horticultura brasileira

Research

FERNANDES, ACG; VALADARES, NR; RODRIGUES, CHO; ALVES, RA; GUEDES, LLM; MAGALHÃES, JR; SILVA, RB; GOMES, LSP; AZEVEDO, AM. 2022. Feasibility of computational vision in the genetic improvement of sweet potato root production. *Horticultura Brasileira* 40: 378-383. DOI: http://dx.doi.org/10.1590/s0102-0536-20220405

Feasibility of computational vision in the genetic improvement of sweet potato root production

Ana Clara G Fernandes [1]; **Nermy R Valadares** [2]; **Clóvis Henrique O Rodrigues** [1]; **Rayane A Alves** [1]; **Lis Lorena M Guedes** [1]; **Jailson R Magalhães** [1]; **Rafael B da Silva** [1]; **Luan S de P Gomes** [1]; **Alcinei M Azevedo** [1]

[1]Universidade Federal de Minas Gerais (UFMG-ICA), Montes Claros-MG, Brasil; anaclaragoncalvesfernandes@gmail.com; clovis.rod1@gmail.com; rayaneaguiaralves@gmail.com; lislorenamelucio@gmail.com; jailsonrmagalhaes@hotmail.com; rafael.bolina2@gmail.com; luanspg@hotmail.com; alcineimistico@hotmail.com (author for correspondence); [2]Instituto Federal do Norte de Minas Gerais (IFNMG), Araçuaí-MG, Brasil; nermyrv@gmail.com

ABSTRACT

The improvement of sweet potato is a costly job due to the large number of characteristics to be analyzed for the selection of the best genotypes, making it necessary to adopt new technologies, such as the use of images, associated with the phenotyping process. The objective of this research was to develop a methodology for the phenotyping of the root production aiming genetic improvement of half-sib sweet potato progenies through computational analysis of images and to compare its performance to the traditional methodology of evaluation. Sixteen half-sib sweet potato families in a randomized block design with 4 replications were evaluated. At plant level, the weight per root and the total number of roots were evaluated. The images were acquired in a "studio" made of mdf with a digital camera model Canon PowerShotSX400 IS, under artificial lighting. The evaluations were carried out using the R software, where a second-degree polynomial regression model was fitted to predict the root weight (in grams) and the genetic values and expected gains were obtained. It was possible to predict the root weight at plant and plot level, obtaining high coefficients of determination between the predicted and observed weight. Computer vision allowed the prediction of root weight, maintaining the genotype ranking and consequently the similarity between the expected gains with the selection. Thus, the use of images is an efficient tool for sweet potato genetic improvement programs, assisting in the crop phenotyping process.

Keywords: *Ipomoea batatas*, image analysis, high efficiency phenotyping.

RESUMO

Viabilidade da visão computacional no melhoramento genético na produção de raízes de batata-doce

O melhoramento da batata-doce é um trabalho oneroso em decorrência do grande número de características analisadas para a seleção dos melhores genótipos, fazendo-se necessária a adoção de novas tecnologias, como o uso de imagens, associadas ao processo de fenotipagem. Objetivou-se desenvolver uma metodologia para a fenotipagem da produção de raízes para o melhoramento genético de progênies de meios irmãos de batata-doce por meio da análise computacional de imagens e comparar seu desempenho com a metodologia tradicional de avaliação. Foram avaliadas 16 famílias de meios irmãos de batata-doce em delineamento de blocos casualizados com 4 repetições. Avaliou-se a nível de plantas o peso por raiz. As imagens foram adquiridas em um "estúdio" feito com mdf com uma câmera digital modelo Canon PowerShotSX400 IS, sob iluminação artificial. As avaliações foram realizadas no software R, onde para a predição do peso das raízes (em gramas) foi ajustado um modelo de regressão polinomial de segundo grau e foram obtidos os valores genéticos e ganhos esperados. Foi possível predizer o peso das raízes a nível de plantas e de parcelas, obtendo altos coeficientes de determinação entre o peso predito e observado. A visão computacional permitiu a predição do peso das raízes, mantendo o ranqueamento dos genótipos e consequentemente a similaridade entre os ganhos esperados com a seleção. Assim, o uso de imagens é uma ferramenta eficiente para programas de melhoramento genético da batata-doce, auxiliando no processo de fenotipagem da cultura.

