

## DECISION TREE AS A TOOL IN THE CLASSIFICATION OF LIMA BEAN ACCESSIONS<sup>1</sup>

RAFAEL DA COSTA ALMEIDA<sup>2\*</sup>, WILSON VITORINO DE ASSUNÇÃO NETO<sup>2</sup>, VERÔNICA BRITO DA SILVA<sup>3</sup>,  
LEONARDO CASTELO BRANCO CARVALHO<sup>3</sup>, ÂNGELA CELIS DE ALMEIDA LOPES<sup>3</sup>,  
REGINA LUCIA FERREIRA GOMES<sup>3</sup>

**ABSTRACT** - Morpho-agronomic characterization studies aiming at the discrimination and classification of lima bean accessions in relation to the centers of domestication and biological status have been of great importance for conserving the biodiversity of this species. For this purpose, researchers have widely used the multivariate analysis called discriminant analysis, which is not always capable of producing satisfactory results. Computational intelligence-based classifiers are additional tools for understanding complex classification problems. In this study, the objective was to test the use of the decision tree in the classification of lima bean according to the centers of domestication and biological status (cultivated and wild), based on eight phenotypic traits of the seed. Sixty accessions of lima bean from the *Phaseolus* Germplasm Bank of Universidade Federal do Piauí (BGP / UFPI) were evaluated, and classification was performed using two approaches: conventional statistics with discriminant analysis of principal components (DAPC) and computational intelligence through decision tree (DT). The results showed that the use of DT was efficient to identify patterns in the classification of lima bean accessions, due to its comprehensibility. Seed weight was one of the main descriptors used to explain the origin and diversity of the species. The results found will be useful for studies that involve the conservation of genetic resources, mainly for the maintenance of germplasm banks and in breeding programs. In addition, it is recommended to integrate machine learning algorithms in studies aimed at classifying lima bean.

**Keywords:** *Phaseolus lunatus* L. Machine learning. Computational intelligence. Multivariate methods.

## ÁRVORE DE DECISÃO COMO FERRAMENTA NA CLASSIFICAÇÃO DE ACESSOS DE FEIJÃO-FAVA

**RESUMO** - Estudos de caracterização morfoagronômica que visam a discriminação e classificação de acessos de feijão-fava quanto aos centros de domesticação e estado biológico têm sido de grande importância para a conservação da biodiversidade da espécie. Para esse fim, é muito utilizada a análise multivariada denominada análise discriminante, que nem sempre é capaz de produzir resultados satisfatórios. Classificadores baseados em inteligência computacional constituem-se ferramentas adicionais para a compreensão de problemas complexos de classificação. Neste estudo, objetivou-se testar o uso da árvore de decisão na classificação do feijão-fava de acordo com os centros de domesticação e estado biológico (cultivado e silvestre), com base em oito caracteres fenotípicos da semente. Foram avaliados 60 acessos de feijão-fava do Banco de Germoplasma de *Phaseolus* da Universidade Federal do Piauí (BGP/UFPI), em cuja classificação foram utilizadas duas abordagens: estatística convencional com análise discriminante sob componentes principais (DAPC) e inteligência computacional por meio da árvore de decisão (AD). Os resultados mostraram que o uso da AD foi eficiente para identificar padrões na classificação dos acessos de feijão-fava, devido à sua compreensibilidade. O peso da semente foi um dos principais descritores utilizado para explicar a origem e a diversidade da espécie. Os resultados encontrados serão úteis para estudos que envolvem a conservação de recursos genéticos, principalmente para a manutenção de bancos de germoplasma e em programas de melhoramento. Além disso, recomenda-se a integração de algoritmos de aprendizado de máquina em estudos voltados à classificação de feijão-fava.

**Palavras-chave:** *Phaseolus lunatus* L. Aprendizado de máquina. Inteligência computacional. Métodos multivariados.

\*Corresponding author

<sup>1</sup>Received for publication in 03/31/2020; accepted in 09/03/2020.

Paper extracted from the doctoral thesis of the first author.

<sup>2</sup>Universidade Federal do Piauí, Teresina, PI, Brazil; rafa\_scoutt@hotmail.com – ORCID: 0000-0001-8067-546X, wilsonassuncao90@gmail.com – ORCID: 0000-0002-0499-6058.

<sup>3</sup>Department of Plant Science, Universidade Federal do Piauí, Teresina, PI, Brazil; verabritosl@hotmail.com – ORCID: 0000-0001-7739-5061, cbcleonardo@gmail.com – ORCID: 0000-0001-5722-9322, acalopes@ufpi.edu.br – ORCID: 0000-0002-9546-5403, rlfomes@ufpi.edu.br – ORCID: 0000-0002-7700-6959.

