant correlation with fresh BGB at EL and LBK stages (**Figure 5a, Figure 6a**). Our results revealed that the later stage (DMA) of cassava crop life was least correlated, attributing the fact that at the later crop stages (i.e. when the roots are actively accumulating dry matter), cassava canopy tends to senescence.

In both the trials, NDRE, NDVI, GNDVI, BNDVI, NDREI, NPCI, GRVI indices showed significant positive correlations with fresh root biomass with *r* values ranging from 0.18 to 0.72 during the EL to LBK stage, where the highest correlation coefficient (r=0.72) correspond to NDRE at the EL stage at trial two (**Figure 5a, Figure 6a**). On the other hand, canopy metrics (CCuav and CVuav) exhibited highest and stronger correlations with BGB at LBK in trial one with r = 0.70 and r = 0.70, respectively (**Figure 5a)**. Also, we found that the DMA stage showed poor and no significant correlation for some VIs, CHuav and CVuav metrics (**Figure 6a**). In addition, the multi-temporal analysis showed improved correlations with BGB, where we observed that the combination of VIs at [EL+EBK] stages showed highly significant correlation (r = 0.77) for GNDVI (**Figure 6a, Figure 10b**). Generally, from three to five months after planting (MAP), intense development of the photosynthetic apparatus and aerial part of the cassava plants is observed. Consequently, a vigor in this phase causes the greatest enhancement of AGB with consequent reflection in fresh root yield [13]. The relationship between aerial imaging features and BGB obtained from this study are encouraging and it can be an add-on feature for our ongoing Ground penetrating Radar (GPR) research predicting BGB in cassava. Furthermore, all the data produced from above (UAV multispectral) and below ground sensors (GPR) could be merged using high precision Geographic Information System (GIS) to achieve more comprehensive estimation of BGB.

Cassava root yield predictions using ML models

Accurate estimation of crop yield is essential for plant breeders. Yield is a very important harvest trait observation that involves the cumulative effect of weather and management practices throughout the entire growing cycle. [79]. Early detection and crop management associated with yield limitations can help increase productivity [4,23,80]. Crop yield prediction models could aid in early decision-making, optimizing the time required for field evaluation, thus reducing the resources allocated to the research programs [81]. Furthermore, the predicted yield maps could also be used to implement variable rate technology (VRT) systems in spatial databases, thereby accomplishing precise field-level inputs through the entire field [82]. Traditional cassava growth models have certain limitations, such as high input cost required to run the models, the lack of spatial information, or the actual quality of input data [13]. Remote sensing approaches can provide growers with final yield assessments and show variations across the field [79]. In remote sensing, MS imagery can describe crop development for potato tuber yield forecasting, across time and space, in a cost-effective manner [81,82].

To our knowledge, there are no predictive models for cassava root yield using aerial imaging and ML techniques. Therefore, ML technique was explored to provide a means of early prediction of cassava root yield using MS UAV remote sensing on a field scale. A PCA and PCR analysis was used to establish, with which more than 600 predictor variables were retained to train the models. The PCA results showed that the contribution of the first 10 components explains 90% of variance (**Figure 11**) and PCR after a 10 fold cross validations can achieve a R^2 of 0.89. With PCA, the most important component was PC1, explained 55.6% of total variance (**Table 3**). Using the first four components provided by PCA (80% of the total variance) and PCR, SVM, RF, kNN, and ANN models were built to predict BGB using multi-temporal VIs combinations and canopy metrics (**Figure 12**). Among the four developed ML models, the results showed consistent performance with small differences between PCA and PCR techniques ranging from 0 to 9% along the metrics (**Table 4**). PCA was performed little better than PCR in terms of RRMSE and R2 ranging from 20.51% to 22.73% and 0.61 to 0.67, respectively. In this case, the RF model gave the most well-adjusted results, with high R^2 and lowest RMSE, indicating the importance of VIs and canopy metrics to predict BGB by MS sensors. Even though the accuracy of developed models is not very high, considering the laborious cassava phenotyping efforts, CIAT Pheno-i will still be handy for breeders to reduce their time and efforts. This model accuracy can be easily improved by adding other features such as climate, soil, and more timing points.

Conclusions And Future Directions

The use of UAV platforms in rapid acquisition of phenotypic information, such as key phenological stages and vegetation indices as described in this work, have great potential to be used as a selection tool in cassava breeding programs. Automated image analytical framework (CIAT Pheno-i) developed in this study showed promising results and could be applied to other crops than cassava to accelerate germplasm and varietal selection. Machine learning model to predict cassava root yield using MS UAV imagery is encouraging however further validation in diverse sets of germplasm in different environments is necessary. Furthermore, the validation of this ML models in large cassava core collection is currently under progress. In summary, UAVs equipped with MS sensors rapidly monitored canopy metrics, VIs and effectively predicted cassava root yield in a non-destructive and cost effective way. As of now, we are also exploring other ground sensor technologies such as Ground penetrating radar (GPR) to predict cassava root yield more accurately by integrating above and below-ground time series information. Through different innovative remote sensing and image technologies it is highly possible to find out the hidden secrets of below-ground information in cassava which eventually bring higher accuracy in yield prediction.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors agreed to publish this manuscript.

Availability of data and material

The data used in this study is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This work was partially funded by the NSF-BREAD program through Texas A&M University.

Author's contributions

Michael Gomez Selvaraj designed the study and wrote the paper and the other authors read, edited and approved the manuscript. Diego Guzmán and Milton Valencia helped to analyze the UAV images. Michael Gomez Selvaraj and Milton Valencia designed the field experiments and collected the ground-truth data. Manuel Valderrama and Henry Ruiz helped to develop CIAT Pheno-i Image Analysis Platform. Animesh Acharjee guided Manuel Valderrama to develop ML models.

Acknowledgments:

The authors would like to thank Hernan Ceballos, Luis Augusto Becerra, Ishitani Manabu and Joe Tohme, from the International Center for Tropical Agriculture (CIAT) for their support in this research. The authors would like to acknowledge Frank Montenegro, Alejandro Vergara, Cristhian Delgado, Jorge Casas, Sandra Salazar and all the phenomics lab members from CIAT for their support in conducting this study. As well as Angela Fernando, CIAT consultant, for her support in formatting and technical editing. Thanks to CGIAR Big Data and CGIAR Research Program on Roots, Tubers and Bananas (RTB) for their continuous encouragement to develop CIAT Pheno-i software and phenomics platform. Animesh Acharjee acknowledged support from National Institute of Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre (SRMRC). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Medical Research Council or the Department of Health, UK.