Palavras-chave: *Ipomoea batatas*, análise de imagens, fenotipagem de alta eficiência.

Received on May 10, 2022; accepted on August 30, 2022

Among the most important vegetables in Brazil is the sweet potato (*Ipomoea batatas*) (Vargas *et al*., 2017). Due to its rusticity, easy handling and low production cost, sweet potato is widely used by small producers, mainly due to its multiple uses in human and animal food. Thus, the vegetable stands out for its great socioeconomic importance, as well as for the wide genotypic and phenotypic variety (Otoboni *et al*., 2020). In this sense, efforts for the genetic improvement of the crop are necessary, aiming at increasing the productivity and quality of the roots.

The genetic improvement of plants is considered one of the most important strategies in modern agriculture, aiming to maximize the productivity and quality of food through individuals superior to those existing today (Amabile *et al.*, 2018). Improved varieties are generally more productive, more resistant, have higher nutraceutical quality and, consequently, reduced cost of production (Rincker *et al.*, 2014; Huang *et al.*, 2016).

The genetic improvement of sweet potato is an expensive task due to the large number of characteristics to be analyzed, since inferences about genetic materials in field experiments, in order to classify those candidates to be released as cultivars, is not an easy task, because these inferences should be based on true genotypic values and not just phenotypic values (Borges *et al.*, 2010). Thus, it demands a large amount of labor, funding and time availability, especially in work with half-sib progenies, since the character analysis must be performed at the plant level. In view of the difficulty in analyzing sweet potato phenotypic data, it is necessary to use alternative phenotyping tools. Thus, the development of high-throughput technologies is of paramount importance for plant genetic improvement, as it allows large-scale data collection with minimal effort (Zhou *et al.*, 2018). In this sense, images analysis allows to increase the performance and accuracy of phenotyping activities, making it possible to evaluate qualitative and productivity attributes at a lower cost, greater speed, efficiency and precision. This method also aids in the prediction of characteristics and decision making in addition to allowing the reproducibility of the results after discarding the plant material, since the image bank is preserved. The computer vision approach has already been used to characterize sweet potato size and shape phenotypes using high-throughput images (Haque *et al.*, 2021), however for quantitative characters this analysis is still scarce.

The objective of this research was to develop a methodology for the phenotyping of root production for the genetic improvement of sweet potato half-sib progenies through computational image analysis and to compare its performance with the traditional evaluation methodology.

MATERIAL AND METHODS

Installation and evaluation of the research

The research was carried out in the municipality of Montes Claros-MG (16°40’58”S, 43°50’20”W, 625 m altitude) where 16 progenies of sweet potato half-sibs were evaluated [BELGARD (F2), CAMBRAIA (F4), LICURI (F5), UFVJM40 (F6), UFVJM01 (F7), ARRUBA (F8), UFVJM05 (F10), UFVJM15 (F13), UFVJM56 (F16), UFVJM31 (F20), UFVJM37 (F22), UFVJM54 (F24), UFVJM25 (F26), UFVJM29 (F27), TCARRO02 (F29), UFVJM09 (F25)], in a haplic cambisol under irrigation conditions.

Sweet potato half-sibs were obtained by collecting seeds from the germplasm bank composed of elite accessions brought from the Federal University of the Jequitinhonha and Mucuri Valleys (UFVJM) and cultivated at the Institute of Agricultural Sciences (UFMG-ICA). Seeds were collected daily between April and October 2018 and stored in a refrigerator at 4°C. Subsequently, the seeds were subjected to mechanical scarification with sandpaper to break dormancy (tegumentary impermeability) and planted in 72-cell polystyrene trays with commercial substrate composed of sphagnum peat, expanded vermiculite and roasted rice husk. The trays were kept in a greenhouse and irrigated daily for 2 months, when the seedlings were ready for planting.

Planting was carried out in windrows, in a randomized block design with 16 treatments (different progenies) and 4 replications, with rows spaced 1 m apart and 0.4 m between plants. As the assessments were carried out at the plant level, a larger spacing was used in order to facilitate the identification of each plant and facilitate harvesting. Fertilization and cultural treatments were carried out as recommended for the crop by Filgueira (2008). In order to meet the crop’s requirements, 180 kg/ha of phosphorus and 30 kg/ha of nitrogen were used. 30 days after planting the seedlings, a top dressing with 30 kg/ha of nitrogen was carried out. Potassium fertilization was not necessary according to the chemical analysis of the soil.

