Computers and Electronics in Agriculture 232 (2025) 110096

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Factors affecting deep learning model performance in citizen science–based image data collection for agriculture: A case study on coffee crops

Juan C. Rivera-Palacio [a,b,c], Christian Bunn [b], Masahiro Ryo [a,c,*]

[a] Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, Müncheberg 15374 Germany
[b] The International Center for Tropical Agriculture (CIAT), Km 17 Recta Cali-Palmira, AA6713 763537 Cali Colombia
[c] Brandenburg University of Technology Cottbus-Senftenberg, Platz Der Deutschen Einheit 1 Konrad-Wachsmann-Allee 13 Cottbus 03046 Germany

ARTICLE INFO

Keywords:
Deep learning
Bias
Citizen science
Smartphone imaging
Coffee
Prediction

ABSTRACT

Citizen science is an effective approach for collecting extensive data scalable for deep learning, although data quality is debatable. However, few studies have determined the factors associated with data collection that affect model performance and potential sampling bias. This study aims to identify the factors that significantly influence the performance of a deep learning object detection model in agricultural prediction tasks. To do so, we analyzed errors in a You Only Look Once (YOLO v8) model trained for counting the number of coffee cherries in mobile pictures. The model was trained with 436 images taken in Colombia and Peru collected by local farmers as a citizen science approach. We analyzed the prediction errors of the model using 637 additional pictures. We then applied a linear mixed model (LMM) and a decision tree machine learning model to regress the model's error against predictor variables related to the following categories: photographer influence, geographic location, mobile phone characteristics, picture characteristics, and coffee varieties. Our results show the strong influence of photographer identity and adherence (whether the image collection protocol was followed or not) on model prediction error. Following the protocol can increase model performance from an R^2 of 0.48 to 0.73. Additionally, model performance varied significantly depending on photographer identity, with R^2 ranging from 0.45 to 0.93. In contrast, factors such as mobile phone characteristics (e.g., frontal camera resolution, flash type, and screen size), using the screen behind the branch to obscure other cherries, coffee varieties, and geographic location did not significantly affect prediction error. These findings demonstrate that data quality in citizen science–based data collection for enhancing model prediction can be achieved through straightforward and comprehensive protocols, customized volunteer training, and regular feedback from experts. Such measures collectively support the robust application of deep learning models in agriculture. Furthermore, this study demonstrated that any mobile device with a camera can contribute to citizen science initiatives, underscoring the potential and scalability of this approach in agricultural research.

1. Introduction

Collecting a large number and covering good-quality images to train a deep learning algorithm is a common challenge in computer vision tasks. Capturing RGB images using smartphone cameras is one of the most promising approaches regarding logistic cost and scalability for large-scale data collection in natural science (Ryan et al., 2018). In particular, a citizen science–based data collection approach can be effective in countries with emerging and developing economies (Baig and Straquadine, 2014; Ryan et al., 2018), as investigators do not need to purchase measurement devices. There are more than five billion unique mobile subscribers, equivalent to 76 % of the world's population (GSMA Intelligence, 2024). Therefore, a mobile phone–based data collection approach makes it possible to cooperate with a large number of local citizens for scalable image data collection and to cover a wide range of conditions (Barbedo, 2018; GSMA Intelligence, 2024; Rivera-Palacio et al., 2024; Ryo et al., 2023).

In agriculture, for example, nearly 60 %–70 % of the variation in the crop yields of rice and coffee can be predicted using the combination of a smartphone camera and deep neural networks across countries (Rivera-Palacio et al., 2024; Tanaka et al., 2023). This citizen science approach of local contributors can drastically scale up the extent and speed of data

* Corresponding author.
E-mail addresses: juancamilo.rivera@zalf.de (J.C. Rivera-Palacio), c.bunn@cgiar.org (C. Bunn), masahiro.ryo@zalf.de (M. Ryo).

https://doi.org/10.1016/j.compag.2025.110096
Received 15 August 2024; Received in revised form 4 February 2025; Accepted 5 February 2025
Available online 13 February 2025

collection while maintaining high model performance. Another advantage of this approach is its suitability for crop types that cannot be monitored from a vertical point of view using unmanned aerial vehicles, aircraft, and satellites (e.g., cherries under leaves and tree canopies in agroforestry settings).

On the other hand, a critical downside of a citizen science approach is uncertainty in the quality of the dataset collected by various contributors. Operators can cause differences in the way they hold the camera, frame the region of interest, and maintain hand steadiness, even when they follow a uniform protocol to capture images (Barbedo, 2018). Cameras have a variety of parameters that can be manually or automatically set to produce images (Barbedo, 2022). The background screen behind the target objects can also affect model performance. However, few studies have investigated the relative influence of these factors on model performance.

This study aims to identify and quantify the factors that significantly influence the performance of the object detection algorithm in agriculture prediction tasks using image data collected through citizen science. More specifically, we analyze the sources of the error of the developed deep learning model, YOLO v8, that identifies and counts the number of coffee cherries on a tree branch in a picture in our previous study Rivera-Palacio, et al (2024). The previous study achieved high performance, attaining an R^2 of 0.72 at an unprecedented geographic scale by training an object detection model YOLO v8 in collaboration with farmers in Peru and Colombia. We assume that the identification process is subject to the combined influence of various factors about the contributors' conditions, including mobile phone characteristics, geographic location, coffee varieties, photographer influence, and picture characteristics. Furthermore, the inexperience of farmers in capturing images alongside the complex agroforestry coffee landscape adds additional layers of challenge to the process.

2. Materials and methods

2.1. Data and protocol for taking pictures

As the full details of the method are listed in the work of Rivera-Palacio, et al. (2024), we provide the essential details here to satisfy the reproducibility requirements of this study.

From March to November 2022, images of coffee cherries on tree branches were collected in the departments of Cauca and Quindio, in southwest Colombia (Genova, Cajibio, El Tambo, Morales, Piendamo, and Popayan; 2°–4°N, 75°–77°W), and in the northern and central parts of the coffee districts in Peru (Chinchaque, Chirinos, Cañariz, Lalaquiz, Pongoa, and San José Lourdes; latitude-longitude of 5°–6°S, 6°–7°W). In October 2022, samples were collected in the municipality of Cajibio, Cauca Department, to test angles for taking pictures at the branch. The total database consists of 8,904 mobile pictures of coffee trees collected by 977 farmers with 53 survey personnel (hereafter, enumerators) that were used in a previous study by Rivera-Palacio et al. (2024). Fig. 1 shows an example of the collected images. For this study, we selected a random sample from the dataset, consisting of 13 enumerators, referred to as photographers, and 49 randomly pictures per enumerator (637 pictures in total). The protocol for taking pictures of coffee branches followed the approach outlined by Rivera-Palacio et al. (2024). Each enumerator selected three branches per tree and took one photo per branch. Each photo was taken during daylight to capture as many cherries as possible and not capture cherries from other branches or trees. Sometimes, a screen was used to conceal other branches in the background.

2.2. Object detection model

We used an object detection model developed by Rivera-Palacio et al. (2024) to detect coffee cherries. This model employs the You Only Look Once version 8 (YOLO v8), which was provided by Ultralytics in 2023