## INTRODUCTION

The species *Phaseolus lunatus* L., popularly referred to as lima beans, can be found in the form of two botanical varieties: *P. lunatus* var. *lunatus*, which includes domesticated populations, and *P. lunatus* var. *silvester*, made up of wild populations (BAUDET, 1977). Classification of this species is carried out according to its geographical origin and seed characteristics.

Studies conducted by Motta-Aldana et al. (2010), from the chloroplast DNA of 109 lima bean accessions, pointed to two major domestication centers: 1) Andean (A), located in Ecuador and northern Peru, in which the plants have large seeds and are adapted to high altitudes; and 2) Mesoamerican (M), which extends from central-western Mexico to Honduras, where plants have small seeds and occur at lower altitudes.

Discrimination of populations has been of great importance for biodiversity conservation and the development of breeding programs. Thus, analyses of genetic diversity, through phenotypic or genetic characteristics, have guided the choice of appropriate parents in breeding programs, leading to the optimization of selective gains, due to the variability found in the offspring of crosses between divergent groups (SANT'ANNA et al., 2018).

One of the techniques that allow allocating a new individual to one of the previously known distinct populations is the multivariate analysis called discriminant analysis (NOGUEIRA et al., 2008). However, biometric analyses are not always able to produce satisfactory results, mainly because the populations analyzed are not sufficiently divergent or the quantity and quality of the variables used in the study are not adequate.

In this context, conducting analyses through methods based on computational intelligence, which are able to go through stages of learning and generalization from all information, being tolerant to noise, represents a major advance for studies involving statistical procedures and for genetic improvement (SANT'ANNA et al., 2018).

Artificial intelligence has allowed a new approach in the decision-making process in several areas of agriculture (LIAKOS et al., 2018), with great potential in the conservation of genetic resources.

Among the techniques of artificial intelligence, decision tree (DT) stands out as they are widely used in various fields, mainly in machine learning, to solve complex classification problems.

Its success is adequately explained by the ability to provide simple representations, which are easily understandable by experts and even by ordinary users (TRABELSI; ELOUEDI; LEFEVRE, 2019).

Several studies show that the traits related to the seed are one of the main contributors to the understanding of genetic diversity and origin of the species (CHACÓN-SÁNCHEZ; MARTÍNEZ-CASTILLO, 2017; SILVA et al., 2017). Considering that DTs are attractive for practical applications due to their comprehensibility and that studies using machine learning with the species *P. lunatus* are still incipient, the aim of this study was to apply the decision tree in the classification of lima beans according to the centers of domestication and biological status (cultivated and wild) of the species, by means of phenotypic traits of the seed.

## MATERIAL AND METHODS

The 60 lima bean accessions used in this study belong to the *Phaseolus* Germplasm Bank of Universidade Federal do Piauí (BGP/UFPI), coming from exchanges with other germplasm banks in the United States (United States Department of Agriculture - USDA) and Colombia (Centro Internacional de Agricultura Tropical - CIAT).

Based on information from previous studies (MOTTA-ALDANA et al., 2010; MONTERO-ROJAS et al., 2013; CHACÓN-SÁNCHEZ; MARTÍNEZ-CASTILLO, 2017), the selected accessions were classified according to the domestication centers (Andean - A and Mesoamerican - M), biological status (cultivated - C and wild - W) and geographic origin (Table 1), thus forming four groups: 33 cultivated Mesoamerican accessions (CM); 12 wild Mesoamerican accessions (WM); 06 cultivated Andean accessions (CA); and 09 wild Andean accessions (WA).

Eight quantitative descriptors were measured, as recommended by the International Plant Genetic Resources Institute (IPGRI, 2001), and their respective measurements, within parentheses, were: 1) seed length (SL, mm); 2) seed width (SW, mm); 3) seed weight (100SW, g); 4) seed thickness (ST, mm); 5) seed area (SAR, mm<sup>2</sup>); 6) seed length perimeter (SLP, mm); 7) seed length to width ratio (L/W); 8) seed thickness to width ratio (T/W). The measurements were taken with a digital caliper and using *Smartgrain* software (TANABATA et al., 2012; ASSUNÇÃO-NETO et al., 2018).

**Table 1.** List of 60 lima bean accessions from the *Phaseolus* Germplasm Bank of Universidade Federal do Piauí (BGP/UFPI), characterized in Teresina, PI, Brazil, 2018.