To keep the soil with adequate moisture, initially, sprinkler irrigation was applied every day. After the critical period of crop establishment, irrigation was applied twice a week.

Plants were harvested manually, 165 days after planting, and analyzed at plant level, where the excess soil was initially removed from the roots and obtained their individual weights.

Image acquisition and processing

Image acquisition was carried out in a “studio” made with mdf with dimensions at the bottom of 0.50x1.00 m and height of 1.0 m (Figure 1). A digital camera model Canon PowerShotSX400 IS was used, under artificial lighting, with a fluorescent lamp. The camera was attached to a support, aiming to standardize the acquisition of images, so that all images were obtained from the same height (70 cm) and angle (90°). The roots were placed on a black background and spaced apart without overlapping.

Image analysis followed the steps shown in Figure 2, performed in the R s*oftware* (R Core Team, 2019). To compress the original images (4608 x 3456 pixels), in order to enable the computational analysis in less time, the *read_image* and *resize_image* functions from the *ExpImage* package were used. Images were compressed to 1000 x 750 pixels in width and height. After compression, the images were converted to grayscale using the *gray_scale* function, obtaining a matrix of 1000 columns and 750 lines with values ranging between 0 and 1. For the segmentation of each image, in order to separate the root of the bottom, a threshold (*treshold*) of 0.35 was used. Thus, all pixels with a value below this threshold were considered as background (black), attributing a value of 1, to the other pixels, corresponding to the roots, a value of 0. After segmentation, to obtain the area in pixels of each of the roots was used the *meansure_image* function.

Prediction of root weight and estimation of genetic parameters and selection gains

For the prediction of root weight (in grams) the adjustment of the first and second degree polynomial regression model was tested. To select the best fit, the significance of the regression coefficients by the t test was considered ($p \leq 0.05$). For this, the weight of each of the roots was considered as the response variable and the number of pixels as the explanatory variable. The regression analysis was also adjusted between the actual and predicted weight at the average level of the plot and at the level of average progeny, in order to verify the efficiency of these predictions.

For the genetic-statistical analysis of the real and predicted weights (at the root level), the $y = Xr + bCov + Za + e$ model was used, where: **y** is the data vector, **r** is the repetition effects vector (assumed as fixed) added to the general average; **a** is the vector of individual additive genetic effects (assumed to be random), **e** is the vector of errors or residuals (random). Coefficient *b* refers to the regression associated with the covariate **Cov** (number of plants per plot). Capitalized roman letters represent the incidence matrices for these purposes. Capital letters represent the incidence matrices for these purposes. From this model, the variance components were estimated and the breeding values were predicted by the REML-BLUP mixed model method with the aid of the *sommer* package (R Core Team, 2019). Subsequently, with the breeding values for the observed and predicted weight, the direct and indirect gains expected with the selection of the best progenies were estimated, as well as the Spearman correlation between the breeding values and the percentage of coincidence between the selected genotypes.

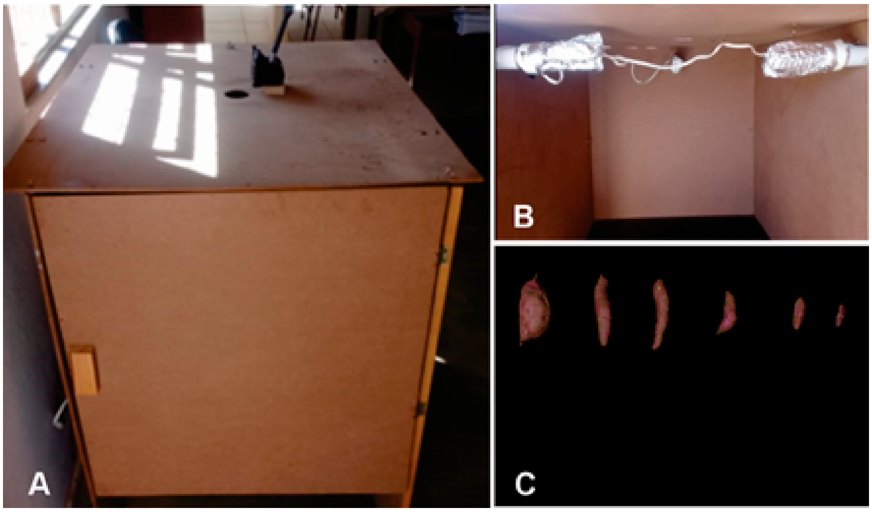

Figure 1. Image acquisition: A) "studio" for image acquisition; B) artificial lighting with fluorescent lamp; C) image acquired in the studio. Montes Claros, UFMG, 2022.

