ISSN 2007-9737

Discovering Diagnostic Features Used by a CNN in Plant Species Identification

Geovanni Figueroa-Mata[1,*], Erick Mata-Montero[2], Luis Acosta-Vargas[3]

[1] Costa Rica Institute of Technology, School of Mathematics, Cartago,
Costa Rica

[2] Costa Rica Institute of Technology, School of Computing, Cartago,
Costa Rica

[3] Costa Rica Institute of Technology, School of Forestry Engineering, Cartago,
Costa Rica

{gfigueroa, lacosta}@tec.ac.cr, erick_mata@yahoo.com

Abstract. An approach to improve the explainability (interpretability) of convolutional neural networks that identify plant species from leaf images is proposed. Specifically, a methodology is established to discover the most determining diagnostic features used by a convolutional neural network (CNN) in the identification of 63 native plant species from Costa Rica. The result is a CNN that not only identifies plant species but also provides an explanation through a heat map and a translation of that map into a table of diagnostic features used in classical taxonomy, each with a weight that describes the relative importance of each trait (e.g., apex, primary vein, and leaf base). To achieve this, a CNN was trained using leaf images from 63 vascular plant species from Costa Rica. Once the network was trained, the Layer-wise Relevance Propagation (LRP) technique was applied to a subset I of 50 leaves images distributed uniformly across a set of 10 species to visualize the representations (heat maps) learned by the internal layers of the CNN. Then, a taxonomist was asked to perform an equivalent task manually, annotating the same 50 leaf images in I by graphically highlighting the most significant features according to their expert judgment (feature map). Finally, algorithmic comparisons were made between the heat maps and feature maps to determine the similarity between the hottest areas used by the CNN and the features used in classical taxonomy.

Keywords. Convolutional neural network, heat map, layer-wise relevance propagation, deep learning, interpretability, automated plant species identification.

1 Introduction

Motivated by the fact that deep learning algorithms have poor explanatory power, this research takes on the challenge of discovering the most determining features used by a Convolutional Neural Network (CNN) in the identification of plant species. In general, this is a formidable challenge that, as far as we know, has not been addressed before, even for small sets of species. Consequently, we limit the scope of this exploratory work to a small subset of the estimated 12,000 native vascular plant species of Costa Rica, namely, the 255 species represented in the CRLeaves dataset [13]. Having explanations that justify the responses of such deep learning algorithms is important for:

1. Demonstrating that the approach is robust from a taxonomic perspective and therefore more reliable.
2. Improving dichotomous or polytomous identification keys used in traditional taxonomic work, in case the CNN detects features not being used by experts.
3. Enhancing the CNN in case very discriminative features used in traditional taxonomy are detected but not used by the CNN.