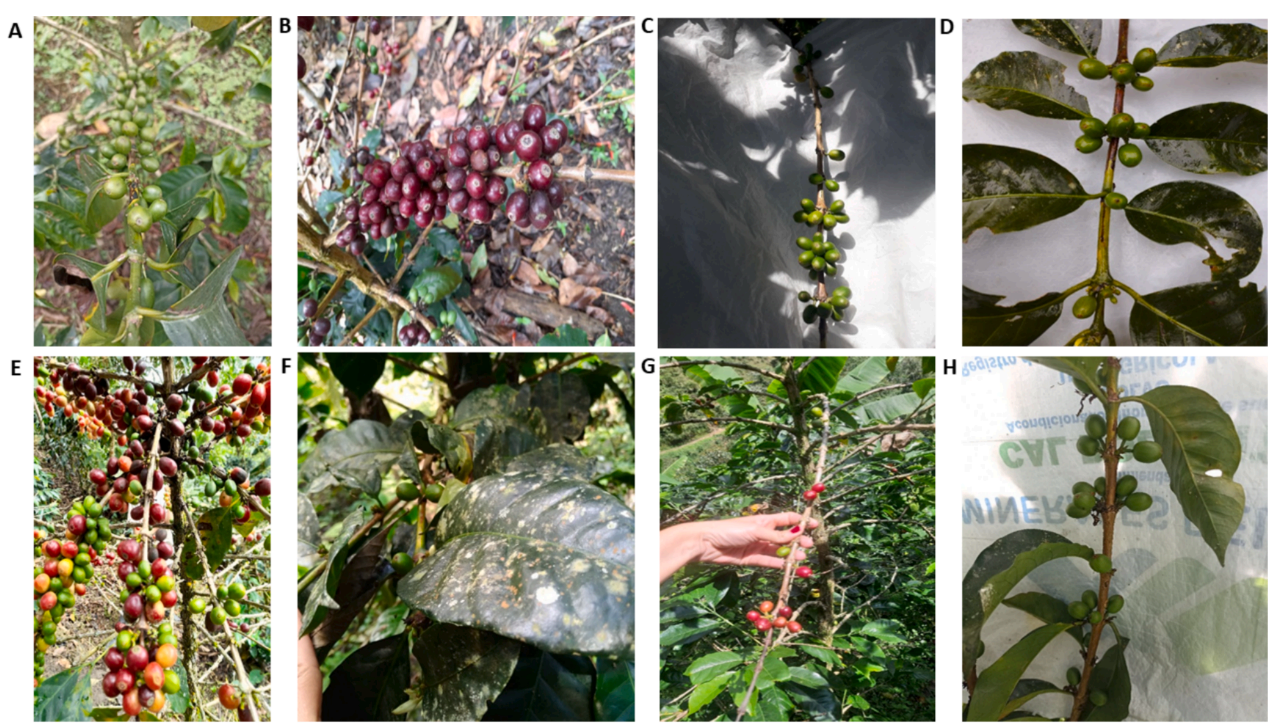

Fig. 1. Representative images of coffee cherries taken based on a citizen science approach. A) a picture taken of the front of the branch; B) a picture taken from above the branch; C) a picture with a white screen and shadows due to sunlight; D) a picture with a white screen behind the branch; E) a picture without a screen, with other branches behind; F) a picture of large leaves that hide the cherries; G) external objects in the pictures, such as a red nail, which resembles a cherry; H) a screen with different colors behind the branch. Additional examples of pictures coffee cherries can be found in Fig. A1 of Appendix 1.

(Jocher et al., 2023) model, originally trained with the Common Objects in Context (COCO) dataset (Lin et al., 2014), underwent further fine-tuning with 436 smartphone images of coffee branches (each image showed up to 120 coffee cherries). The dataset consisted of the following objects in the 436 images: 35,247 green coffee cherries, 342 red coffee cherries, and 105 black coffee cherries. To enhance the dataset's quality, data augmentation techniques, including brightness adjustment and geometric transformation, were applied to the images. The dataset was divided into training (80 %, n = 346), validation (10 %, n = 43), and test (10 %, n = 43) sets, where n is the number of mobile pictures. For a comprehensive outline of the performance metrics and hyperparameters used, see Rivera-Palacio et al. (2024). The model was used in this study to detect coffee cherries on an additional 637 randomly selected smartphone images of coffee branches from 13 enumerators. Descriptions of these pictures are shown in Table A1 **of Appendix 1.** The development process took place on a Windows 10 platform using Python 3.7.0 and PyTorch 1.7, running on 4 GPU Nodes (in particular, two Nvidia Tesla V100 GPUs).

We set up to scenarios to preliminarily evaluate the performance of YOLO v8 in detecting coffee cherries in mobile phone images. In the first scenario, we compared the total number of coffee cherries detected by YOLO v8 with the total number visually counted in each mobile phone image, achieving an R^2 of 0.80. In the second scenario, we compared the manual count of coffee cherries per branch, as recorded by field enumerators, against the total detected by YOLO v8, which dropped significantly, yielding an R^2 of 0.65 (Fig. 2). Ideally, the results of the first and second scenarios should be equal; however, a difference in R^2 of 0.15 was observed. In the the first scenario suggested that YOLO v8 demonstrated good accuracy in detecting cherries in mobile images. However, its performance was less accurate in the second scenario, that is, when compared to the actual number of cherries per branch based on a manual count. This preliminary finding indicates that detection accuracy is influenced not only by the characteristics of YOLO v8 but also by other factors. Therefore, we define the YOLO v8 error as $e(j)$ per each mobile picture j (the red line vertical in Fig. 2):

$$e(j) = y_j - \hat{y}_j \tag{1}$$

Where $e(j)$ is the error for the jth picture, y_j = the total of coffee cherries detected per YOLO v8 $\hat{y}_j$ = the number of coffee cherries manually counted on a branch, and j = 1,.......637 is the index for a mobile picture.

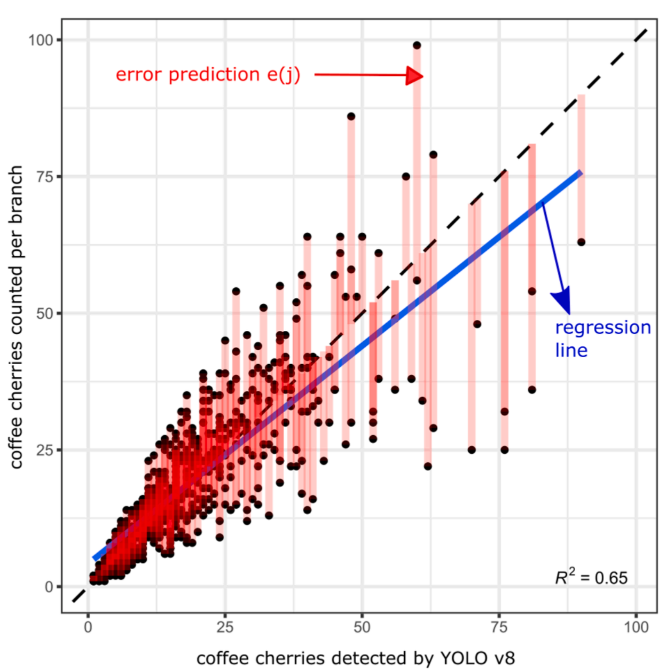

Fig. 2. Evaluating of the performance of the cherry detection model YOLO v8 against the number of cherries manually counted on the branch.

2.3. Factors explaining model prediction error

To investigate the sources of error $e(j)$, for each image, we assigned several attributes that can affect the quality of the image. These attributes are divided into five categories (Fig. 3): photographer influence, mobile phone characteristics, features of mobile pictures, geographic location, and coffee varieties. Table A1 **in Appendix 1** describes these variables in detail.

2.4. Statistical analysis

2.4.1. Analyzing variable associations among explanatory variables

We used Pearson correlation (r) to evaluate whether there was a linear relationship between the quantitative variables (Table A1 **in Appendix 1**). r measures the linearity between two continuous variables and utilizes the covariance matrix of the data to assess the strength of the relationship between two vectors (Sedgwick, 2012). Variables that were highly correlated were discarded for the LMM approach.

Principal Components Analysis (PCA) is a technique that transforms the data into a smaller set for easier interpretation and reveals hidden patterns through orthogonal (uncorrelated) components that retain the maximum possible variance from the original dataset (Dunteman, 1984). We applied PCA to analyze and identify the intrinsic relationships between quantitative variables that are not easily perceived through pairwise correlation analysis.

Multiple correspondence analysis (MCA) is an unsupervised learning technique used to reduce dimensionality and visualize patterns in multidimensional categorical data (Murtagh, 2007). We applied MCA to evaluate the associations between all qualitative variables (Table A1 **in**

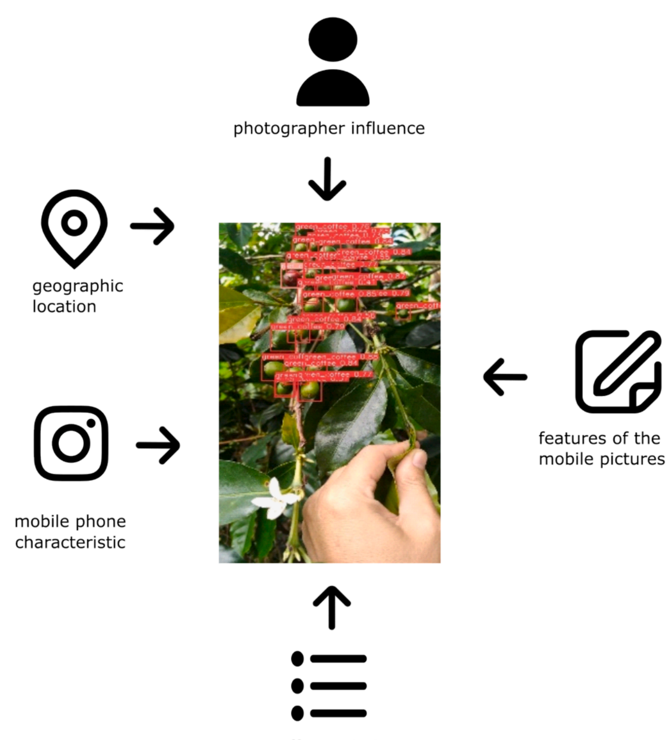

Fig. 3. Categories of variables that affect image data quality and the performance of YOLO v8. These factors can significantly affect the quality of the image data and thereby the task of object (cherry) detection with the trained deep learning model (YOLO v8).