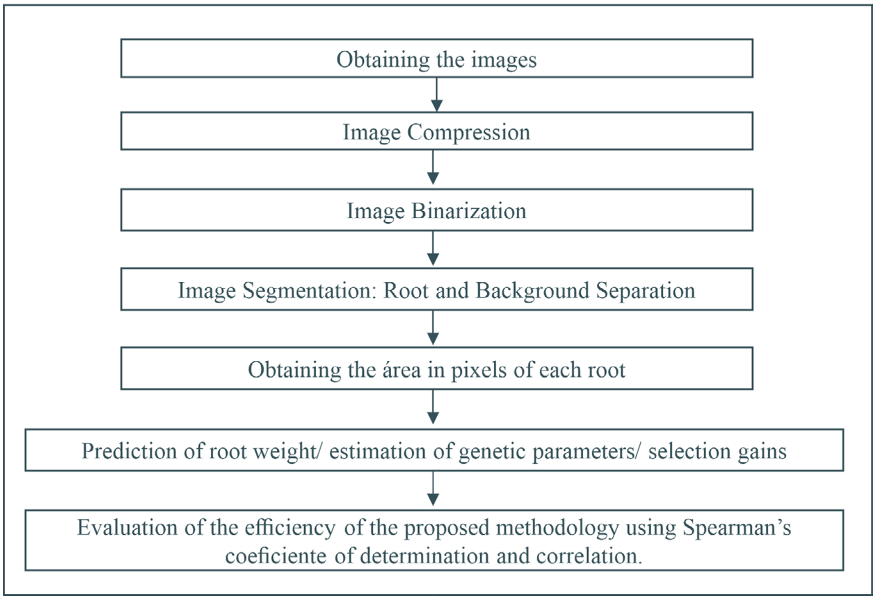

Figure 2. Flowchart of the steps performed after obtaining the images. Montes Claros, UFMG, 2022.

RESULTS AND DISCUSSION

The results indicate high efficiency of the methodology developed in predicting the productivity of roots from images. Since, through the number of pixels corresponding to the roots, estimates of root weight at plant level of 16 sweet potato half-sib progenies were obtained (Figure 3). A high coefficient of determination (R^2= 0.9187) was obtained between the predicted weight and the number of pixels. This ability to predict weight through image analysis assumes that an image is composed of a set of pixels that represent a matrix of (x,y) values (Pelt & Sethian, 2018). From this, the area occupied by a root refers to a specific number of pixels, making it possible to predict its weight. In this sense, it can be inferred that the larger the dimensions of the root, the greater its weight, corroborating what was observed in the field and reflecting the high value of the coefficient of determination obtained.

Analyzing the weight prediction at the plot (Figure 4A) and family (Figure 4B) level, high coefficients of determination were obtained between the predicted weight and the observed weight. Obtaining a higher coefficient of determination at the family level (R^2= 0.9944) (Figure 4B), since at the plot level the total variation of the data is relatively greater, confirming the coefficients of determination obtained through image analysis.

Estimating the genetic parameters for the observed weight and the predicted weight, similar results were observed between the two analyzed characteristics (Table 1), evidencing the efficiency of the sampling carried out in the field, as well as the use of images to predict the yield of sweet potato. This can be confirmed by the low variance values of the prediction error of the genotypic progeny values (PEV) observed for the two traits, as well as the good accuracy obtained for the attributes (ACprog). However, for the predicted weight characteristic, the PEV values (0.71), as well as the standard deviation of the predicted genotypic value (SEP=841.17) for each progeny were slightly lower compared to the observed weight (Table 1). This smaller variation for the predicted weights is due to the tendency of the data set to mitigate the effects of outliers in the prediction process.