Abbreviations

AGB: above-ground biomass

ANN: Artificial neural networks

API: Application program interface

BGB: Below-ground biomass

BNDVI: Blue normalized difference vegetation Index

CCuav: Canopy cover estimate from UAV

CH: Canopy height

CHuav: Canopy height estimate from UAV

CIAT Pheno-i: Automated image analytical framework

CPU: Central processing unit

CSV: Comma separated values

DEM: Digital elevation model

DMA: Dry matter accumulation

EBK: Early bulking

EL: Elongation

ELC: Empirical line calibration

GNDVI: Green normalized difference vegetation index

GCPs: ground control points

GRVI: Green-Red vegetation index

GNSS: Global navigation satellite system

GUI: Graphical user interface

GIS: Geographic information system

GMR: Green minus red

GPR: Ground penetrating radar

GDAL: Geospatial data abstraction library

HTFP: High-throughput field phenotyping

KNN: k-Nearest neighbors

LAI: Leaf area index

LBK: Late bulking

ML: Machine learning

MLP: Multi-Layer perceptron

MS: Multispectral

NIR: near-infrared

NDRE: Normalized difference Red-Edge

NDVI: Normalized difference vegetation index

NDREI: Normalized difference Vegetation Index Red-Edge

NGBDI: Normalized Green-Blue Difference Index

NPCI: Normalized pigment chlorophyll Index

CVuav: Canopy volume estimate from UAV

RTK-GPS: Real-time kinematic global positioning system

RF: Random forest

REST: Representational state transfer

IMU: Inertial measurement unit

RGB: Red, green, blue

RMSE: Root median square error

PCA: Principal component analysis

PCR: Principal component regression

RRMSE: Relative root median square error

SPA: Single page app

SVM: Support Vector Machine

UAV: Unmanned aerial vehicles

VIs: Vegetation indices

VRT: Variable rate technology

References

1. Allem AC, Mendes RA, Salomão AN, Burle ML. The primary gene pool of cassava (Manihot esculenta Crantz subspecies esculenta, Euphorbiaceae). In: Euphytica. Springer Netherlands; 2001. p. 127–32.
2. Oyewole O, Africa SO-J of FT in, 2001 undefined. Effect of length of fermentation on the functional characteristics of fermented cassava'fufu'. ajol.info [Internet]. [cited 2020 Jan 20]; Available from: https://www.ajol.info/index.php/jfta/article/view/19283
3. Pearce F. Cassava comeback. New Sci. 2007 Apr 21;194(2600):38–9.
4. Malik AI, Kongsil P, Nguyen VA, Ou W, Sholihin, Srean P, et al. Cassava breeding and agronomy in Asia - 50 years of history and future directions. Breed Sci (Accept MS). 2020;
5. Lynam J, Byerlee D. Forever Pioneers: CIAT: 50 Years Contributing to a sustainable food future and counting. CIAT [Internet]. Vol. 2009, Cali, Colombia. 2017 [cited 2020 Jan 20]. 156 p. Available from: https://cgspace.cgiar.org/bitstream/handle/10568/89043/CIAT50_FOREVER_PIONEERS.pdf?sequence=3
6. Walter J, Edwards J, Cai J, McDonald G, Miklavcic SJ, Kuchel H. High-Throughput Field Imaging and Basic Image Analysis in a Wheat Breeding Programme. Front Plant Sci [Internet]. 2019 Apr 24 [cited 2020 Jan 21];10. Available from: https://www.frontiersin.org/article/10.3389/fpls.2019.00449/full
7. Qiu Q, Sun N, Bai H, Wang N, Fan Z, Wang Y, et al. Field-based high-throughput phenotyping for maize plant using 3d LIDAR point cloud generated with a "phenomobile." Front Plant Sci. 2019 Apr 16;10.
8. Delgado A, Hays DB, Bruton RK, Ceballos H, Novo A, Boi E, et al. Ground penetrating radar: a case study for estimating root bulking rate in cassava (Manihot esculenta Crantz). Plant Methods [Internet]. 2017 Dec 7 [cited 2020 Jan 20];13(1):65. Available from: http://plantmethods.biomedcentral.com/articles/10.1186/s13007-017-0216-0
9. Ceballos H, Pérez JC, Joaqui Barandica O, Lenis JI, Morante N, Calle F, et al. Cassava Breeding I: The Value of Breeding Value. Front Plant Sci [Internet]. 2016 Aug 29 [cited 2020 Jan 21];7. Available from: http://journal.frontiersin.org/Article/10.3389/fpls.2016.01227/abstract
10. Nduwumuremyi A, Melis R, Shanahan P, Theodore A. Analysis of phenotypic variability for yield and quality traits within a collection of cassava (Manihot esculenta) genotypes. South African J Plant Soil. 2018 May 27;35(3):199–206.
11. Selvaraj MG, Montoya-P ME, Atanbori J, French AP, Pridmore T. A low-cost aeroponic phenotyping system for storage root development: Unravelling the below-ground secrets of cassava (Manihot esculenta). Plant Methods. 2019 Nov 9;15(1).
12. Atanbori J, Montoya-P ME, Selvaraj MG, French AP, Pridmore TP. Convolutional Neural Net-Based Cassava Storage Root Counting Using Real and Synthetic Images. Front Plant Sci [Internet]. 2019 Nov 26 [cited 2020 Jan 21];10. Available from: https://www.frontiersin.org/article/10.3389/fpls.2019.01516/full
13. Vitor AB, Diniz RP, Morgante CV, Antônio RP, de Oliveira EJ. Early prediction models for cassava root yield in different water regimes. F Crop Res. 2019 Jun 1;239:149–58.
14. Zhao C, Zhang Y, Du J, Guo X, Wen W, Gu S, et al. Crop phenomics: Current status and perspectives. Vol. 10, Frontiers in Plant Science. Frontiers Media S.A.; 2019.
15. Haghighattalab A, González Pérez L, Mondal S, Singh D, Schinstock D, Rutkoski J, et al. Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods [Internet]. 2016 [cited 2020 Jan 21];12:35. Available from: www.planner.ardupilot.com
16. Delgado A, Novo A, Hays DB. Data Acquisition Methodologies Utilizing Ground Penetrating Radar for Cassava (Manihot esculenta Crantz) Root Architecture. Geosciences [Internet]. 2019 Apr 15 [cited 2020 Jan 21];9(4):171. Available from: https://www.mdpi.com/2076-3263/9/4/171
17. Quirós Vargas JJ, Zhang C, Smitchger JA, McGee RJ, Sankaran S. Phenotyping of Plant Biomass and Performance Traits Using Remote Sensing Techniques in Pea (Pisum sativum, L.). Sensors [Internet]. 2019 Apr 30 [cited 2020 Jan 21];19(9):2031. Available from: https://www.mdpi.com/1424-8220/19/9/2031
18. Jay S, Baret F, Dutartre D, Malatesta G, Héno S, Comar A, et al. Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. Remote Sens Environ. 2019 Sep 15;231.
19. Furbank RT, Jimenez-Berni JA, George-Jaeggli B, Potgieter AB, Deery DM. Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. New Phytol [Internet]. 2019 Sep 26 [cited 2020 Jan 21];223(4):1714–27. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.15817
20. Hassan MA, Yang M, Rasheed A, Yang G, Reynolds M, Xia X, et al. A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. Plant Sci. 2019 May 1;282:95–103.
21. Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric. 2016 Feb 1;121:57–65.
22. Everingham Y, Sexton J, Skocaj D, Inman-Bamber G. Accurate prediction of sugarcane yield using a random forest algorithm. Agron Sustain Dev. 2016 Jun 1;36(2).
23. Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Vol. 151, Computers and Electronics in Agriculture. Elsevier B.V.; 2018. p. 61–9.
24. Jeffries GR, Griffin TS, Fleisher DH, Naumova EN, Koch M, Wardlow BD. Mapping sub-field maize yields in Nebraska, USA by combining remote sensing imagery, crop simulation models, and machine learning. Precis Agric. 2019;
25. Kamir E, Waldner F, Hochman Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J Photogramm Remote Sens. 2020 Feb 1;160:124–35.