Appendix 1).

MCA and PCA identify intrinsic patterns between variables that are not observable through correlation analysis. This suggests that the LMM results, particularly with respect to important variables, may be influenced by confounding factors that affect the outcomes.

2.4.2. Analyzing model error sources

We used LMM to identify the factors explaining the variability in $e(j)$. LMM allows to analice non-independent, multilevel/hierarchical, longitudinal or correlated data (Carnero-Alcázar et al., 2022). LMM is particularly suited for our study because the data were organized hierarchically and exhibited non-independence. The data structure is hierarchical because we analyzed mobile pictures taken by various independent photographers. Mobile pictures taken by the same photographer are not independent of each other, as these pictures are assumed to be more similar to one another than to those taken by other photographers.

LMMs are linear in terms of their parameters, and the covariates or independent variables may use both random and fixed effects (Pinheiro and Bates, 2000). The fixed effects are constant parameters associated with the entire population. Random effects are parameters exclusive to some individuals' random responses and are normally distributed (Pinheiro and Bates, 2000). The photographer is considered a random effect because each photographer influences the pictures differently due to factors such as personal style, level of digital literacy, smartphone features, or environmental conditions.

The general model structure is formulated in matrix notation as follows:

$$y = X\beta + Zb + e \tag{2}$$

Where y is a vector. The response variable in our case is $e(j)$. β is the vector of fixed effects. X is the matrix used to describe fixed effects related to observations to β. Z is the matrix related to random effects b, and e is the experimental error. The column "Type of variable for LMM" in Table A1 **in Appendix 1** lists the fixed effects and random effects used LMM. The photographer identification number (Photographer) was designated as the random variable, assuming that the factor could randomly modulate the intercepts of the fixed variables. The intercept refers to the baseline level of the dependent variable $e(j)$, when all fixed and random predictor variables are set to zero. We set $p < 0.05$ to be statistically significant.

We reported the results of LMM using *p-value* and effect size. Effect size refers to the magnitude of the difference between groups (Sullivan and Feinn, 2012). We used the index eta-squared η^2 that describes the proportion of variance explained by group membership (Albers and Lakens, 2018) and is defined as follows:

$$\eta^2 = \frac{{\sigma_B}^2}{{\sigma_T}^2} = \frac{SS_B}{SS_T} \tag{3}$$

Where σ_B^2 is the variance between the samples, σ_T^2 es la variance of within and between samples. SS_B is the sum of squares between samples and SS_T is the sum total of the sum of squares of between and within squares.

The indicators of LMM performance were $ConditionalR^2$, which takes into account both the fixed and random effects, and $MarginalR^2$, which considers only the variance of the fixed effects.

Moreover, we anticipated some strong interaction effects between photographer identity and the other variables. To explore such interaction effects, we utilized a decision tree machine learning model. This model was also used to explain $e(j)$ using all the variables in Table A1 **of Appendix 1**, including photographer identification number. The specific decision tree model employed was a conditional inference tree, which falls under the non-parametric class of decision trees (Hothorn et al., 2016).

The conditional inference tree uses a recursive algorithm in which data are divided using an algorithm that creates subsets based on statistically significant variables and values (Hothorn et al., 2016). Each node represents a point where the algorithm splits the data. The algorithm continues splitting the data until the null hypothesis of independence between the response and any of the variables can no longer be rejected (Hothorn et al., 2016). This tree model is robust, unbiased, and easy to interpret. Consequently, this method helped us understand the interactions between mixed categorical and numerical data in mobile pictures, where relations are nonlinear and unclear.

We used R v4.4.1 for data analysis. The statistical methods were implemented using the partykit v1.2–20 library for the ctree algorithm (Hothorn et al., 2016), lmerTest v3.1–3 for the LMM, FactoMineR v2.11 for PCA and MCA, and corrplot v0.95 for the Pearson correlation.

3. Results

3.1. The relationship between explanatory variables

The dimensions of image height and image width showed a strong relationship ($r = -0.74$), similar to the relationship between x-axis and y-axis resolution ($r = 1$), and the relationship between screen width and x-axis and y-axis resolution ($r = -0.95$) (Fig. 4). Image height, as well as x-axis and y-axis resolution, were discarded from the LMM to avoid multicollinearity in further analysis.

We applied the PCA technique to the quantitative variables (Table A1 **in Appendix 1**). The first five orthogonal and uncorrelated factors together explained 89.7 % of the total variance. The first component (dimension 1) explained 32.8 % (eigenvalue = 1.7) (x-axis in Fig. 5) and was primarily influenced by camera resolution (29.01 %) (Table 1). The second principal component (dimension 2) explained 17.5 % of the total variance (eigenvalue = 1.0) (y-axis in Fig. 5). The moment of inertia, representing the sum of variances in dimensions 1 and 2, was 47.2 %. Dimension 2 was strongly correlated with longitude (78.33 %). The third principal component was primarily associated with latitude (91.8 %), while the fourth principal component was influenced by image width (51.85 %). The fifth component was primarily influenced by screen width (51.5 %) (Table 1).

We applied MCA to evaluate all qualitative variables (Table A1 **in Appendix 1**). The first five orthogonal and uncorrelated factors together explained 49.7 % of the total variance. The variance explained by the first principal component (dimension 1) was 13.9 % (eigenvalue = 0.26), while the second principal component (dimension 2) explained 10 % (eigenvalue = 0.19). The moment inertia was 23.9 %.

Fig. 6 presents the results of the MCA algorithm in a two-dimensional plot (x- and y- representing dimensions 1 and 2, respectively). For the first component (dimension 1), the key category was the absence of a screen behind the branch to avoid hiding other cherries (Nonscreen: 18.37 %). The second component (dimension 2) was mainly associated with position in front of the branch (front: 17.61 %). Table A2 **in Appendix 1** presents the key categories associated with each of the first five principal components. The third principal component was linked to the Colombia variety (Variety:VColombia: 16.12 %). The fourth component corresponded to the position in the lower part of the tree (LowerTree: 19.39 %). The fifth component was characterized by the other varieties (Variety:Others 37.16 %).

The MCA revealed associations between the categories of variables. The use of a non-screen behind the branch (**"NonScreen"**, indicated in red on the x-axis) and the presence of the entire branch along with cherries from other branches (**"EntireBranch + O"**) were observed, regardless of the variety. The variety **Colombia** (**"V.Colombia"**) was associated with a random position of the camera (**"Random"**). Other varieties (**"V.Other"**), which refers to varieties other than Supremo, Variedad Colombia, or Castillo, were correlated with the camera being positioned above (**"Above"**) and with images displaying only a portion of the branch (**"PortionBranch"**).

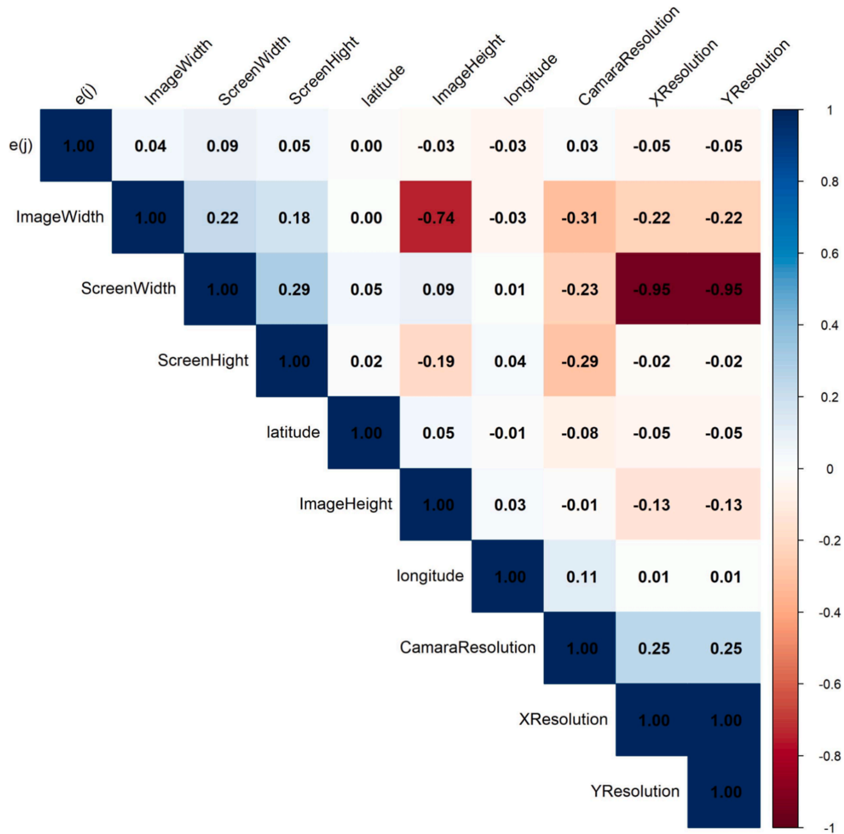

Fig. 4. Pearson coefficients *(r)* of all variable qualitative variables in Table A1 of Appendix 1.