Observing the estimates of additive genetic variance (Va), higher values were obtained for the observed weight variable (obs weight) (6.45) in relation to the predicted weight (pred weight), and this behavior was also observed for phenotypic variance (Vf) (7.61) (Table 1). In addition, the additive and phenotypic genetic variances were higher than the environmental variance (Ve) for the two evaluated traits, demonstrating that the studied families may be viable for breeding (Badu *et al*., 2017; Narasimhamurthy *et al*., 2018). The great variability among the studied sweet potato genotypes is highlighted by the superiority of the individual additive genetic coefficients of variation (CVgi= 63.31 and 60.12%) in relation to the residual coefficients of variation (CVe= 61.04 and 58.51%) for the two analyzed characteristics.

In addition to the coefficients of variance, heritability information (h^2a) is also of paramount importance for the efficient selection of parents (Rahajeng *et al*., 2020). Since heritability is the portion of genetic variance in relation to phenotypic variance, this parameter is used to estimate the reliability of the phenotypic value as an indicator of reproductive value (Ramalho *et al*., 2012). Thus, both for the observed weight in the field and for the predicted weight through image analysis, high values of h^2a were obtained for the studied progenies, which were similar for the two evaluated characteristics. This indicates that the phenotype is more influenced by genetic than environmental factors (Rahajeng *et al*., 2020). In this way, selection becomes more reliable by associating high heritabilities to CVg, since it indicates that the analyzed character has a greater additive effect.

The equivalence of the traditional technique with the selection via image analysis proved the high performance of the methodology to predict the root weights. This can be confirmed by the high correlation obtained between predicted and observed weights (r= 0.991**). In addition, the proximity between the genetic values is verified considering the predicted and observed weights for all families (Figure 5), so, when selecting the families there was 100% coincidence. Identifying the F26 (UFVJM25), F6 (UFVJM40), F10 (UFVJM05), F7 (UFVJM01) and

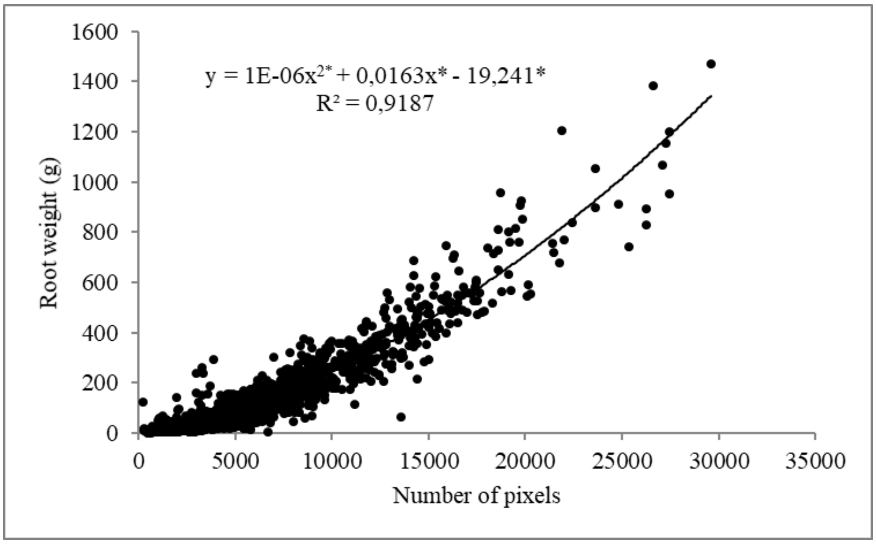

Figure 3. Adjusted regression in the prediction of root weight (g) from the number of pixels in images of 2233 roots from 16 progenies of sweet potato half-sibs. *significant at the 5% level by the T test. Montes Claros, UFMG, 2022.

Table 1. Estimates of genetic parameters in the study of observed and predicted weight in 16 sweet potato half-sib progenies. Montes Claros, UFMG, 2022.

Parameters	Va	Ve	Vf	h^2a	h^2mp	Acprog	h^2ad	CVgi%	CVgp%	CVe%	PEV	SEP
Observed weight	6.45	1.16	7.61	0.85	0.52	0.72	0.81	63	32	61	0.78	881.4
Predicted weight	5.82	1.15	6.97	0.84	0.51	0.72	0.79	60	30	59	0.71	841.2

Va= additive genetic variance; Ve= environmental variance between plots; Vf= phenotypic variance; h^2a= heritability; h^2mp= heritability of the average of families; ACprog= accuracy of family selection; h^2ad= additive heritability within a plot; CVgi=individual additive genetic variation coefficient; CVgp= coefficient of genetic variation between families; CVe= residual coefficient of variation; PEV= variance of the prediction error of progeny genotypic values; SEP= standard deviation of predicted genotypic progeny value.