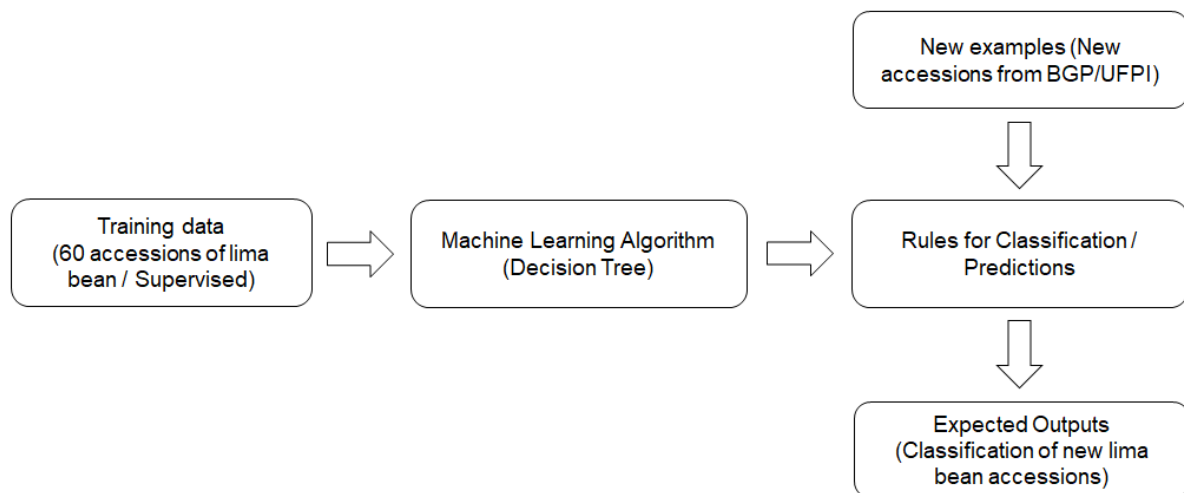
Accession	Biological Status	Country	Classification
UFPI 1029	Wild	Costa Rica	WM
UFPI 1030	Wild	Mexico	WM
UFPI 1031	Wild	Mexico	WM
UFPI 1033	Cultivated	Guatemala	CM
UFPI 1039	Cultivated	Argentina	CM
UFPI 1040	Wild	Mexico	WM
UFPI 1044	Wild	Costa Rica	WM
UFPI 1045	Wild	Ecuador	WA
UFPI 1046	Cultivated	Trinidad and Tobago	CM
UFPI 1047	Wild	Colombia	WA
UFPI 1048	Cultivated	Mexico	CM
UFPI 1050	Wild	Mexico	WM
UFPI 1051	Cultivated	Peru	CA
UFPI 1053	Wild	Colombia	WA
UFPI 1054	Wild	Colombia	WA
UFPI 1055	Wild	Ecuador	WA
UFPI 1057	Wild	Honduras	WA
UFPI 1058	Wild	Colombia	WA
UFPI 1061	Cultivated	Colombia	CM
UFPI 1065	Cultivated	Dominican Republic	CM
UFPI 1066	Wild	El Salvador	WM
UFPI 1067	Wild	Mexico	WM
UFPI 1071	Wild	Peru	WA
UFPI 1078	Cultivated	Argentina	CM
UFPI 1081	Cultivated	Mexico	CM
UFPI 1082	Cultivated	Mexico	CM
UFPI 1083	Wild	Guatemala	WM
UFPI 1084	Cultivated	El Salvador	CM
UFPI 1085	Cultivated	El Salvador	CM
UFPI 1087	Cultivated	Guatemala	CM
UFPI 1089	Cultivated	Mexico	CM
UFPI 1091	Cultivated	Cuba	CM
UFPI 1092	Cultivated	Mexico	CM
UFPI 1095	Cultivated	Peru	CA
UFPI 1096	Cultivated	Colombia	CM
UFPI 1098	Cultivated	Peru	CA
UFPI 1099	Cultivated	Colombia	CM
UFPI 1100	Cultivated	Colombia	CM
UFPI 1101	Cultivated	Costa Rica	CM
UFPI 1102	Cultivated	Argentina	CM
UFPI 1103	Cultivated	Argentina	CM
UFPI 1104	Wild	Peru	WA
UFPI 1105	Cultivated	Peru	CA
UFPI 1107	Cultivated	Bolivia	CA
UFPI 1114	Cultivated	Guatemala	CM
UFPI 1115	Cultivated	Guatemala	CM
UFPI 1116	Cultivated	Guatemala	CM
UFPI 1117	Cultivated	Mexico	CM
UFPI 1118	Cultivated	Guatemala	CM
UFPI 1119	Cultivated	Costa Rica	CM
UFPI 1124	Cultivated	Ecuador	CA
UFPI 1127	Cultivated	Mexico	CM
UFPI 1128	Cultivated	Mexico	CM
UFPI 1130	Cultivated	Mexico	CM
UFPI 1132	Cultivated	Mexico	CM
UFPI 1133	Cultivated	Mexico	CM
UFPI 1138	Wild	Guatemala	WM
UFPI 1143	Wild	Mexico	WM
UFPI 1144	Wild	Mexico	WM
UFPI 1150	Cultivated	Argentina	CM

Legend: CM: cultivated Mesoamerican; WM: wild Mesoamerican; CA: cultivated Andean; WA: wild Andean.