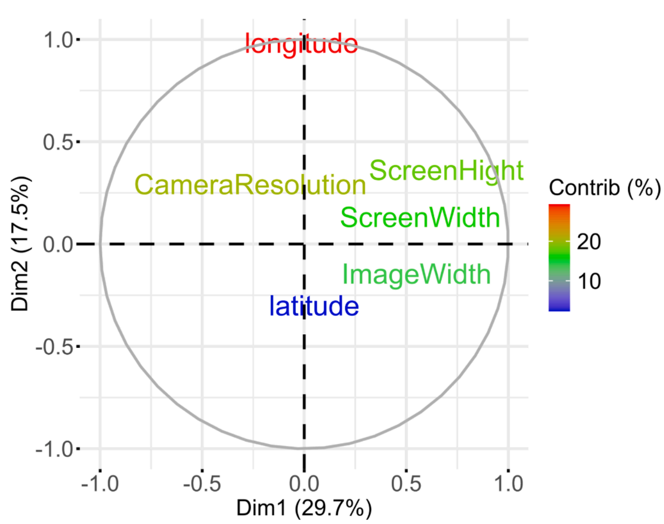

Fig. 5. Two-dimensional PCA plot of the first and second factors and the correlation between variables.

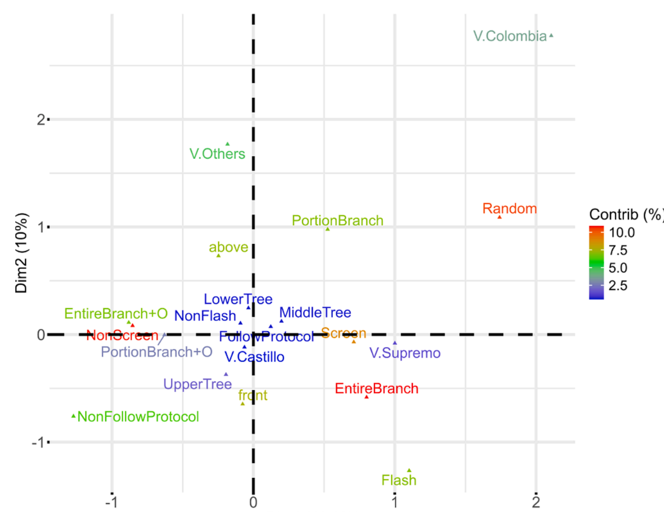

Fig. 6. Two-dimensional MCA plot of the first and second factors and the correlation between categorical variables.

Table 1
The contribution of each variable to each first five dimensions in PCA after removing the variables resolution in x, resolution in y and image height.

Variables	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
ImageWidth	22.17	0.42	3.86	51.85	2.63
ScreenWidth	23.52	4.23	0.15	10.37	56.73
ScreenHight	23.93	7.32	0.01	23.99	20.48
latitude	0.91	4.69	91.8	0.93	0.08
longitude	0.46	78.33	4.17	11.15	1.61
Camera Resolution	29.01	5.02	0.02	1.7	18.48

3.2. Factors influencing model performance

The distribution of $e(j)$. exhibits a bell-shaped curve, indicating a normal distribution, with the majority of values concentrated near the mean of −0.45. The data span a range from −39 to 51 (Fig. 7). There is a slight tendency for values to be more concentrated in the negative range, which suggests that YOLO v8 tends to detect fewer cherries than the actual number of cherries present.

The LMM model showed $ConditionalR^2$ of **0.23** and $MarginalR^2$ of **0.08**, indicating the strong influence of photographer identity. Table 2

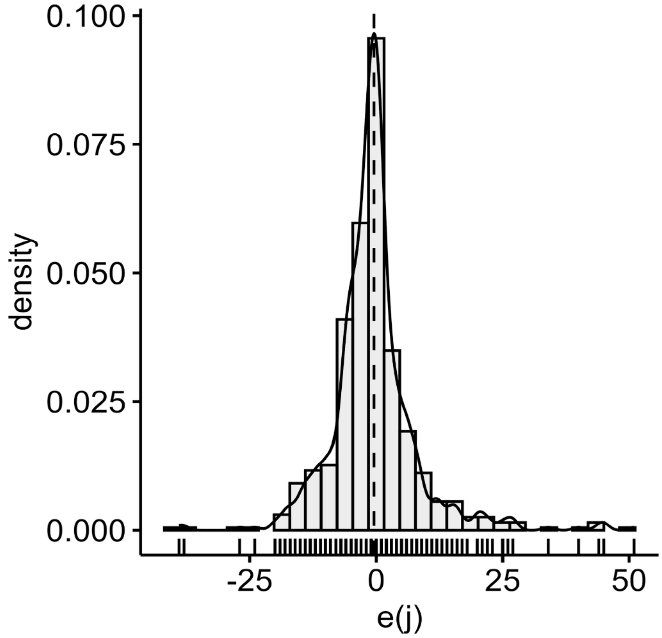

Fig. 7. Distribution of $e(j)$. The minimum value and maximum values of $e(j)$ are −39 and 51, respectively, with a median of −1 and a mean of −0.45.

presents the fixed coefficients from the LMM, revealing that the protocol for taking pictures played an important role in prediction accuracy (Instructions: NonFollowProtocol, $p = 0.00$**)**. The relative height of the branch (LocationBranch: UpperTree: $p = 0.05$) and the presence of other branches in the picture different from the entire main branch ('Features: EntireBranch_OtherBranches', $p = 0$) were also identified as important features for explaining the variability in $e(j)$. The conditional means, which represent the deviations of the intercepts for each level of the random effect relative to the overall intercept, demonstrate significant variability in $e(j)$ across photographers (Table 3).

As with the p-values, the effect sizes η^2 also showed that differences in instructions explain the largest portion of the variance (Instructions: $\eta^2 = 0.03$), followed by the features of mobile pictures (Features: $\eta^2 = 0.02$).

The influence of the photographer was also confirmed with the conditional inference tree (Fig. 8). The tree partitions the data based on the photographer. The group on the left side of the tree consists of five photographs and is further split by instructions, resulting in two groups: those who followed the instructions ($n = 208$, $\overline{e(y)} = 2.9$) and those who did not ($n = 37$, $\overline{e(y)} = 8.2$). In both cases, the median $e(j)$. is greater than zero, indicating that their pictures, when used in YOLO v8, tend to result in the model counting more cherries than the actual number of cherries based on a manual count—an effect of overestimation.

The group on the right side is further split by photographer, resulting in two subgroups of photographers: one group composed of three photographers ($n = 147$, $\overline{e(j)} = -1.2$) and another group composed of five photographers ($n = 245$, $\overline{e(j)} = -4.2$). In both cases, the median of $e(j)$ is less than zero, suggesting that images from these photographers, when processed by YOLO v8, tend to result in underestimation, with the model detecting fewer cherries than are actually present. Overall, overestimation occurs more frequently (67 %) than underestimation (33 %).

3.3. Post hoc analysis of key factors

We analyzed the impact of the photographer on model performance

Table 3
Results of conditional means for random effects (photographer). These are deviations of the intercepts for each level of the random effect relative to the overall effect.

photographer	conditional mean
1.00	5.35
10.00	−0.43
11.00	−1.90
12.00	−3.62
13.00	0.08
2.00	−1.33
3.00	−1.52
4.00	0.07
5.00	4.82
6.00	−2.55
7.00	2.03
8.00	−2.12
9.00	1.13

Table 2
Results of the linear mixed model (LMM) for fixed effects. The fixed variables are fully described in Table A1 of Appendix 1. The random effect is the person who took the picture (photographer). Alpha (significance level) = 0.05. The variables with $p < 0.05$ are emphasized in bold. The description column names are sourced from the lm4e library (Bates, 2014) in R. Estimate is the estimated effect of each fixed-effect variable on the target variable, holding other variables constant. Std. Error refers to the standard error of the estimate. df denotes the degrees of freedom for the fixed effects. t value: The ratio of the estimate to the standard error used to evaluate the statistical significance of fixed effects. Pr(>|t|) is the p-value associated with each t-statistic, indicating the probability that the observed effect occurred by chance.