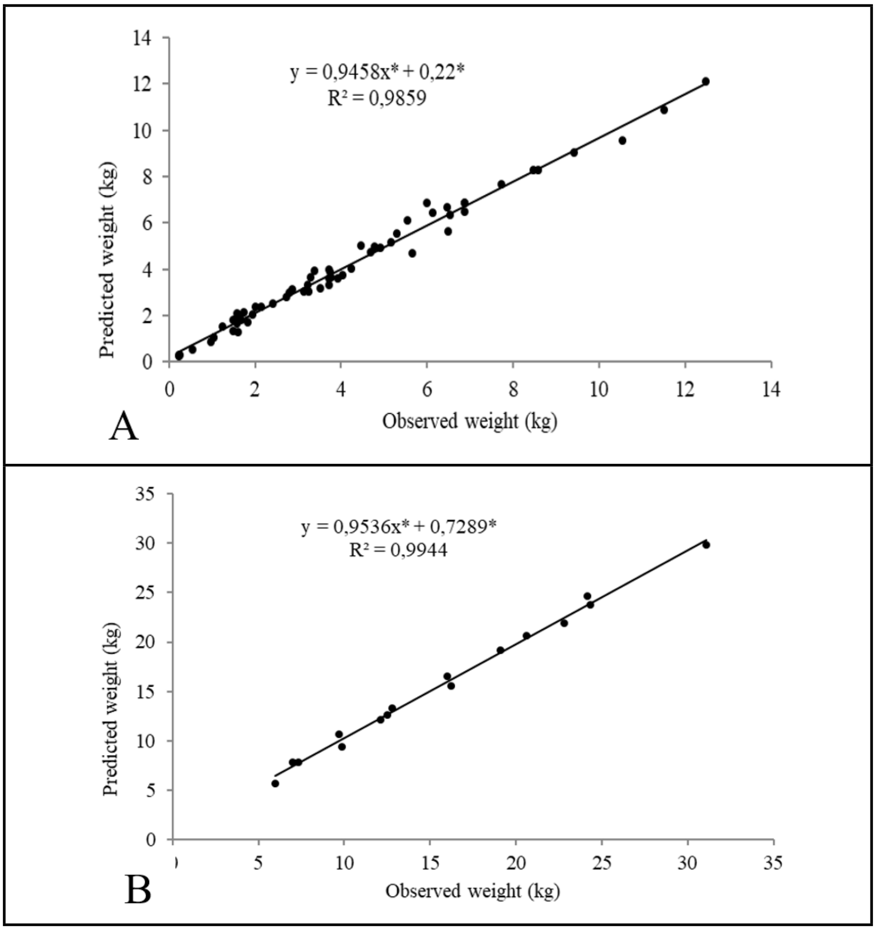

Figure 4. Graphical dispersion between predicted and observed total weight values at the plot (A) and family (B) level in 16 sweet-potato half-sib progenies. *significant at the 5% level by the T test. Montes Claros, UFMG, 2022.