26. Li J, Veeranampalayam-Sivakumar AN, Bhatta M, Garst ND, Stoll H, Stephen Baenziger P, et al. Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. Plant Methods. 2019 Nov 1;15(1).
27. Zhou X, Zheng HB, Xu XQ, He JY, Ge XK, Yao X, et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS J Photogramm Remote Sens [Internet]. 2017 [cited 2020 Jan 21];130:246–55. Available from: http://dx.doi.org/10.1016/j.isprsjprs.2017.05.003
28. Naik HS, Zhang J, Lofquist A, Assefa T, Sarkar S, Ackerman D, et al. A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. Plant Methods [Internet]. 2017 Dec 8 [cited 2020 Jan 21];13(1):23. Available from: http://plantmethods.biomedcentral.com/articles/10.1186/s13007-017-0173-7
29. Cen H, Wan L, Zhu J, Li Y, Li X, Zhu Y, et al. Dynamic monitoring of biomass of rice under different nitrogen treatments using a lightweight UAV with dual image-frame snapshot cameras. Plant Methods. 2019 Mar 27;15(1).
30. Han L, Yang G, Dai H, Xu B, Yang H, Feng H, et al. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods. 2019 Feb 4;15(1).
31. Zahid A, Abbas HT, Ren A, Zoha A, Heidari H, Shah SA, et al. Machine learning driven non - invasive approach of water content estimation in living plant leaves using terahertz waves. Plant Methods [Internet]. 2019 [cited 2020 Jan 21];1–13. Available from: https://doi.org/10.1186/s13007-019-0522-9
32. Roitsch T, Cabrera-Bosquet L, Fournier A, Ghamkhar K, Jiménez-Berni J, Pinto F, et al. Review: New sensors and data-driven approaches—A path to next generation phenomics. Vol. 282, Plant Science. Elsevier Ireland Ltd; 2019. p. 2–10.
33. Fiorani F, Schurr U. Future Scenarios for Plant Phenotyping. Annu Rev Plant Biol [Internet]. 2013 Apr 29 [cited 2020 Jan 21];64(1):267–91. Available from: http://www.annualreviews.org/doi/10.1146/annurev-arplant-050312-120137
34. Czedik-Eysenberg A, Seitner S, Güldener U, Koemeda S, Jez J, Colombini M, et al. The 'PhenoBox', a flexible, automated, open-source plant phenotyping solution. New Phytol [Internet]. 2018 Jul [cited 2020 Jan 21];219(2):808–23. Available from: http://doi.wiley.com/10.1111/nph.15129
35. Xue J, Su B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. hindawi.com [Internet]. 2017 [cited 2020 Apr 2]; Available from: https://doi.org/10.1155/2017/1353691
36. Fukuda WMG, Guevara CL, Kawuki R, Ferguson ME. Selected morphological and agronomic descriptors for the characterization of cassava [Internet]. [cited 2020 Jan 21]. Available from: www.iita.org
37. LI-2200C Plant Canopy Analyzer [Internet]. [cited 2020 Mar 31]. Available from: https://www.licor.com/env/products/leaf_area/LAI-2200C/
38. Barnes E, Clarke T, Richards S, Colaizzi P, Haberland J, Kostrzewski M, et al. COINCIDENT DETECTION OF CROP WATER STRESS, NITROGEN STATUS AND CANOPY DENSITY USING GROUND-BASED MULTISPECTRAL DATA. 2000.
39. Tucker CJ. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Vol. 8, REMOTE SENSING OF ENVIRONMENT. 1979.
40. Moges SM, Raun WR, Mullen RW, Freeman KW, Johnson G V., Solie JB. Evaluation of Green, Red, and Near Infrared Bands for Predicting Winter Wheat Biomass, Nitrogen Uptake, and Final Grain Yield. J Plant Nutr [Internet]. 2005 Jan 2 [cited 2020 Feb 10];27(8):1431–41. Available from: http://www.tandfonline.com/doi/abs/10.1081/PLN-200025858
41. WANG F, HUANG J, TANG Y, WANG X. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. Rice Sci. 2007 Sep 1;14(3):195–203.
42. Gitelson A, Merzlyak MN. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J Photochem Photobiol B Biol. 1994 Mar 1;22(3):247–52.
43. Peñuelas J, Gamon JA, Fredeen AL, Merino J, Field CB. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sens Environ. 1994 May 1;48(2):135–46.
44. Wang Y, Wang D, Zhang G, Wang J. Estimating nitrogen status of rice using the image segmentation of G-R thresholding method. F Crop Res. 2013 Aug 1;149:33–9.
45. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9(1):62–6.
46. Kuhn M, Johnson K. Applied predictive modeling. Applied Predictive Modeling. Springer New York; 2013. 1–600 p.
47. Jolliffe IT. A Note on the Use of Principal Components in Regression. Appl Stat [Internet]. 1982 Nov 1 [cited 2020 Apr 3];31(3):300. Available from: https://www.jstor.org/stable/10.2307/2348005?origin=crossref
48. Gago J, Fernie AR, Nikoloski Z, Tohge T, Martorell S, Escalona JM, et al. Integrative field scale phenotyping for investigating metabolic components of water stress within a vineyard. Plant Methods. 2017 Oct 30;13(1).
49. Nagasubramanian K, Jones S, Sarkar S, Singh AK, Singh A, Ganapathysubramanian B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. Plant Methods. 2018 Oct 3;14(1).
50. Schirrmann M, Giebel A, Gleiniger F, Pflanz M, Lentschke J, Dammer K-H. Monitoring Agronomic Parameters of Winter Wheat Crops with Low-Cost UAV Imagery. Remote Sens [Internet]. 2016 Aug 27 [cited 2020 Jan 21];8(9):706. Available from: http://www.mdpi.com/2072-4292/8/9/706
51. Albon C. Machine learning with python cookbook: Practical solutions from preprocessing to deep learning [Internet]. 2018 [cited 2020 Feb 12]. Available from: https://books.google.com/books?hl=en&lr=&id=klhQDwAAQBAJ&oi=fnd&pg=PT80&dq=Machine+learning+with+python+cookbook:+Practical+solutions+from+preprocessing+to+deep+lea
52. Ho TK. Random decision forests. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. IEEE Computer Society; 1995. p. 278–82.
53. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995 Sep;20(3):273–97.

54. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;46(3):175–85.
55. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines.
56. Minervini M, Giuffrida M V., Perata P, Tsaftaris SA. Phenotiki: an open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. Plant J [Internet]. 2017 Apr [cited 2020 Jan 21];90(1):204–16. Available from: http://doi.wiley.com/10.1111/tpj.13472
57. Gago J, Douthe C, Coopman RE, Gallego PP, Ribas-Carbo M, Flexas J, et al. UAVs challenge to assess water stress for sustainable agriculture. Vol. 153, Agricultural Water Management. Elsevier; 2015. p. 9–19.
58. Sankaran S, Khot LR, Espinoza CZ, Jarolmasjed S, Sathuvalli VR, Vandemark GJ, et al. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. Vol. 70, European Journal of Agronomy. Elsevier; 2015. p. 112–23.
59. van der Meij B, Kooistra L, Suomalainen J, Barel JM, De Deyn GB. Remote sensing of plant trait responses to field-based plant–soil feedback using UAV-based optical sensors. Biogeosciences [Internet]. 2017 Feb 15 [cited 2020 Jan 22];14(3):733–49. Available from: https://www.biogeosciences.net/14/733/2017/
60. Araus JL, Cairns JE. Field high-throughput phenotyping: The new crop breeding frontier. Vol. 19, Trends in Plant Science. 2014. p. 52–61.
61. Zaman-Allah M, Vergara O, Araus JL, Tarekegne A, Magorokosho C, Zarco-Tejada PJ, et al. Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. Plant Methods. 2015 Jul 1;11(1).
62. Aasen H, Burkart A, Bolten A, Bareth G. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS J Photogramm Remote Sens. 2015 Oct 1;108:245–59.
63. Domingues Franceschini M, Bartholomeus H, van Apeldoorn D, Suomalainen J, Kooistra L. Intercomparison of Unmanned Aerial Vehicle and Ground-Based Narrow Band Spectrometers Applied to Crop Trait Monitoring in Organic Potato Production. Sensors [Internet]. 2017 Jun 18 [cited 2020 Jan 22];17(6):1428. Available from: http://www.mdpi.com/1424-8220/17/6/1428
64. Jin X, Liu S, Baret F, Hemerlé M, Comar A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens Environ. 2017 Sep 1;198:105–14.
65. Duan B, Fang S, Zhu R, Wu X, Wang S, Gong Y, et al. Remote estimation of rice yield with unmanned aerial vehicle (uav) data and spectral mixture analysis. Front Plant Sci. 2019 Mar 7;10.
66. Singh D, Wang X, Kumar U, Gao L, Noor M, Imtiaz M, et al. High-Throughput Phenotyping Enabled Genetic Dissection of Crop Lodging in Wheat. Front Plant Sci [Internet]. 2019 Apr 3 [cited 2020 Jan 22];10. Available from: https://www.frontiersin.org/article/10.3389/fpls.2019.00394/full
67. Shafian S, Rajan N, Schnell R, Bagavathiannan M, Valasek J, Shi Y, et al. Unmanned aerial systems-based remote sensing for monitoring sorghum growth and development. PLoS One. 2018 May 1;13(5).
68. Okogbenin E, Setter TL, Ferguson M, Mutegi R, Ceballos H, Olasanmi B, et al. Phenotypic approaches to drought in cassava: Review. Front Physiol. 2013;4 MAY.
69. Mulualem T, Ayenew B. Cassava (Mannihot esculenta Cranz) Varieties and Harvesting Stages Influenced Yield and Yield Related Components [Internet]. Vol. 2, Journal of Natural Sciences Research www.iiste.org ISSN. Online; 2012 [cited 2020 Jan 22]. Available from: www.iiste.org
70. Xu R, Liid C, Paterson AH. Multispectral imaging and unmanned aerial systems for cotton plant phenotyping. 2019 [cited 2020 Jan 22]; Available from: https://doi.org/10.1371/journal.pone.0205083
71. Verma N, Lamb D, Reid N, Wilson B. Comparison of Canopy Volume Measurements of Scattered Eucalypt Farm Trees Derived from High Spatial Resolution Imagery and LiDAR. Remote Sens [Internet]. 2016 May 5 [cited 2020 Jan 22];8(5):388. Available from: http://www.mdpi.com/2072-4292/8/5/388
72. Ballesteros R, Ortega JF, Hernandez D, Moreno MA. Onion biomass monitoring using UAV-based RGB imaging. Precis Agric. 2018 Oct 1;19(5):840–57.
73. Cock J. Cassava: New Potential for a Neglected Crop. Westview Press Inc, Color USA. 1985 Jan 1;
74. Ekanayake IJ, Bank W. Production and delivery systems of good quality cassava planting materials in sub-Saharan Africa Zambia Agricultural Development Support Project View project Potato Research in the Warm Tropics View project [Internet]. 2018 [cited 2020 Jan 22]. Available from: https://www.researchgate.net/publication/290449258
75. Ramirez-Garcia J, Almendros P, Quemada M. Ground cover and leaf area index relationship in a grass, legume and crucifer crop. Plant, Soil Environ. 2012;58(8):385–90.
76. Campillo C, Garcia M, Daza C, Hortscience MP-, 2010 undefined. Study of a non-destructive method for estimating the leaf area index in vegetable crops using digital images. journals.ashs.org [Internet]. [cited 2020 Jan 22]; Available from: https://journals.ashs.org/hortsci/view/journals/hortsci/45/10/article-p1459.xml
77. Adeel Hassan M, Yang M, Rasheed A, Jin X, Xia X, Xiao Y, et al. Time-Series Multispectral Indices from Unmanned Aerial Vehicle Imagery Reveal Senescence Rate in Bread Wheat. mdpi.com [Internet]. 2018 [cited 2020 Feb 11]; Available from: www.mdpi.com/journal/remotesensing
78. El-Sharkawy MA. Prospects of photosynthetic research for increasing agricultural productivity, with emphasis on the tropical C4 Amaranthus and the cassava C3-C4 crops. Vol. 54, Photosynthetica. Kluwer Academic Publishers; 2016. p. 161–84.
79. Pinter PJ, Hatfield JL, Schepers JS, Barnes EM, Moran MS, Daughtry CST, et al. Remote sensing for crop management. Vol. 69, Photogrammetric Engineering and Remote Sensing. 2003. p. 647–64.
80. Selvaraj MG, Vergara A, Ruiz H, Safari N, Elayabalan S, Ocimati W, et al. AI-powered banana diseases and pest detection. Plant Methods. 2019 Aug 12;15(1):1–11.
81. Gómez, Salvador, Sanz, Casanova. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. Remote Sens [Internet]. 2019 Jul 24 [cited 2020 Jan 22];11(15):1745. Available from: https://www.mdpi.com/2072-4292/11/15/1745