Name	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	−2.20	1.81	79.89	−1.22	0.23
Variety: OTHERS	−1.16	1.52	616.40	−0.76	0.45
Variety: SUPREMO	0.03	1.52	616.32	0.02	0.99
Variety: VCOLOMBIA	−0.09	3.26	613.14	−0.03	0.98
BehindBranch: Screen	−1.14	0.93	479.68	−1.22	0.22
LocationBranch: MiddleTree	−0.31	0.75	608.03	−0.41	0.68
LocationBranch: UpperTree	**1.53**	**0.78**	**608.59**	**1.97**	**0.05**
Instructions: NonFollowProtocol	**4.61**	**1.12**	**610.53**	**4.12**	**0.00**
longitude	−0.30	0.32	614.57	−0.96	0.34
latitude	0.20	0.32	615.89	0.62	0.53
ImageWidth	−0.51	0.42	447.35	−1.22	0.22
AnglePicture: Front	0.86	0.70	617.00	1.22	0.22
AnglePicture: Random	1.19	1.32	616.36	0.90	0.37
Flash: NonFlash	−0.15	1.28	599.77	−0.12	0.91
ScreenWidth	0.46	1.01	9.47	0.46	0.66
ScreenHight	0.26	1.03	8.63	0.26	0.80
CameraResolution	0.16	0.99	10.33	0.16	0.88
Features: EntireBranch_OtherBranches	**3.04**	**0.96**	**616.13**	**3.17**	**0.00**
Features: PortionBranch	1.04	0.94	613.76	1.11	0.27
Features: PortionBranch_OtherBranches	1.01	0.97	616.56	1.04	0.30

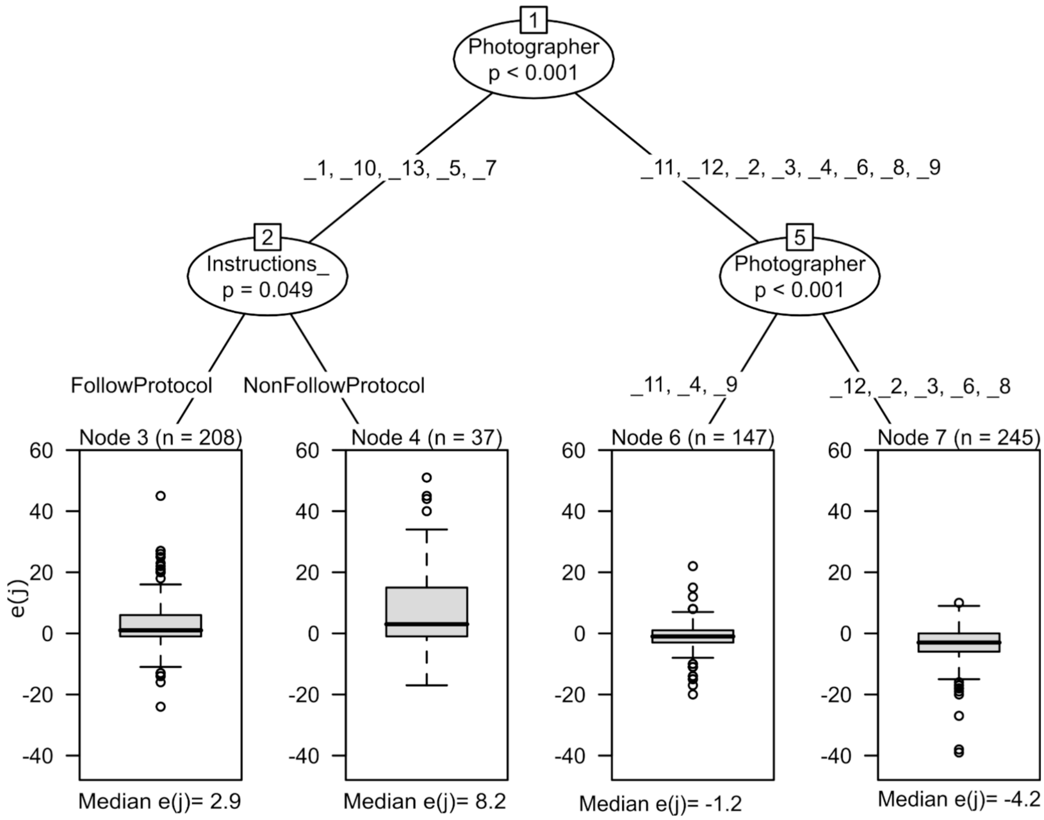

Fig. 8. Conditional inference tree examining model error using the same variables as LMM. "Photographer" refers to the person who took the picture, while "Instructions" indicates whether the picture follows the given directions.

by categorizing the pictures they took according to the Instructions variable (Table A1 **in Appendix 1**). These pictures were evaluated based on whether they followed the instructions to capture as many cherries as possible, capture them during daylight, and not capture cherries from other branches or trees. Consequently, we divided the dataset into two groups: pictures adhering to the protocol ("FollowProtocol") and pictures not adhering to the protocol ("NonFollowProtocol").

The performance of each photographer was evaluated along with their contributions to model accuracy. These results varied substantially, with R^2 values ranging from 0.45 to 0.93, and the number of detected cherries is densely below 25 coffee cherries per image. Statistical analysis confirmed differences in error prediction among photographers (Fig. 9**)**.

We compared the performance of YOLO v8 obtained from the mobile pictures that followed the protocol for taking pictures (88 %, n = 564) to those from mobile pictures that did not follow the protocols (12 %, n = 73). The comparison revealed a significant difference (R^2 of 0.48 and R^2 of 0.73) (Fig. 10).

4. Discussion

The results from the LMM and conditional inference tree reveal that variables associated with photographer influence were key predictors of $e(j)$. Specifically, whether the photographer followed the instructions for taking pictures (which included capturing as many cherries from the branch as possible during daylight) and whether the image contained cherries from other branches were significant factors. The MCA also revealed that not using a screen was associated with the presence of more cherries from other trees or branches and that the camera position was linked to the tree variety. This suggests that, in addition to the influence of the photographer, there are also tree variety–specific characteristics that affected the outcome. These differences between coffee trees may be attributed to management practices, vegetative stage, or topography, depending on the tree's location. For example, pictures of coffee trees located on hills are more difficult to take than pictures of coffee trees on flat ground.

The errors associated with the influence of the photographer and adherence to protocol emphasize the need for proper training and quality control measures for data collection (Ebitu et al., 2021; Lovell et al., 2009; Torney et al., 2019). The post hoc analysis suggests that following the protocol can improve the predictive performance from 48 % to 73 % (an increase of 25 %), and photographer identity can vary between 45 % and 93 %. This suggests that the design of a simple and easy protocol for data collection, coupled with appropriate training, can enhance deep learning performance in citizen science–based data collection.

Therefore, investigating how to enhance and secure the quality and quantity of data is critical for any citizen science–based study. Despite the high potential to generate enormous data and impressive scientific discoveries, there are active debates as to whether relying on a citizen science–based approach is valid to inform policy due to unreliable data (Hunter et al., 2013; Kosmala et al., 2016). This skepticism has led scientists to investigate bias and data quality in citizen science projects. Kosmala et al. (2016) found that data collected by unpaid volunteers were of the same quality as those produced by professionals. Swanson et al. (2016) found that collaboration between volunteers for classification objects can improved the performance model from 88.6 % to 97.9 %.

Adequate training, feedback, and customization of protocols regarding volunteers' skills facilitate high-quality data collection. Relevant training sessions must meet the needs of volunteers, and protocols must be customized to these needs (Hunter et al., 2013; Kosmala et al., 2016; Sullivan et al., 2014). We found that short sessions of around one hour using the local language appropriate for the agronomy of coffee plants, regular feedback on current data collection, and the use of social media channels such as WhatsApp or YouTube were highly