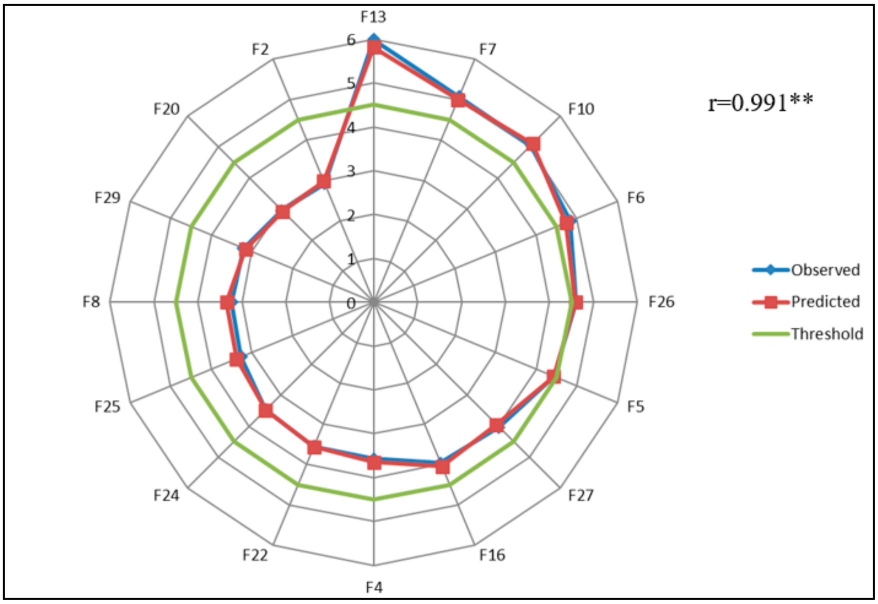

Figure 5. Radar scatter plot and Spearman correlation (r) of the additive breeding values of 16 sweet potato half-sib progenies. **significant at the 1% level by the T test. Montes Claros, UFMG, 2022.

F13 (UFVJM15) families that obtained additive values higher than expected, considering the 30% selection index, for the data set studied.

In addition to the possibility of identifying superior families, the methodology was also effective according to the direct and indirect selection gains for the two analyzed characteristics. As the selected families between the two methodologies were the same, the selection gains were identical, with estimates around 27.50 (obs weight) and 25.93% (pred weight) (Table 2). A relatively higher gain (1.6%) can be observed for the observed weight.

This characteristic represents an advantage for crop genetic improvement programs, since its main objective is to obtain roots within the appropriate weight for commercial purposes. Therefore, it can be inferred that the developed methodology has high potential for use in high-efficiency phenotyping through image analysis to predict the weight of sweet potato roots.

The increase in genomic information available in plant biology has led to the need for high-throughput phenotyping workflows to bridge the growing gap between genotype and phenotype (Pound *et al.*, 2017), since traditional phenotyping tools often rely on manual measurement of several selected traits from a small sample of plants. This fact can make the phenotyping process limited, delaying the knowledge of how phenotypes correlate with genetic and environmental factors, needing a more comprehensive analysis (Ubbens & Stavness, 2017). In this sense, the use of images has great potential to increase the yield and accuracy of phenotyping activities in plants (Pound *et al.*, 2017). Phenotyping through image analysis is already being used in tuberous roots, where it is possible to observe the evaluation of the broad genomic association of cassava roots and genomic prediction for image-extracted size and shape characteristics (Yonis *et al.*, 2020); as well as the analysis of size and shape in sweet potato roots (Haque *et al.*, 2021).

As for the prediction of sweet potato

Table 2. Direct (bold) and indirect selection gain in sweet potato half-sib progenies. Montes Claros, UFMG, 2022.

Selected variables	**Selection gain (%)**	
	Observed weight	**Predicted weight**
Observed weight	28	26
Predicted weight	28	26

root weight, there are few studies. In this sense, the study serves as a fundamental step, demonstrating that computer vision allows the prediction of root weight, maintaining the classification of genotypes and consequently the similarity between the expected gains with the selection. Thus, the use of images is an efficient tool for sweet potato genetic improvement programs, assisting in the crop phenotyping process. In addition, it is a first approach to automated large-scale phenotyping, allowing the creation of a database with production information extracted from the images.

ACKNOWLEDGMENTS

We thank CAPES (Council for the Improvement of Higher Education Personnel, code 001), FAPEMIG (Research Support Foundation of the State of Minas Gerais) and CNPq (National Council for Scientific and Technological Development) for supporting this study.

REFERENCES

AMABILE, RF; VILELA, MS; PEIXOTO, JR. 2018. Melhoramento de plantas: variabilidade genética, ferramentas e mercado. *Embrapa Cerrados*, *Livro técnico* (INFOTECA-E).

BADU, M; ASHOK, P; KIRANPATRO, TSKK; SASIKALA, K. 2017. Studies on genetic variability, heritability and genetic advance for growth, yield and quality parameters among orange flesh sweet-potato [*Ipomoea batatas* (L.) Lam.] genotypes. *International Journal of Current Microbiology and Appied Science* 6: 1894-1903.