Discriminant analysis of principal components (DAPC) and computational intelligence through the decision tree (DT) were used for classifying the accessions.

Classification of accessions was performed based on DAPC and on the probability of association, using the a priori information of the country of origin and biological status of the accession (Table 1). For the application of the decision tree (DT) algorithm, the response variable

was the classification presented in Table 1 and the predictor variables were the measured phenotypic traits of the seed (Figure 1), the division index used was Gini (BREIMAN et al., 1984). Data analyses were performed with the R program (R DEVELOPMENT CORE TEAM, 2018), through the following packages: *adeigenet* (JOMBART; AHMED, 2011), *rpart* (THERNEAU; ATKINSON, 2019) and *rpart.plot* (MILBORROW, 2019).



**Figure 1.** Schematic representation of the application of the decision tree in the classification of 60 lima bean accessions.

## RESULTS AND DISCUSSION

The principal component analysis, performed from the means of the eight quantitative traits evaluated, provided the data of the eigenvalues

(variance) of each principal component, the percentage of explained variance and accumulated proportion (Table 2). The first two principal components (PCs) were sufficient to explain 91.34% of the total variation contained in the set of phenotypic traits of the seed.

**Table 2.** Estimates of the main components, eigenvalues, percentage of explained variance and cumulative proportion (%) for eight quantitative descriptors evaluated in 60 lima bean accessions from the *Phaseolus* Germplasm Bank of Universidade Federal do Piauí (BGP/UFPI), characterized in Teresina, PI, Brazil, 2018.

Principal component	Variance (Eigenvalues)	Variance (%)	Accumulated variance (%)
PC1	2.42	73.04	73.04
PC2	1.21	18.30	91.34
PC3	0.77	7.45	98.79
PC4	0.29	1.04	99.83
PC5	0.09	0.10	99.93
PC6	0.05	0.04	99.97
PC7	0.04	0.02	99.99
PC8	0.03	0.01	100.00

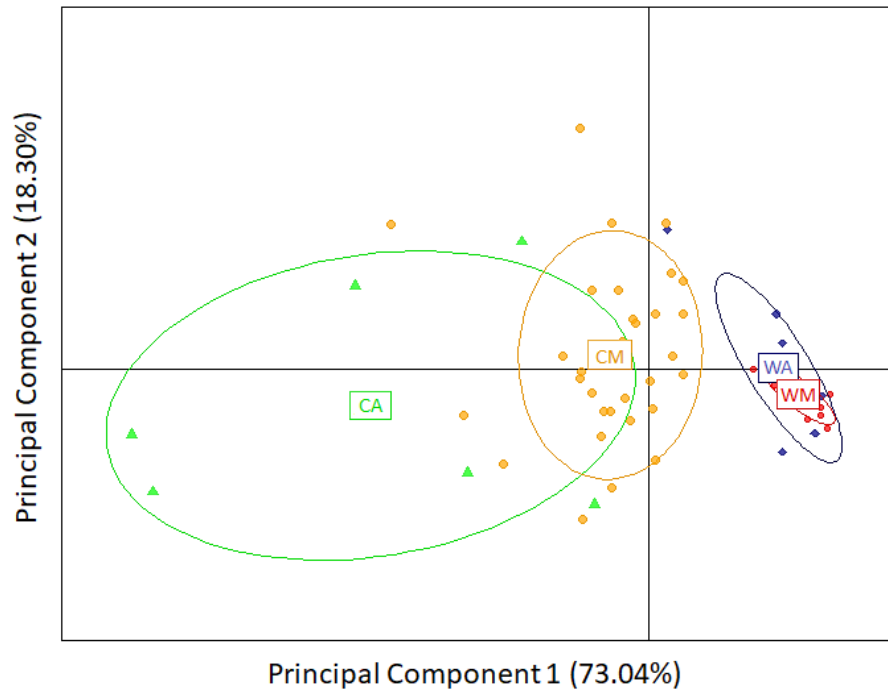
The discriminant analysis of principal components (DAPC) seeks to obtain functions that make it possible to classify an accession, from information of a set of measured characteristics, one among several known groups, seeking to minimize the probability of poor classification (CRUZ; REGAZZI; CARNEIRO, 2014).