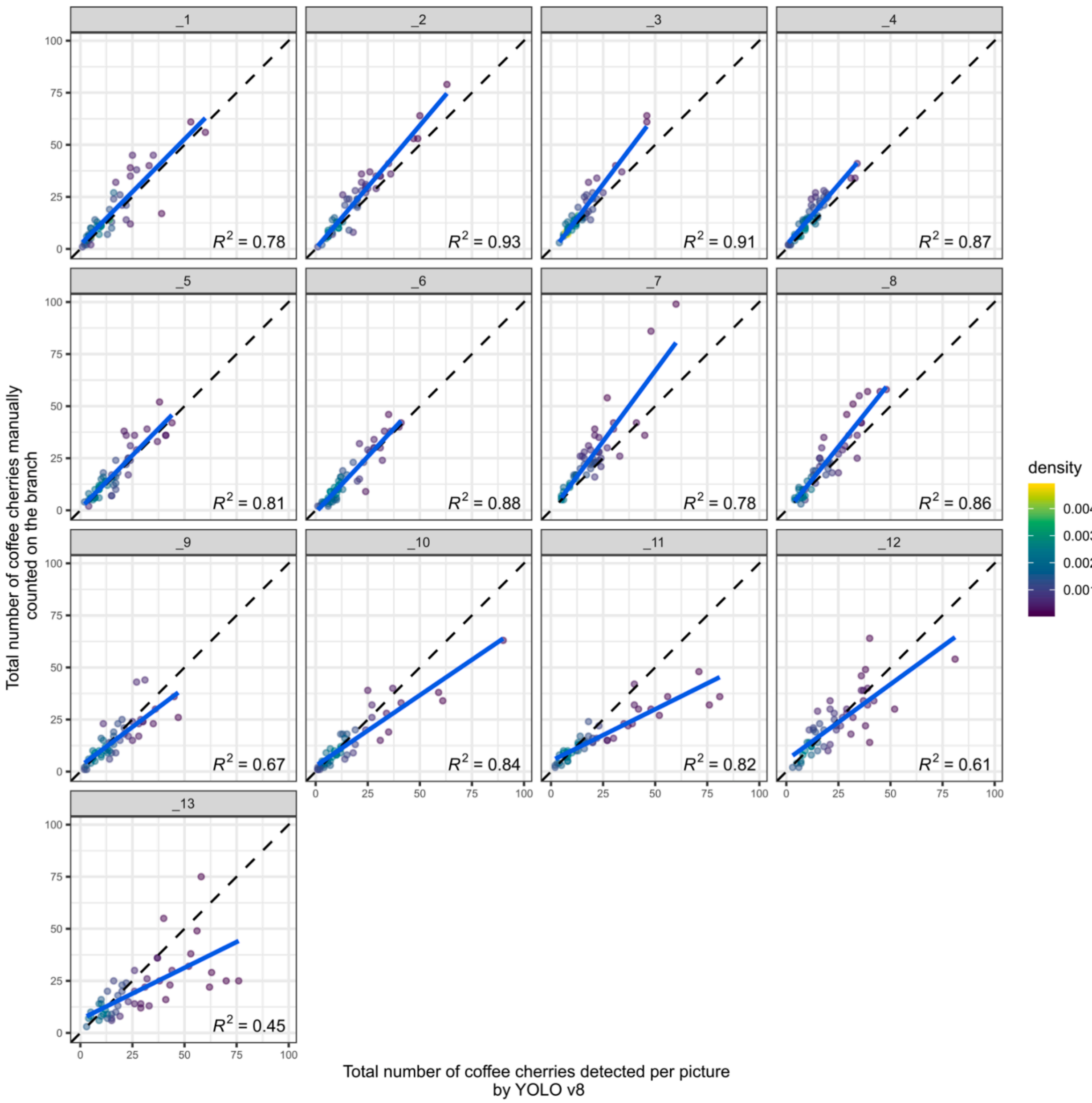

Fig. 9. Evaluation of the model performance of each photographer. Each graph represents the performance of YOLO v8 for each photographer. The lowest R^2 value was 0.45, while the highest was 0.93.

effective at encouraging the adoption of the protocols. We found farmers to maintain a high ability to use their cell phones' cameras, which allowed for more fluent and quick explanation.

One study has suggested that the rewards increased the interest of the volunteers in the project and boosted the volume of data collected (Sullivan et al., 2014). These rewards can help with the retention of volunteers, creating a core group of participants with advanced-level experience and resulting in more reliable data (Cooper et al., 2007). Additionally, these rewards are not necessary for funding. Sullivan et al. (2014) designed a plan that included rewards based on accessing privileged information about bird species. While we did not use rewards in our research, we explained the future benefits of high-precision production, which leads to lower-risk economic agriculture (Benami et al., 2021).

We further argue for other factors that did not reveal statistically significant relationships. We did not find a significant correlation between the two positions of the camera, upper and front of the branch; therefore, prediction errors warrant further investigation. In addition, the use of a screen behind the branch did not significantly influence the prediction errors related to overestimation. The results of the conditional inference tree showed that using a screen was not a crucial factor in preventing overestimation. However, while screens can improve the detection of coffee cherries in mobile pictures, they increase the complexity of the protocol. Taking a picture of only one branch is difficult without physically preventing other branches from being captured in the image. Similar devices, in addition to mobiles, are used in other crops to improve the quality of mobile pictures. For example, rectangular markers are used to delimit areas of interest in the detection of the Kohlrabi crop (Hernández-Hernández et al., 2017), platforms are utilized to hold the cellphones for phenotyping the seeds (Zhihong et al., 2016), and portable white backgrounds are employed to monitor foliar damage in soybean plants (Machado et al., 2016). Therefore, although this approach may improve model performance, it drastically increases the complexity of the protocol, which can render it impractical in many places.

Interestingly, geographic locations, including latitude, longitude,

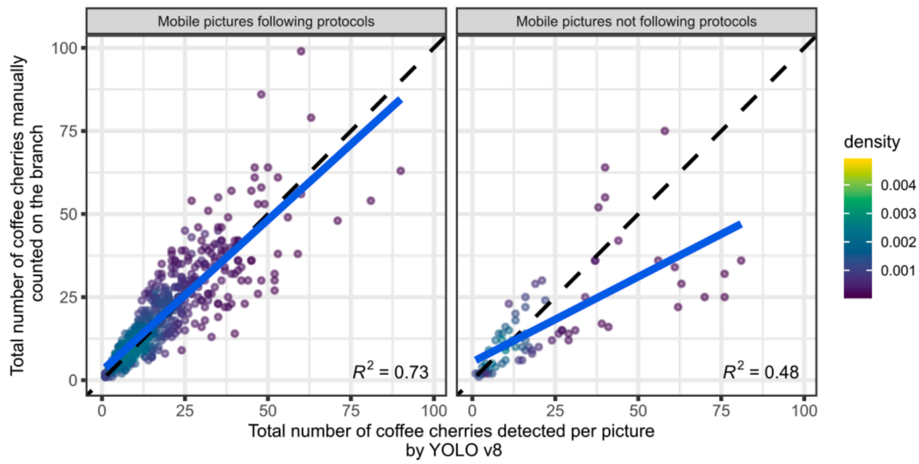

Fig. 10. Evaluating cherry detection model performance: The impact of the instructions given in the protocol for taking pictures ("Instruction_", Table A1 of Appendix 1).

and altitude, were not found to be key factors explaining the identified error. This finding aligns with the work Rivera-Palacio et al. (2024), suggesting that this method can be extended to other regions despite local differences in environment and management conditions. Additionally, this approach can help to address more specific extension programs through citizen science (Ryan et al., 2018). However, the focus of this study is on the Arabica species, which typically reaches an average height of around 2.5 m (Wintgens, 2004). For other species, the protocol may need adjustments to account for differences in height and branch architecture. Furthermore, ensuring high-level expert supervision in the field contributes to achieving sustained data quality (Lovell et al., 2009).

The use of mobile phones wielded citizen science and artificial intelligence learning tools for data collection across large geographic scales has become increasingly widespread in multiple fields of environmental science, such as sustainable agriculture (Ebitu et al., 2021; Rivera-Palacio et al., 2024; Ryan et al., 2018; Sauermann et al., 2020), in biology for monitoring biodiversity in flora and fauna species (Lee and Nel, 2020; Lovell et al., 2009; Sullivan et al., 2014) and in epidemiology for the monitoring, detection, and management of coronavirus diseases (Katapally, 2020). Given the increasing internet connectivity and the availability of low-cost smartphones, this trend seems likely to continue.

Our results did not show a strong relationship between smartphone features and $e(j)$. However, this finding is inconclusive due to two key factors. First, this work's small sample size and the lack of consideration of smartphone features when selecting photographers may have introduced bias into the results. Second, photographers were confounded with smartphone characteristics, as each photographer had a unique smartphone. Thus, the variables related to photographers were implicitly constrained by the smartphone characteristics. Similarly, the effects of sunlight and the photographer's individual skill were correlated because each enumerator took pictures at the same hour. Due to the study design, it was impossible to decorrelate various factors. Therefore, future studies should implement a more rigorous design to disentangle these effects.

There are two considerations regarding the AI model approach. First, the dataset was unbalanced in the categories of cherries; the number of green cherries greatly exceeded the number of other classes, such as red cherries. However, this class imbalance is unlikely to affect our results, as our interest lies in the total number of coffee cherries and not in their classification. Nonetheless, this imbalance could lead to overfitting. The second consideration is that only one detection model was used. It would have been ideal to compare several models and analyze their error structures to achieve greater levels of generalization. Nevertheless, Mengsuwan et al. (2024) demonstrated that YOLO v8 achieves high accuracy that is comparable to those of the state-of-the-art foundation models.

This study focused solely on coffee cherries, which cannot be generalized to various crop types. Coffee cherries have a particular size; they are relatively small and have specific shapes. These findings could be extended to images of other fruits with similar characteristics, such as cocoa or apples.