BORGES, V; FERREIRA, PV; SOARES, L; SANTOS, GM; SANTOS, AMM. 2010. Seleção de clones de batata-doce pelo procedimento REML/BLUP. *Acta Scientiarum. Agronomy 32*: 643-649.

FILGUEIRA, FAR. 2008. *Novo manual de olericultura: agrotecnologia moderna na produção e comercialização de hortaliças*. Viçosa: UFV, 402p.

HAQUE, S; LOBATON, E; NELSON, N; YENCHO, GC; PECOTA, KV; MIEROP, R; WILLIAMS, CM. 2021. Abordagem de visão computacional para caracterizar fenótipos de tamanho e forma de culturas hortícolas usando imagens de alto rendimento. *Computadores e Eletrônicos na Agricultura*, 182: 106011. Available at https://doi.org/10.1016/j.compag.2021.106011.

HUANG, S; WEIGEL, D; BEACHY, RN; LI, J. 2016. A proposed regulatory framework for genome-edited crops. *Nature Genetics* 48: 109-111.

NARASIMHAMURTHY, PN; PATEL, NB; PATEL, AI; KOTESWARA, RAO, G. 2018. Genetic variability, heritability and genetic advance for growth, yield and quality parameters among sweet-potato [*Ipomoea batatas* (L.) lam.] genotypes. *International Journal of Chemical Studies* 6: 2410-2413.

OTOBONI, MEF; OLIVEIRA, DJLSFD; VARGAS, PF; PAVAN, BE; ANDRADE, MI. 2020. Genetic parameters and gain from selection in sweet-potato genotypes with high beta-carotene content. *Crop Breeding and Applied Biotechnology* 20:

PELT, DM; SETHIAN, JA. 2018. A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences* 115: 254-259.

POUND, MP; ATKINSON, JA; TOWNSEND, AJ; WILSON, MH; GRIFFITHS, M; JACKSON, AS; BULAT, A; TZIMIROPOULOS, G; WELLS, DM; MURCHIE, EH; PRIDMORE, TP; FRENCH, AP. 2017. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience* 6: 1-10.

R CORE TEAM. 2019. R: A language and environment for statistical computing. (2019). *R Foundation for Statistical Computing*, Vienna, Austria.

RAHAJENG, W; RESTUONO, J; INDRIANI, FC; PURWONO, P. 2020. Genetic parameters of agronomic traits in sweet-potato accessions. *Biosaintifika: Journal of Biology & Biology Education* 12: 240-246.

RAMALHO, MAP; ABREU, AFB; SANTOS, JB; NUNES, JAR. 2012. *Aplicações da genética quantitativa no melhoramento de plantas autógamas*. UFLA, Lavras, 522p (Ph.D. thesis).

RINCKER, K; NELSON, R; SPECHT, J; SLEPER, D; CARY, T; CIANZIO, SR; CASTEEL, S; CONLEY, S; CHEN, P; DAVIS, V; FOX, C; GRAEF, G; GODSEY, C; HOLSHOUSER, D; JIANG, GL; KANTARTZI, SK; KENWORTHY, W; LEE, C; MIAN, R; MCHALE, L; NAEVE, S; ORF, J; POYSA, V; SCHAPAUGH, W; SHANNON, G; UNIATOWSKI, R; WANG, D; DIERS, B. 2014. Genetic Improvement of U.S. soybean in maturity groups II, III, and IV. *Crop Science.* 54: 1419-1432. DOI: 10.2135/cropsci2013.10.0665.

UBBENS, JR; STAVNESS, I. 2017. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science* 8: 1190.

VARGAS, PF; GODOY, DRZ; ALMEIDA, LCF; CASTOLDI, R. 2017. Agronomic characterization of sweet-potato accessions. *Comunicata Scientiae* 8: 116-125.

YONIS, BO; PINO, DCD; WOLFE, M; JANNINK, JL; KULAKOW, P; RABBI, I. 2020. Improving root characterisation for genomic prediction in cassava. *Scientific Reports* 10 (8003). DOI: 10.1038/s41598-020-64963-9.

ZHOU, N; SIEGEL, ZD; ZARECOR, S; LEE, N; CAMPBELL, DA; ANDORF, CM; FRIEDBERG, I. 2018. Crowd sourcing image analysis for plant phenomics to generate ground truth data for machine learning. *PLoS computational biology* 14:e1006337.