82. Al-Gaadi K, Hassaballa A, Tola E, One AK-P, 2016 U. Prediction of potato crop yield using precision agriculture techniques. ncbi.nlm.nih.gov [Internet]. 2016 [cited 2020 Jan 22]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017787/

Supplementary File Legends

Additional File 1. Video of Pheno-i Image Analysis Platform (https://youtu.be/hnq_ydC1-rw)

Supplementary Figures Legends

Figure S1. Agisoft Metashape automated orthomosaic building pipeline.

Figure S2. CIAT Pheno-i Front-end overview.

Figure S3. Pheno-i image analysis platform design. Back-end: Developed in Python 3 and Flask as a Web Service. Front-end: Developed using React as a single page app, it implements leaflet.js to render maps.

Figure S4. Schematic representation of database implemented in CIAT Pheno-i.

Figure S5. Comparison of time series data between manual and automatic orthomosaic generation with MS and RGB sensors. (**a)** Multispectral trial in manual mode. (**b)** Multispectral trial in auto mode. (**c)** RGB trial in manual mode. (**d)** RGB trial in auto mode. *M1-M8 Manual orthomosaics. *A1-A8 Automatic orthomosaics.

Supplementary Tables Legends

Table S1. Cassava morphological and agronomic descriptors.

Table S2. Phenological stage information (Months) of genotypes listed in this study.

Table S3. List of hardware and software used in this study.

Tables

Table 1. Field experimental conditions and images acquisition.

Trial conditions and images acquisition	Trial one (Dec. 2016 - Nov. 2017)	Trial two (Dec. 2017 - Dec. 2018)
Irrigation	Surface irrigation	Drip irrigation
Soil Type	Clay loam	Clay loam
Experimental Design	Split plot Design	Randomized Complete Block Design
No. of Replication	3	4
Average annual precipitation (mm)	1435.10	1026.50
Average annual temperature (°C)	24.00	23.33
Total solar radiation (W/m)	207.8	222.39
Average annual relative humidity (%)	81.30	78.90
Image acquisition Stages	EL, LBK and DMA:	EL, EBK, LBK and DMA

The following definitions are related to cassava storage root development phases: Elongation **(EL)** stage is the initial growth phase of active fibrous root development. Early bulking (**EBK)** is the root differentiation (from fibrous and storage roots) phase, the beginning of storage root bulking and accumulation of assimilated reserves in the storage roots. Late bulking (**LBK**) stage is the rapid expansion and bulking of storage roots. Dry matter accumulation **(DMA)** stage is the starch accumulation in the storage roots.

Table 2. Summary of vegetation indices used in this study. Camera channels B: blue, G: green, R: red, RE: red-edge, and NIR: near-infrared.

Vegetation Index	Acronym	Formula	References
Normalized Difference Red-Edge	NDRE	(NIR-Rededge)/(NIR+Rededge)	[38]
Normalized Difference Vegetation Index	NDVI	(NIR-Red)(NIR+Red)	[39]
Green Normalized Difference Vegetation Index	GNDVI	(NIR-Green)/(NIR+Green)	[40]
Blue Normalized Difference Vegetation Index	BNDVI	(NIR-Blue)/(NIR+Blue)	[41]
Normalized Difference Vegetation Index Red-Edge	NDREI	(Rededge-Red)/(Rededge+Red)	[42]
Normalized Pigment Chlorophyll Index	NPCI	(Rededge-Blue)/(Rededge+Blue)	[43]
Green-Red Vegetation Index	GRVI	(Green-Red)/(Green+Red)	[17]
Normalized Green-Blue Difference Index	NGBDI	(Green-Blue)/(Green+Blue)	[40]

Table 3. Total variance explained by component.

Component	Total variance
PC1	0.556
PC2	0.145
PC3	0.061
PC4	0.038
PC5	0.027
PC6	0.018
PC7	0.014
PC8	0.014
PC9	0.011
PC10	0.009

Table 4: Root yield ML Model comparison.

		PCA			PCR			PCA vs PCR Difference (%)		
	ML Method	R^2	RMSE	RRMSE	R^2	RMSE	RRMSE	R^2	RMSE	RRMSE
Validation	RF	0.93	443.07	9.88%	0.94	449.09	9.19%	1.07%	1.35%	7.24%
	SVM	0.67	888.23	20.91%	0.63	947.45	22.10%	6.15%	6.45%	5.53%
	kNN	0.64	930.57	21.94%	0.64	953.3	22.01%	0.00%	2.41%	0.32%
	ANN	0.69	893.16	20.26%	0.7	910.21	21.12%	1.44%	1.89%	4.16%
Test	RF	0.66	889.28	20.52%	0.64	891.86	21.12%	3.08%	0.29%	2.88%
	SVM	0.64	916.77	21.23%	0.64	874.13	21.14%	0.00%	4.76%	0.42%
	kNN	0.67	899.8	20.10%	0.67	879.09	20.20%	0.00%	2.33%	0.50%
	ANN	0.61	1021.81	22.73%	0.61	1120.04	22.61%	0.00%	9.17%	0.53%

Figures

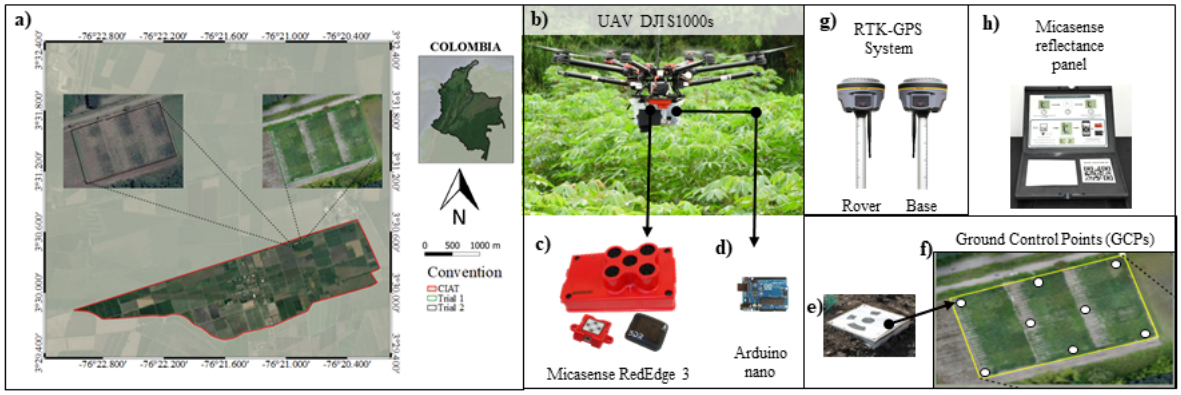

Figure 1.