5. Conclusions

This paper highlighted the external factors that affected the prediction of a coffee cherry detection model based on the YOLO v8 algorithm using mobile pictures taken by farmers. Factors exogenous to the model were categorized into five main groups: photographer influence, mobile picture features, smartphone features, localization, and coffee varieties. Our analysis revealed correlations between these factors and their impact on the prediction error $e(j)$. The results from the LMM and the conditional inference tree indicated that each photographer's unique influence played a crucial role in prediction error. Depending on the photographer's identity, the model's predictive performance varied, with R^2 values ranging from 0.45 to 0.93. Moreover, adherence to protocols resulted in R^2 values ranging from 0.48 to 0.73. Furthermore, correlation analysis using MCA revealed that factors such as the use of a screen and camera position were related to coffee varieties. In addition, PCA and Pearson correlation revealed associations between mobile phone features, such as screen width and resolution (x and y). Mobile phone features were not statistically important; since each farmer used a unique smartphone, its influence was intrinsically linked to the photographer.

These insights underscore the potential of integrating citizen science with smartphone-based image collection, which holds great promise for advancing coffee yield prediction. Ensuring data quality and consistency, providing a simple protocol and proper training to citizen scientists, and providing feedback to individual photographers remain important considerations.

6. Funding sources

This work was supported by the Brandenburg University of Technology Cottbus-Senftenberg (BTU) Graduate Research School (GRS) cluster project "Integrated analysis of Multifunctional Fruit production landscape to promote ecosystem services and sustainable land-use under climate change" [grant number BTUGRS2018_19] and by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) [grant number 81275837].

Declaration of generative AI and AI-assisted technologies in the writing process

The author(s) used ChatGPT, Grammarly, and DeepL to check the grammar of the text during the preparation of this work. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the final content of the publication.

CRediT authorship contribution statement

Juan C. Rivera-Palacio: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Christian Bunn:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Masahiro Ryo:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to the anonymous reviewers for their constructive comments. We appreciate Eric Rahn, Javier Hoyos, Paul Schmidt, Sam Webb, Christian Feil, Deisy Little-Savage, Anton Eltzinger, Romain Gautron, and Francy Viviana Narvaez for their contributions during the development of this research. We also appreciate the entire team of technicians at TECNICAFE. This work was supported by the Brandenburg University of Technology Cottbus-Senftenberg (BTU), Graduate Research School (GRS) cluster project "Integrated analysis of Multifunctional Fruit production landscape to promote ecosystem services and sustainable land-use under climate change" (grant number BTUGRS2018_19) and the Croppie funded by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) (grant number 81275837).

Appendix 1

Table A1
The description of each variable.

Categorie	Variable	Description	Type	Values	Variables for LMM	Number Mobile Pictures
features of mobile pictures	BehindBranch_	If the picture of a branch was taken in front of a screen sheet to hide the other non-targeted objects, such as other branches and scenery	categorical	"Screen" = the picture used a screen for hiding other cherries.		
"NonScreen" = No screen used. "NonScreen" = 289	Fixed Effect	"Screen" = 348				
	Flash_	If the flash was used when the picture was taken.	categorical	"NonFlash" = Flash did not fire.		
"Flash"=Flash fired. "Flash" = 49	Fixed Effect	"NonFlash" = 588				
	AnglePicture_	The position of the camera in relation to the branch, from the front or from above.	categorical	"Front"= the picture was taken in front of the branch.		
"Random" = The picture does not have a properly defined position.						
"Above" = The picture was taken from above the branch. "Random" = 49 "Above" = 237	Fixed Effect	"Front" = 351				
	Features_	Describe the branch in the picture, noting whether it is completely or partially visible. Also, indicate if the picture includes branches from other trees or from the same tree.	categorical	"EntireBranch" = the image capturing the entire branch.		
"EntireBranch_ OtherBranches" = The image includes the entire branch and branches from other trees or the same tree. "PartialBranch" = The image only partially includes the branch. "PortionBranch_						

(continued on next page)

Table A1 (*continued*)

Categorie	Variable	Description	Type	Values	Variables for LMM	Number Mobile Pictures
OtherBranches“ = The image includes a partial branch and other branches. “EntireBranch_ OtherBranches“ = 170 “PartialBranch” = 112 “PortionBranch_ OtherBranches“ = 135	Fixed Effect	“EntireBranch” = 220				
	LocationBranch_ The location of the branch in the tree. In the upper, middle or lower part of the tree.	categorical	“LowerTree”=the branch is located in the upper part of the tree.			
“MiddleTree”=the branch is located in the middle part of the tree. “UpperTree” = the branch is located in the lower part of the tree. “MiddleTree”= 237 “UpperTree” = 207	Fixed Effect	“LowerTree” = = 193				
Features smartphone	ImageWidth	The length of width picture	numeric	Min = 209px Max = 1024px.	Fixed Effect	637
	ImageHeight	The length of height picture	numeric	Min = 460px		
Max = 1024px		637				
	XResolution	Resolution in x	numeric	Min = 1 pixels per inch.		
Max = 72 pixels per inch.		637				
	YResolution	Resolution in y	numeric	Min = 1 pixels per inch,		
Max = 72 pixels per inch		637				
	ScreenWidth	The width of the mobile screen	numeric	Min = 7.0		
Mean = 7.619 Max = 10.1	Fixed Effect	637				
	ScreenHeight	The height of the mobile screen	numeric	Min = 14.34		
Mean = 15.89 Max = 16.96	Fixed Effect	637				
	ResolutionFrontCamerainMP_	Resolution of the frontal camera	numeric	Min = 5 Megapixels (MP)		
Max = 64 MP	Fixed Effect	637				
geographic location	Longitude	The longitude where the picture was taken	numeric	Max = 34.6		
Min = -76.90	Fixed Effect	637				
	Latitude	The latitude where the picture was taken	numeric	Max = 36.85		
Min = 2.37	Fixed Effect	637				
photographer Influence	Photographer	The photographer identification number	categorical	Thirteen photographers	Random Effect	49 picture per photographer
	Instruction_	The image follows the instructions given in the protocol for taking a picture.	categorical	“FollowProtocol”= the picture meets with instructions of the protocol.		
“NonFollowProtocol” = the picture doesn’t follow the instructions. “NonFollowProtocol” = 56	Fixed Effect	“FollowProtocol”= 581				
Coffee varieties	Variety	The name of coffee variety	categorical	“Variety Colombia”, “Castillo”, “Supremo”, “Ophther varieties”	Fixed Effect	“Variety Colombia” = 6
“Castillo” = 572 “Supremo” = 29 “other varieties” = 30						

Fig. A1. A) a picture of an entire coffee branch, B) a picture of an entire coffee branch along with other branches, C) a picture of a partial coffee branch along with other branches, D) a picture of a portion of a branch, E) a picture taken with flash, F) a picture taken without flash, G) a picture not following the protocols, H) a picture adhering to the protocols, I) a picture taken from above the branch, J) a picture taken without a defined position (randomly), K) a picture taken at the lower part of the tree, L) a picture taken in the middle part of the tree, M) a picture taken at the upper part of the tree, N) a picture of the Castillo variety, and O) a picture of the Supremo variety.

Table A2
The contribution of each variable to the first five dimensions in MCA.

variable	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Variety: Castillo	0.2	0.95	0	0.15	1.3
Variety: Others	0.09	11.29	9.37	0.83	37.16
Variety: Supremo	2.53	0.02	1.18	1.3	0.76
Variety: V.Colombia	2.32	5.57	16.12	18.12	0.33
NonScreen	18.37	0.24	0	0.01	2.95
Screen	15.26	0.2	0	0.01	2.45
LowerTree	0.02	1.42	0.53	19.39	0.77
MiddleTree	0.81	0.43	15.28	14.69	0.44
UpperTree	0.68	3.44	12.1	0.02	2.44
FollowProtocol	0.76	0.37	1.06	0.21	0.19
NonFollowProtocol	7.89	3.89	11.02	2.23	1.93
above	1.26	15.25	0.28	11.38	3.57
front	0.18	17.61	0	2.5	0.81
Random	12.93	7	1.25	10.14	3.04
Flash	5.18	9.43	0	10.99	0.7
NonFlash	0.43	0.79	0	0.92	0.06
EntireBranch	12.26	8.99	3.1	0.01	1.74
EntireBranch_OtherBranches	11.51	0.26	10.02	0.72	11.67
PortionBranch	2.68	12.87	10.73	4.9	6.96
PortionBranch_OtherBranches	4.65	0	7.93	1.47	20.74

Data availability

The data is freely available and can be downloaded https://github.com/j-river1/FactorsDeepLearningCitizenScience website.

References

Albers, C., Lakens, D., 2018. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. J. Exp. Soc. Psychol. 74, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004.