doi: 10.13053/CyS-28-4-4720

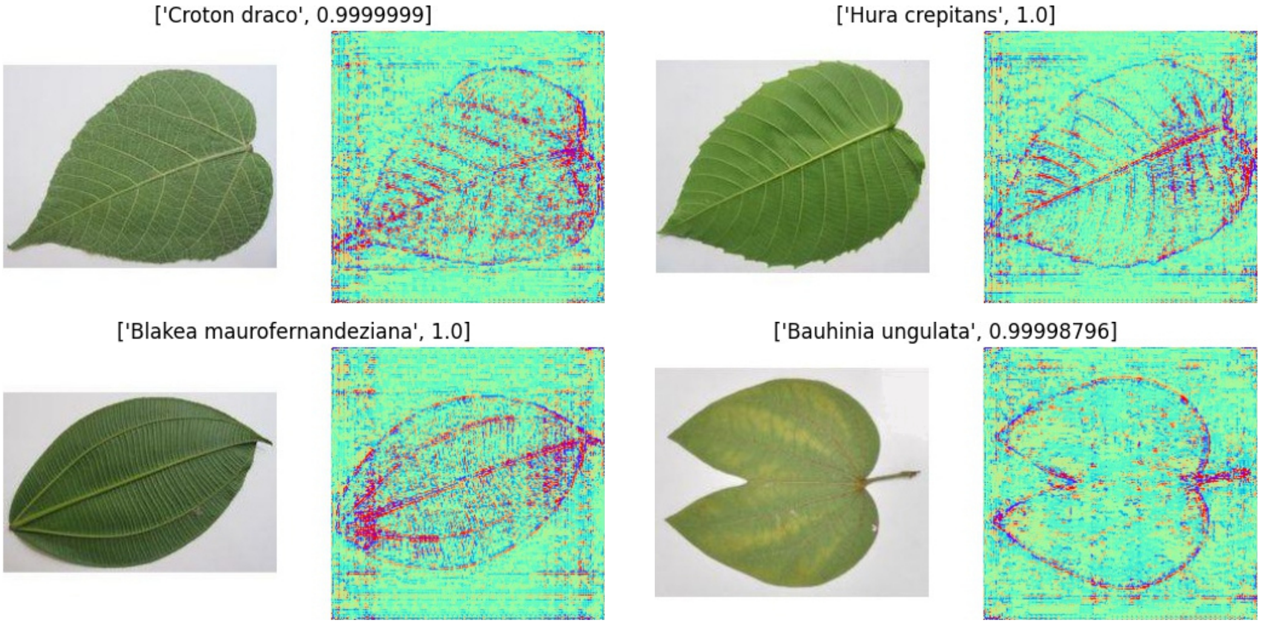

['Croton draco', 0.9999999]
['Hura crepitans', 1.0]
['Blakea maurofernandeziana', 1.0]
['Bauhinia ungulata', 0.99998796]

Fig. 1. LRP for some leaves of Costa Rican plant species

In this research, the CRLeaves dataset was used to train a CNN capable of identifying plant species in Costa Rica from images of their leaves. Then, the Layer-wise Relevance Propagation (LRP) technique was applied to a subset I of leaves images, to visualize the representations (heat maps) learned by the internal layers of the CNN. Subsequently, the heat maps of specimens of each species and a formally defined measure of intra-specific variability were used to asses how similar the “hot regions” in those heat maps were.

Finally, and more importantly, the heat maps were compared to the identification criteria used by an expert in the identification process for the same subset I of leaves images. The rest of this article is organized as follows:

Section 2 presents a summary of the background regarding the visualization of relevant regions used by deep learning algorithms in decision-making and the interpretability of a model based on a CNN. Section 3 summarizes the methodology used. Section 4 presents the results achieved and Section 5 summarizes the conclusions and provides some recommendations for future research.

2 Background

2.1 Plant Identification

Species conservation is closely related to their correct identification [9, 20]. It is estimated that there are around 400,000 species of vascular plants in the world. This enormous biodiversity and its vast intra and interspecific variability make their identification a very complex task even for experts.

Typically, the task of plant identification involves assigning a specific taxon to an individual plant specimen based on the similarity of its discriminative morphological characters (diagnostic characters) to those of a particular species. A human expert (taxonomist) uses these characters to visually identify the species of a particular plant.

The ultimate goal is to assign a plant species to a particular specimen. Due to the inherent complexity of the identification task, it has been approached computationally, first with interactive applications based on identification keys, then with software based on machine learning techniques [1, 2, 10], and more recently by applying deep learning techniques [7, 8, 11, 12, 21].

Table 1. Explanation vectors obtained from CNN

Specimen	Margin	Apex	Main vein	Base	Complement	Secondary veins
`Croton draco_1_2_2`	0.41	0.40	0.41	0.42	0.41	0.40
`Guazuma ulmifolia_1_3_2`	0.38	0.41	0.43	0.35	0.42	0.45
`Hura crepitans_1_2_2`	0.41	0.39	0.40	0.37	0.43	0.44
`Hymenaea courbaril_2_1_2`	0.47	0.51	0.38	0.42	0.45	0