Figure 1

Field trial site and remote sensing platform. (a) Trial one and two were conducted at the International Center for Tropical Agriculture (CIAT). (b) Unmanned aerial vehicle (UAV), DJI S1000s. (c) Multispectral camera, Micasense RedEdge 3. (d) Arduino nano. (e) Ground Control Point (GCPs). (f) GCPs installed in trial

one. (g) RTK-GPS.

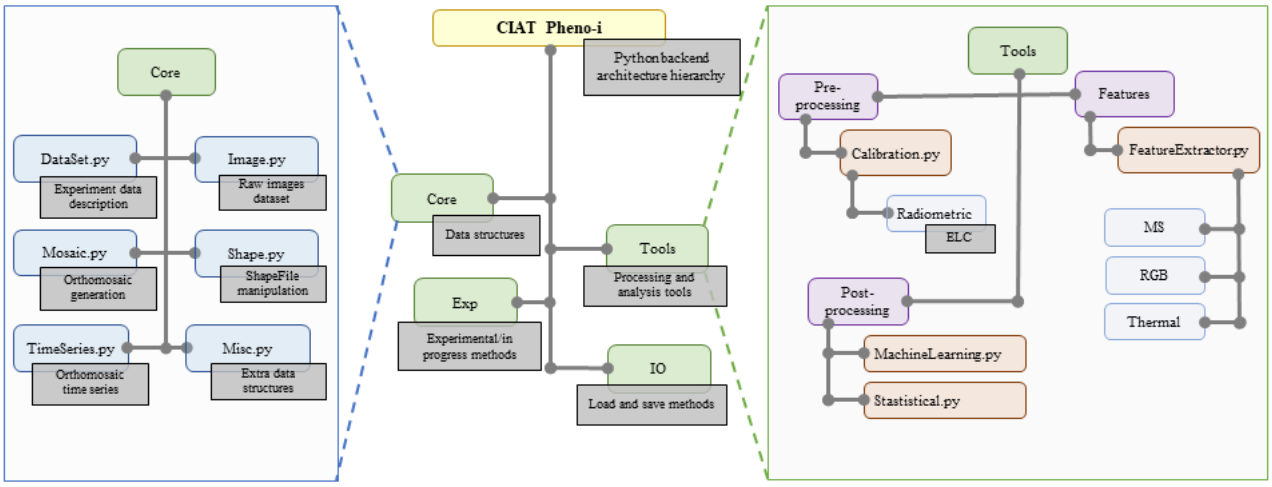

Figure 2.

Figure 2

CIAT Pheno-i data processing back-end Python architecture.

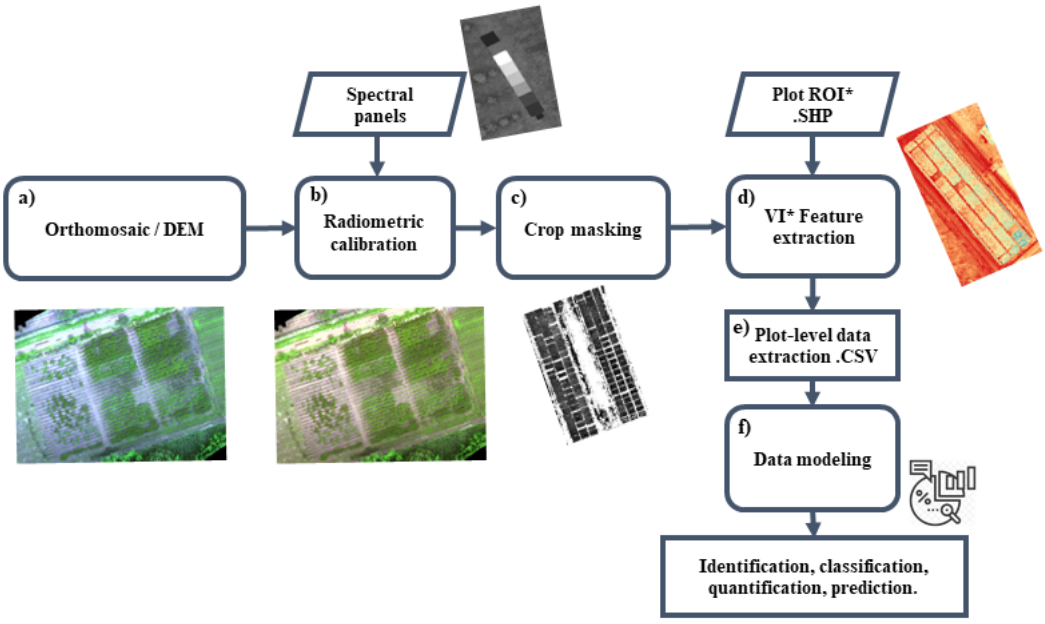

Figure 3.

Figure 3

CIAT Pheno-i workflow: Applying image processing for plot level data generation and use it on identification, classification, quantification and prediction.

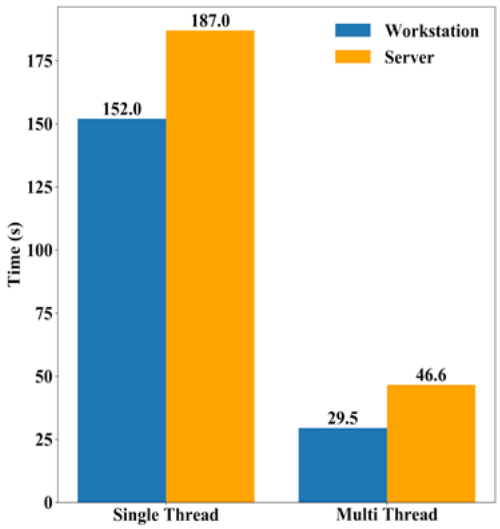

Figure 4.

Figure 4

CIAT Pheno-i back-end performance. Multispectral 5 band image analysis average time of 50 runs with 3870x3739 pixels. Workstation processing time compared against server processing time.