Baig, M.B., Straquadine, G.S., 2014. Sustainable Agriculture and Rural Development in the Kingdom of Saudi Arabia: Implications for Agricultural Extension and Education, in: Behnassi, M., Syomiti Muteng'e, M., Ramachandran, G., Shelat, K.N. (Eds.), Vulnerability of Agriculture, Water and Fisheries to Climate Change: Toward Sustainable Adaptation Strategies. Springer Netherlands, Dordrecht, pp. 101–116. 10.1007/978-94-017-8962-2_7.

Barbedo, J.G.A., 2022. Deep learning applied to plant pathology: the problem of data representativeness. Trop. Plant Pathol. 47, 85–94. https://doi.org/10.1007/s40858-021-00459-9.

Barbedo, J.G.A., 2018. Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng. 172, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013.

Bates, D., 2014. Fitting linear mixed-effects models using lme4. ArXiv Prepr. ArXiv14065823.

Benami, E., Jin, Z., Carter, M.R., Ghosh, A., Hijmans, R.J., Hobbs, A., Kenduiywo, B., Lobell, D.B., 2021. Uniting remote sensing, crop modelling and economics for agricultural risk management. Nat. Rev. Earth Environ. 2, 140–159. https://doi.org/10.1038/s43017-020-00122-y.

Carnero-Alcázar, M., Montero-Cruces, L., Maroto-Castellanos, L., 2022. Mixed models: an essential tool for non-independent data analysis. Eur. J. Cardiothorac. Surg. 62, ezac462. https://doi.org/10.1093/ejcts/ezac462.

Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. Ecol. Soc. 12. https://www.ecologyandsociety.org/vol12/iss2/art11/.

Dunteman, G.H., 1984. Introduction to multivariate analysis. No Title.

Ebitu, L., Avery, H., Mourad, K.A., Enyetu, J., 2021. Citizen science for sustainable agriculture – A systematic literature review. Land Use Policy 103, 105326. https://doi.org/10.1016/j.landusepol.2021.105326.

Intelligence, G.S.M.A., 2024. GSMA intelligence [WWW Document]. Extensive Datasets Glob, Reach https://www.gsmaintelligence.com/data/ (accessed 5.21.24.

Hernández-Hernández, J.L., Ruiz-Hernández, J., García-Mateos, G., González-Esquiva, J. M., Ruiz-Canales, A., Molina-Martínez, J.M., 2017. A new portable application for automatic segmentation of plants in agriculture. Agric. Water Manag. 183, 146–157. https://doi.org/10.1016/j.agwat.2016.08.013.

Hothorn, T., Hornik, K., Zeileis, A., 2016. Unbiased recursive partitioning: A conditional inference framework. J. Comput. Graph. Stat. 15, 651–674. https://doi.org/10.1198/106186006X133933.

Hunter, J., Alabri, A., van Ingen, C., 2013. Assessing the quality and trustworthiness of citizen science data. Concurr. Comput. Pract. Exp. 25, 454–466. https://doi.org/10.1002/cpe.2923.

G. Jocher, A. Chaurasia, J. Qiu, 2023. Ultralytics YOLO.

Katapally, T.R., 2020. A global digital citizen science policy to tackle pandemics like COVID-19. J. Med Internet Res 22, e19357. https://doi.org/10.2196/19357.

Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. Front. Ecol. Environ. 14, 551–560. https://doi.org/10.1002/fee.1436.

Lee, A.T.K., Nel, H., 2020. BirdLasser: The influence of a mobile app on a citizen science project. Afr. Zool. 55, 155–160. https://doi.org/10.1080/15627020.2020.1717376.

T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, 2014. Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 740–755. 10.48550/arXiv.1405.0312.

Lovell, S., Hamer, M., Slotow, R., Herbert, D., 2009. An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. Biodivers. Conserv. 18, 3295–3307. https://doi.org/10.1007/s10531-009-9642-2.

Machado, B.B., Orue, J.P.M., Arruda, M.S., Santos, C.V., Sarath, D.S., Goncalves, W.N., Silva, G.G., Pistori, H., Roel, A.R., Rodrigues-Jr, J.F., 2016. BioLeaf: A professional mobile application to measure foliar damage caused by insect herbivory. Comput. Electron. Agric. 129, 44–55. https://doi.org/10.1016/j.compag.2016.09.007.

Mengsuwan, K., Rivera-Palacio, J.C., Ryo, M., 2024. ChatGPT and general-purpose AI count fruits in pictures surprisingly well without programming or training. Smart Agric. Technol. 9, 100688. https://doi.org/10.1016/j.atech.2024.100688.

Murtagh, F., 2007. Multiple correspondence analysis and related methods. Psychometrika 72, 275–277. https://doi.org/10.1007/s11336-006-1579-x.

Pinheiro, J.C., Bates, D.M., 2000. Linear mixed-effects models: basic concepts and examples. Mix.-Eff. Models -plus 3–56. https://doi.org/10.1007/978-1-4419-0318-1_1.

J.C. Rivera-Palacio, 2024. Factors affecting deep learning model performance in citizen science-based image data collection in agriculture. https://github.com/j-river1/FactorsDeepLearningCitizenScience.

Rivera-Palacio, J.C., Bunn, C., Rahn, E., Little-Savage, D., Schmidt, P., Ryo, M., 2024. Geographic-scale coffee cherry counting with smartphones and deep learning, 0165 Plant Phenomics Wash. DC 6. https://doi.org/10.34133/plantphenomics.0165.

Ryan, S.F., Adamson, N.L., Aktipis, A., Andersen, L.K., Austin, R., Barnes, L., Beasley, M. R., Bedell, K.D., Briggs, S., Chapman, B., Cooper, C.B., Corn, J.O., Creamer, N.G., Delborne, J.A., Domenico, P., Driscoll, E., Goodwin, J., Hjarding, A., Hulbert, J.M., Isard, S., Just, M.G., Kar Gupta, K., López-Uribe, M.M., O'Sullivan, J., Landis, E.A., Madden, A.A., McKenney, E.A., Nichols, L.M., Reading, B.J., Russell, S., Sengupta, N., Shapiro, L.R., Shell, L.K., Sheard, J.K., Shoemaker, D.D., Sorger, D.M., Starling, C., Thakur, S., Vatsavai, R.R., Weinstein, M., Winfrey, P., Dunn, R.R., 2018. The role of citizen science in addressing grand challenges in food and agriculture research. Proc. R. Soc. B Biol. Sci. 285, 20181977. https://doi.org/10.1098/rspb.2018.1977.

Ryo, M., Schiller, J., Stiller, S., Rivera Palacio, J.C., Mengsuwan, K., Safonova, A., Wei, Y., 2023. Deep learning for sustainable agriculture needs ecology and human involvement. J. Sustain. Agric. Environ. 2, 40–44. https://doi.org/10.1002/sae2.12036.

Sauermann, H., Vohland, K., Antoniou, V., Balázs, B., Göbel, C., Karatzas, K., Mooney, P., Perelló, J., Ponti, M., Samson, R., Winter, S., 2020. Citizen science and sustainability transitions. Res. Policy 49, 103978. https://doi.org/10.1016/j.respol.2020.103978.

Sedgwick, P., 2012. Pearson's correlation coefficient. Bmj 345.

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J. W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliff, M.J., Lagoze, C., Sorte, F.A.L., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K.V., Trautmann, N.M., Wiggins, A., Winkler, D.W., Wong, W.-K., Wood, C.L., Yu, J., Kelling, S., 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biol. Conserv. 169, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003.

Sullivan, G.M., Feinn, R., 2012. Using effect size-or why the P value is not enough. J. Grad. Med. Educ. 4, 279–282. https://doi.org/10.4300/JGME-D-12-00156.1.

Swanson, A., Kosmala, M., Lintott, C., Packer, C., 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol. J. Soc. Conserv. Biol. 30, 520–531. https://doi.org/10.1111/cobi.12695.

Tanaka, Y., Watanabe, T., Katsura, K., Tsujimoto, Y., Takai, T., Tanaka, T.S.T., Kawamura, K., Saito, H., Homma, K., Mairoua, S.G., Ahouanton, K., Ibrahim, A., Senthilkumar, K., Semwal, V.K., Matute, E.J.G., Corredor, E., El-Namaky, R., Manigbas, N., Quilang, E.J.P., Iwahashi, Y., Nakajima, K., Takeuchi, E., Saito, K., 2023. Deep learning enables instant and versatile estimation of rice yield using ground-based RGB images, 0073 Plant Phenomics 5. https://doi.org/10.34133/plantphenomics.0073.

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M., Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. Methods Ecol. Evol. 10, 779–787. https://doi.org/10.1111/2041-210X.13165.

J.N. Wintgens, 2004. Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers. 10.1002/9783527619627.ch1.

Zhihong, M., Yuhan, M., Liang, G., Chengliang, L., 2016. Smartphone-Based visual measurement and portable instrumentation for crop seed phenotyping. IFAC-Pap. 49, 259–264. https://doi.org/10.1016/j.ifacol.2016.10.048.