The scatter plot (Figure 2) that corresponds to the grouping based on the discriminant analysis of principal components (DAPC), plotted from the first two principal components, used information of the country of origin and biological status of the accession (Table 1). Each color symbolizes the groups obtained a priori and, therefore, provides the

possibility of verifying the consistency of this grouping in relation to DAPC.

Graphic analysis (Figure 2) showed the dispersion of lima bean accessions and the separation according to the biological status (cultivated and

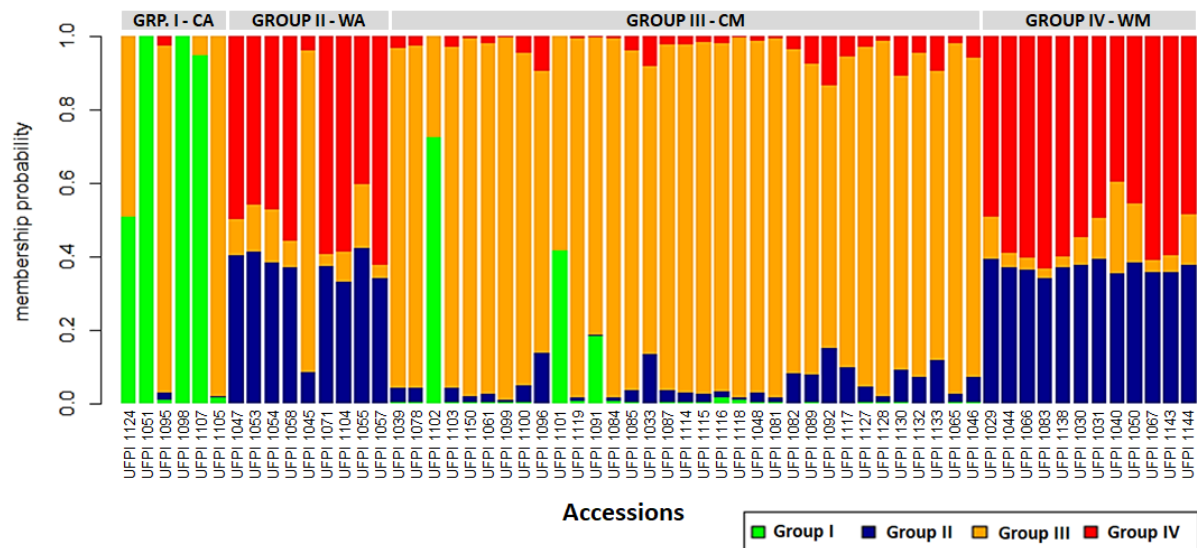
wild). It was also possible to observe that there is a lot of overlap between the wild groups (WA and WM), showing a certain proximity between the groups, which refers to a certain degree of difficulty in pointing out a difference between them.



**Figure 2.** Scatter plot of the 60 lima bean accessions from BGP/UFPI in the discriminant analysis of principal components (DAPC). The graph represents the accessions as geometric shapes and groups as ellipses. CA: Cultivated Andean; WA: Wild Andean; CM: Cultivated Mesoamerican; WM: Wild Mesoamerican.

The graph of the association probability (Figure 3), that is, the probability of an accession belonging to the group to which it was classified based on country of origin and biological status

(Table 1), shows the presence of accessions with characteristics of other groups, functioning as an indicator of consistency and may even show diversity within accessions.



**Figure 3.** Probability of an accession belonging to the group to which it was associated based on country of origin and biological status. Colors represent the four groups. CA: Cultivated Andean; WA: Wild Andean; CM: Cultivated Mesoamerican; WM: Wild Mesoamerican.