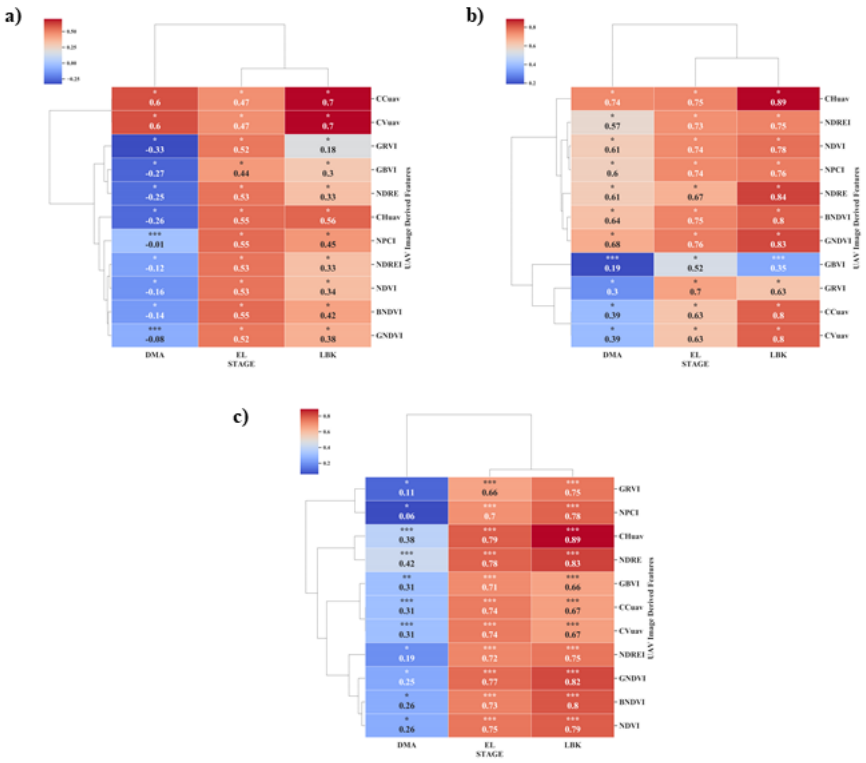

Figure 5.

Figure 5

Pearson correlation analysis between remote sensing features versus shoot and root biomass at different cassava phenological stages under surface irrigation management during the trial one. (a) BGB. (b) AGB. (c) PH. P<0.05: *, P<0.01: **, P<0.005: ***

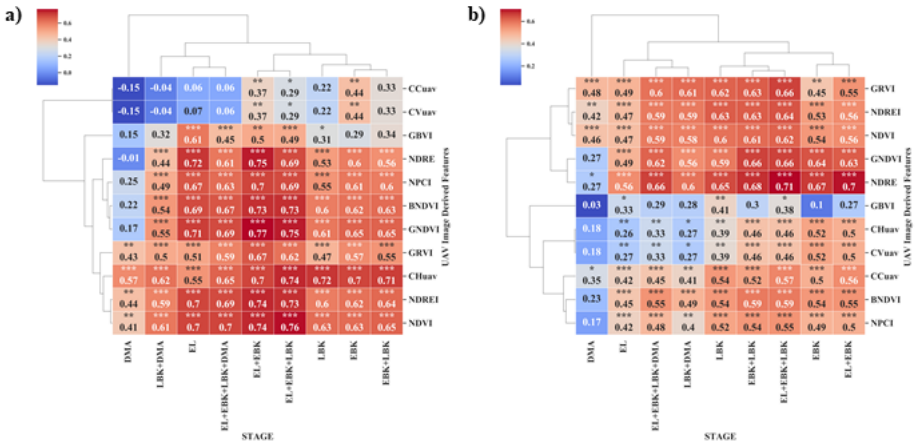

Figure 6.

Figure 6

Pearson correlation analysis between remote sensing features versus shoot and root traits at different cassava phenological stages under drip irrigation management during the trial two. (a) BGB. (b) AGB. P<0.05: *, P<0.01: **, P<0.005: ***

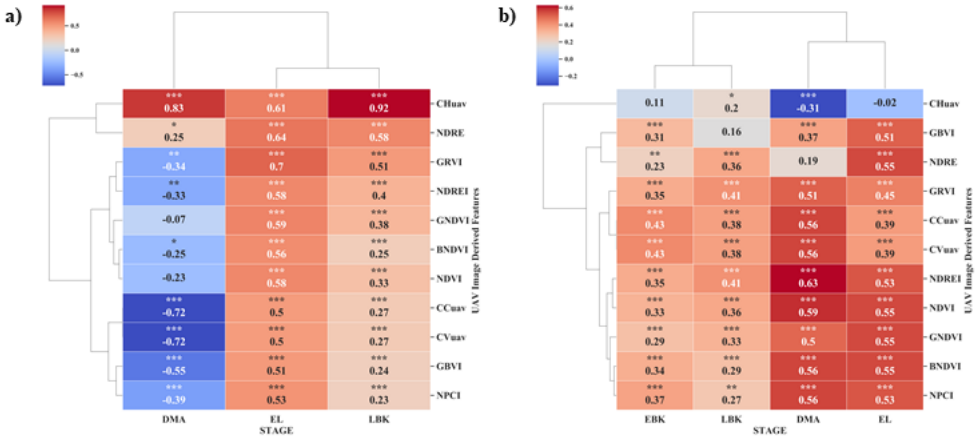

Figure 7.

Figure 7

Pearson correlation analysis between remote sensing features versus canopy metric traits at different cassava phenological stages under drip irrigation management during the trial two. (a) CH. (b) LAI. P<0.05: *, P<0.01: **, P<0.005: ***

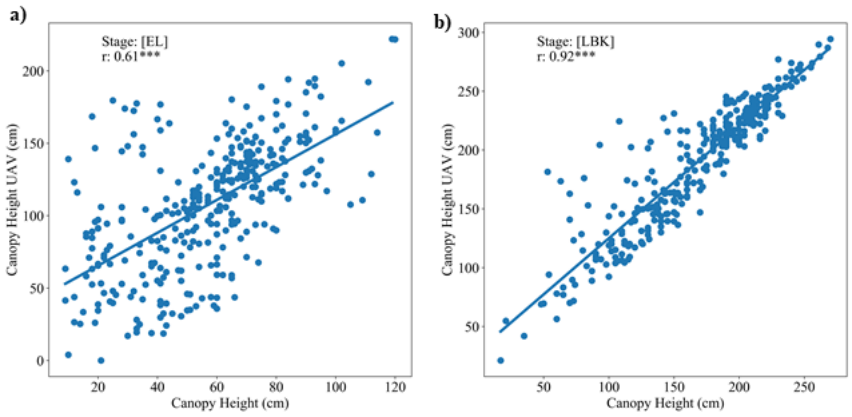

Figure 8.

Figure 8

Comparison of canopy height UAV versus canopy height of cassava at EL (Elongation) and LBK (Late Bulking) during trial two.

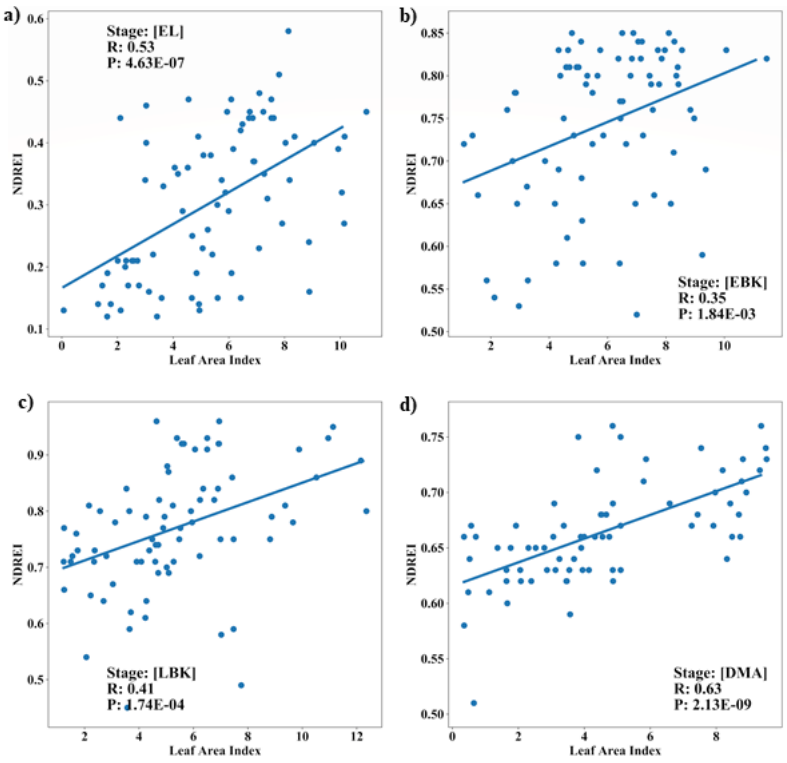

Figure 9.

Figure 9

Comparison of Normalized Difference Vegetation Index Red-Edge NDREI versus Leaf Area Index (LAI) at EL (Elongation), EBK (Early Bulking), LBK (Late Bulking), and DMA (Dry Matter Accumulation) of cassava during the trial two.

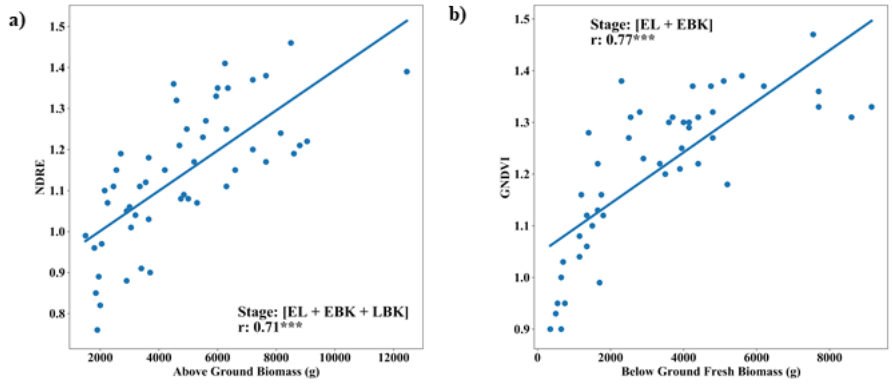

Figure 10.

Figure 10

Relationship between fresh above-ground biomass (AGB) and fresh below ground biomass (BGB) of cassava with multi-temporal VIs (Normalized Difference Red-Edge, NDRE and Green Normalized Difference Vegetation Index GNDVI) during trial two. . (a) Ground truth AGB versus multi-temporal NDRE index at EL (Elongation), EBK (Early Bulking), and LBK (Late Bulking stage). (b) Ground truth BGB and multi-temporal GNDVI index at EL (Elongation) and LBK (Late Bulking) stage.

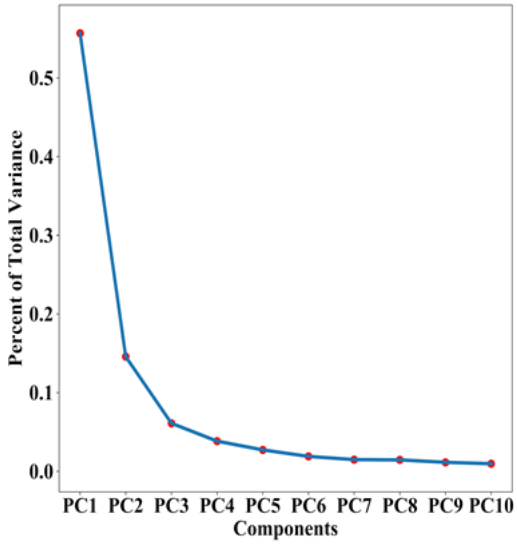

Figure 11.

Figure 11

PCA scree plot of the percent of aforementioned variance during trial two.

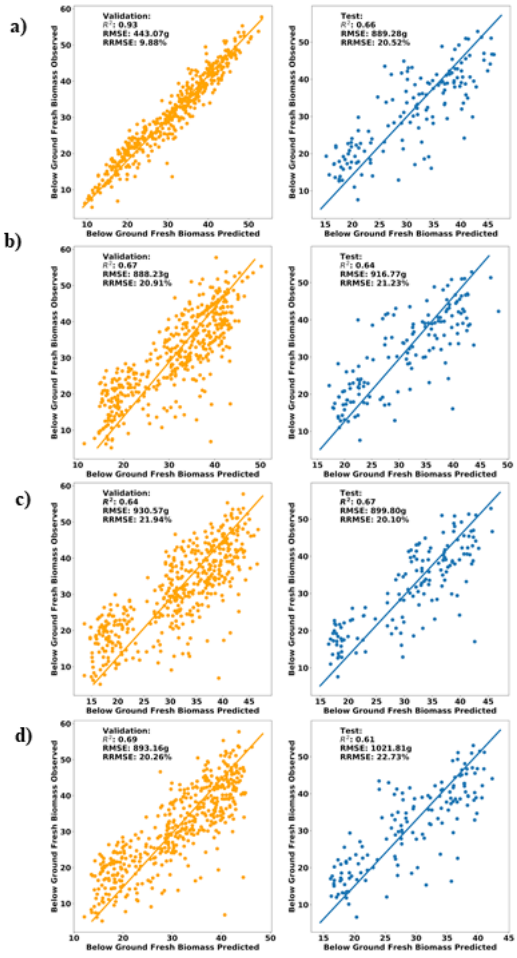

Figure 12.

Figure 12

Plots based on regression methods, validation dataset on the left and test dataset on the right. (a) RF parameters (max_features:4, trees: 100). (b) SVM parameter (C:2.1, kernel:"rbf"). (c) kNN parameters (algorithm: ball_tree, K: 38, weights: uniform). (d) ANN.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pptx
- SupplementaryTables.docx
- MachinelearningMLmodeldevelopment.docx