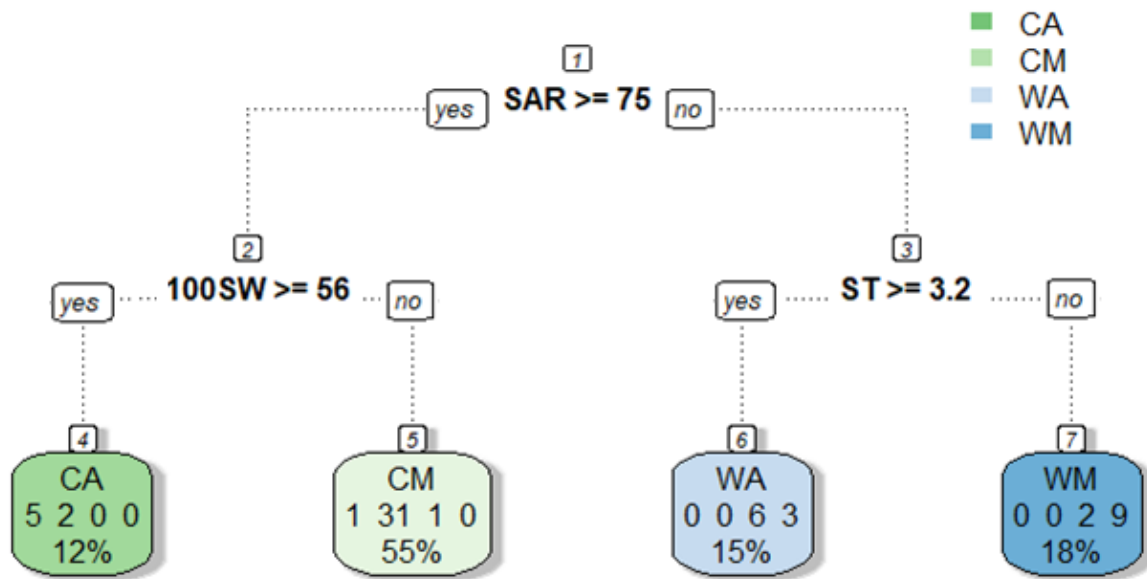
Group I, cultivated Andean, had high values for the probability of association, except for the accessions UFPI 1124, UFPI 1095 and UFPI 1105, which had low values for the traits related to the seed, indicating characteristics associated with the Mesoamerican domestication center. Some possible explanations for these exceptions within each group are: the natural hybridization of wild and domesticated forms, which are observed throughout Latin America, generating intermediate forms (BAUDOIN et al., 2004), or the phenotypic plasticity of the crop.

Group III, cultivated Mesoamerican, also showed a high probability of association with values higher than 90%, except for the accession UFPI 1102. The contrast of this accession in group III is mainly due to the high values for the traits related to the seed, except for thickness. This causes the accession to be highly associated with group I, cultivated Andean, since larger and heavy seeds identify the Andean domestication center (MOTTA-ALDANA et al., 2010).

Groups II and IV showed a lot of mixture, because they are of the wild biological status, with very similar phenotypic values related to the seed,

which made differentiation difficult. Some possible explanations for the efficiency of discrimination have not been so high are: I) the groups are not in fact so differentiated and consequently there will be overlap; II) the quantity and quality of the variables are not sufficient to differentiate the groups; III) the statistical approach used for discriminating the accessions is not adequate.

The decision tree (DT) (Figure 4) was used as a tool for better dissection of the classification of lima bean accessions regarding domestication centers and biological status of the crop, based on phenotypic traits of the seed. According to the results, the DT demonstrated that, if the seed has an area larger than or equal to 75 mm<sup>2</sup>, it is classified in the cultivated biological status, otherwise it is classified as wild. When the seed classified as of cultivated biological status has 100SW higher than or equal to 56 g, it is classified in the cultivated Andean (CA) group of seeds and, if 100SW is lower, it is classified as cultivated Mesoamerican (CM). Otherwise, when the seed has an area smaller than 75 mm<sup>2</sup> and has a thickness greater than or equal to 3.2 mm, it is classified as WA and, if it is smaller, the seed is classified as WM.



**Figure 4.** Decision tree (DT) based on the Gini index (BREIMAN et al., 1984) for the training data set of the 60 lima bean accessions from BGP/UFPI. Leaf nodes can assume the following labels: CA: Cultivated Andean; WA: Wild Andean; CM: Cultivated Mesoamerican; WM: Wild Mesoamerican. Acronyms of the descriptors: Seed area (SAR, mm<sup>2</sup>); Seed weight (100SW, g); seed thickness (ST, mm).

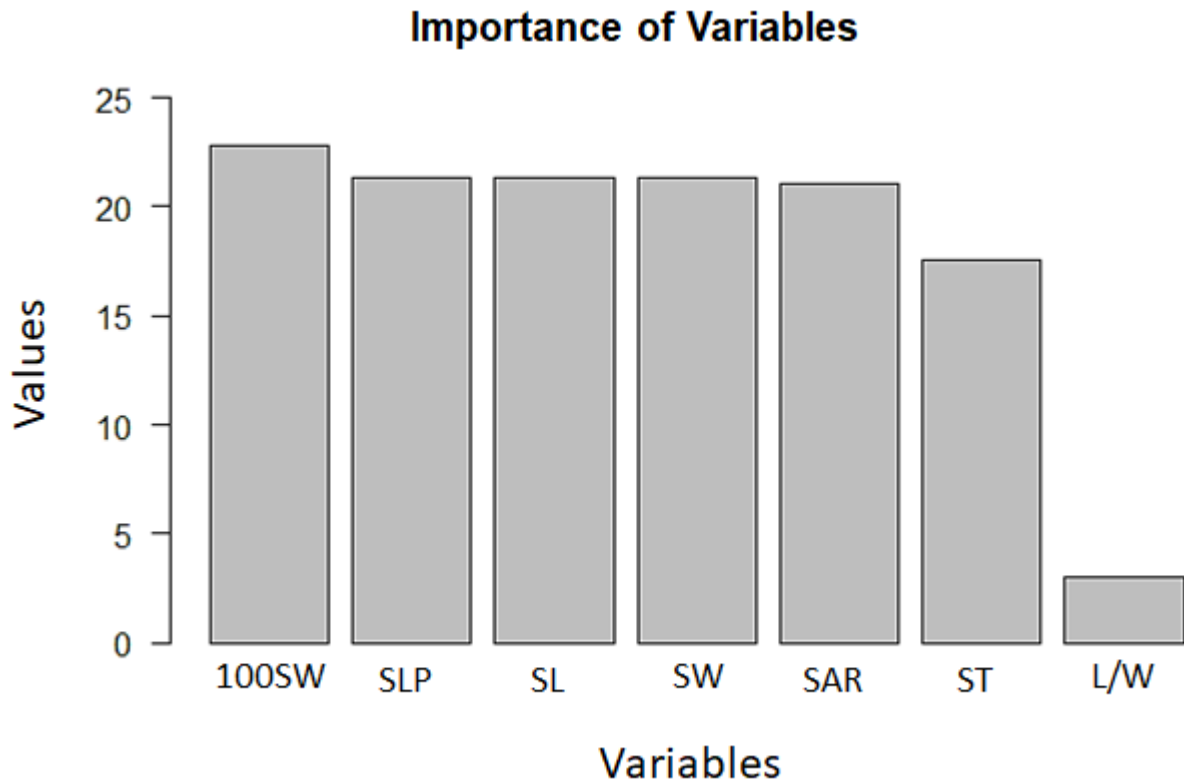
According to the decision tree, the most important variable for the classification of a new lima bean accession with respect to the domestication center and biological status was seed weight (100SW) (Figure 5). Seeds classified as of cultivated biological status and with 100SW greater than or equal to 56 g are classified in the group of cultivated Andean seeds, while lighter seeds are

related to the Mesoamerican domestication center (Figure 4).

The traits seed length to width ratio (L/W) and seed thickness to width ratio (T/W) contributed little and can be discarded in new studies on classification of this species (Figure 5). The results obtained corroborated the literature, demonstrating that 100SW is one of the main descriptors used to

explain the origin and diversity of lima beans (GUTIÉRREZ-SALGADO; GEPTS; DEBOUCK, 1995; MOTTA-ALDANA et al., 2010; CHACÓN-SÁNCHEZ; MARTÍNEZ-CASTILLO, 2017). These

studies found that larger seeds (weight ranging from 58 to 122 g / 100 seeds) are classified in the Andean domestication center.



**Figure 5.** Importance of the phenotypic variables of the seed used in the decision tree (DT) for the classification of a new lima bean accession. 100SW - seed weight; SAR - seed area; SLP - seed length perimeter; SL - seed length; SW - seed width; ST - seed thickness; L/W - seed length to width ratio.

The study conducted by Silva et al. (2017), with the objective of characterizing 166 lima bean accessions cultivated in Brazil, using the Ward – MLM (Modified Location Model), in order to analyze the organization of genetic diversity and origin of this species, detected high genetic variability. The traits seed length and width were the main contributors to the genetic divergence between the evaluated accessions. The traits SL and SW were also important in the construction of the DT. The confusion matrix was obtained to evaluate the accuracy of the DT, and it demonstrated an 85% accuracy efficiency.

By evaluating the use of discriminant analysis of principal components (DAPC) and the decision tree (DT) in the lima bean classification study, it can be seen that the use of DT enabled a greater capacity to understand the logic behind the differentiation between the accessions. This methodology was also efficient to classify wild lima bean accessions. According to Piltaver et al. (2016), the decision tree is one of the most understandable machine learning

techniques to solve complex classification problems.

This approach has considerable potential for studies involving the conservation of genetic resources, mainly for the maintenance of germplasm banks of the species and in breeding programs of the crop. The explanation of the differentiation between groups is a subject of great relevance in the conservation of genetic resources, in the identification of duplicates, and in genetic improvement, when the objective is to establish heterotic groups and identify hybrid combinations of higher vigor.

## CONCLUSIONS

Artificial intelligence-based methods are important alternatives in studies aimed at discrimination of populations.

The use of decision tree in the classification of lima beans in the domestication centers and biological status (cultivated and wild) of the crop,

based on phenotypic traits of the seed, proved to be efficient.

Seed weight is a key character that can be used to clarify the origin and diversity of lima beans, considering that seeds classified as of cultivated biological status and having 100SW greater than or equal to 56 g belong to the Andean domestication center, while lighter seeds are related to the Mesoamerican center.

## REFERENCES

- ASSUNÇÃO-NETO, W. V. et al. Genetic diversity in *Phaseolus lunatus* L. based on morphological seed characters. **Annual Report of the Bean Improvement Cooperative**, 61: 199-200, 2018.
- BAUDET, J. C. The taxonomic status of the cultivated types of lima bean (*Phaseolus lunatus* L.). **Tropical Grain Legume**, 7: 29-30, 1977.
- BAUDOIN, J. P. et al. Ecogeography, demography, diversity and conservation of *Phaseolus lunatus* L. in the Central Valley of Costa Rica. **Systematic and Ecogeographic Studies on Crop Gene Pools 12**, Rome, Italy: International Plant Genetic Resources Institute, 2004. 85 p.
- BREIMAN, L. et al. **Classification and regression trees**. 1. ed. Belmont, CA: WADSWORTH, 1984. 358 p.
- CHACÓN-SÁNCHEZ, M. I.; MARTÍNEZ-CASTILLO, J. Testing Domestication Scenarios of Lima Bean (*Phaseolus lunatus* L.) in Mesoamerica: Insights from Genome-Wide Genetic Markers. **Frontiers in Plant Science**, 8: 1-20, 2017.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. vol. 2. 3 ed. Viçosa, MG: UFV, 2014. 668 p.
- GUTIÉRREZ-SALGADO, A.; GEPTS, P.; DEBOUCK, D. G. Evidence for two gene pools of the Lima bean, *Phaseolus lunatus* L., in the Americas. **Genetic Resources and Crop Evolution**, 42: 15-28, 1995.
- IPGRI. Descritores para *Phaseolus lunatus* (feijão-espadinho). **International Plant Genetic Resources Institute**, Bioversity Technical Bulletin Series. Rome: Bioversity International. 2001. 51 p.
- JOMBART, T.; AHMED, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. **Bioinformatics**, 27: 3070-3071, 2011.
- LIAKOS, K. G. et al. Machine learning in agriculture: a review. **Sensors**, 18: 2674-2703, 2018.
- MILBORROW, S. **rpart.plot: plot “rpart” models: an enhanced version of “plot.rpart”**, R package version 3.0.8. 2019. Disponível em: <<https://cran.r-project.org/package=rpart.plot>>. Acesso em: 04 dez. 2019.
- MONTERO-ROJAS, M. et al. Genetic, morphological and cyanogen content evaluation of a new collection of Caribbean Lima bean (*Phaseolus lunatus* L.) landraces. **Genetic Resources and Crop Evolution**, 60: 2241-2252, 2013.
- MOTTA-ALDANA, J. R. et al. Multiple Origins of Lima Bean Landraces in the Americas: Evidence from Chloroplast and Nuclear DNA Polymorphisms. **Crop Science**, 50: 1773-1787, 2010.
- NOGUEIRA, A. P. O. et al. Novas características para diferenciação de cultivares de soja pela análise discriminante. **Ciencia Rural**, 38: 2427-2433, 2008.
- PILTAVER, R. et al. What makes classification trees comprehensible? **Expert Systems With Applications**, 62: 333-346, 2016.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2018.
- SANT’ANNA, I. C. et al. RNA - Aplicações em estudos classificatórios. In: CRUZ, Cosme Damião; NASCIMENTO, Moysés. (Eds.). **Inteligência computacional aplicada ao melhoramento genético**. 1. ed. Viçosa: UFV, 2018. cap. 7. p. 189-214.
- SILVA, R. N. O. et al. Phenotypic diversity in lima bean landraces cultivated in Brazil, using the Ward-MLM strategy. **Chilean Journal of Agricultural Research**, 77: 35-40, 2017.
- TANABATA, T. et al. SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. **Plant Physiology**, 160: 1871-1880, 2012.
- THERNEAU, T.; ATKINSON, B. **rpart: Recursive partitioning and regression trees**, R package version 4.1-15. 2019. Disponível em: <<https://cran.r-project.org/package=rpart>>. Acesso em: 04 dez. 2019.
- TRABELSI, A.; ELOUEDI, Z.; LEFEVRE, E. Decision tree classifiers for evidential attribute values and class labels. **Fuzzy Sets and Systems**, 366: 46-62, 2019.