Table 2. Explanation vectors assigned by the expert

Specimen	Margin	Apex	Main vein	Base	Complement	Secondary veins
`Croton draco_1_2_2`	0.22	0.33	0.55	0.65	0.11	0.33
`Guazuma ulmifolia_1_3_2`	0.23	0.18	0.35	0.88	0.035	0.088
`Hura crepitans_1_2_2`	0.45	0.23	0.23	0.57	0.23	0.57
`Hymenaea courbaril_2_1_2`	0.68	0.39	0.19	0.58	0.097	0

2.2 Interpretability of a Model based on a CNN

The main challenge in the field of eXplainable AI (XAI) is explaining the decisions of an intelligent system. This is referred to as the problem of interpretability or explainability. In the context of Machine Learning (ML), interpretability is defined as the ability to explain results in understandable terms to a human [6]. Miller [14] defines interpretability as the degree to which a human can understand the cause of a decision. Therefore, we could say that an ML model is interpretable if a human can understand its operations, either through introspection or through a produced explanation [4].

In general, two types of models can be distinguished: those that are interpretable by design (e.g., decision trees, Bayesian models, k-nearest neighbors, and rule-based learning models) and those that can be explained through an external XAI technique (post-hoc explainability technique). The latter are aimed at models that are not easily interpretable, as is the case of CNNs. Some examples of external XAI techniques are text explanations, visualizations, local explanations, and example-based explanations [16].

Visual explanation techniques add explainability to a model by highlighting which input variables (in this case, pixels) have contributed to classifying the given image by the CNN.

This scoring of pixels can be visualized as a heat map that overlays the original image to highlight the regions of the image that were relevant in the prediction [17].

2.3 Visualizing the Decisions of a CNN

Visual explanation techniques aim to produce class activation maps for input images. A class activation map is a 2D grid of scores associated with a specific output class and calculated for each pixel in the input image, indicating the importance of each pixel with respect to the considered class.

Some of the developed techniques include: Class Activation Map (CAM) [18, 23], Saliency Map (SM) [19, 22]) and more recently, Layer-Wise Relevance (LRP). Layer-wise Relevance Propagation (LRP) was originally described by Bach et al. [3]. The idea is to calculate the contribution of each pixel x_d in an input image x to the prediction $f(x)$ made by a classifier f in an image classification task.

Here classifier $f : \mathbb{R}^v \to \mathbb{R}$ has real valued outputs which are thresholded at zero. For this purpose, a relevance measure (R_d) is defined for each pixel x_d in the input image x, such that the prediction $f(x)$ is expressed as the sum of the values of R_d:

$$f(x) \approx \sum_{d=1}^{v} R_d, \tag{1}$$

doi: 10.13053/CyS-28-4-4720

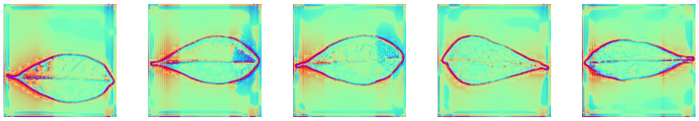
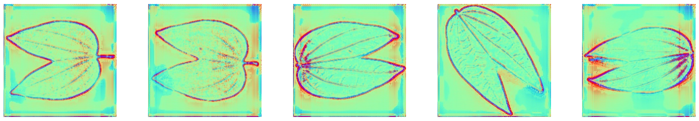
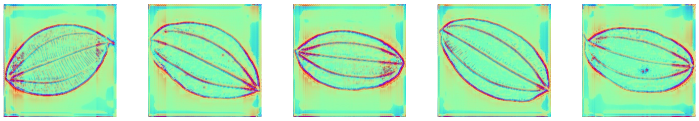
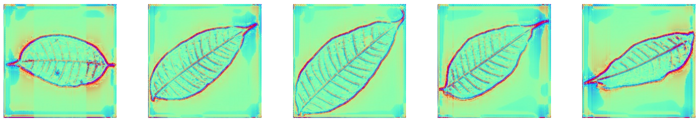
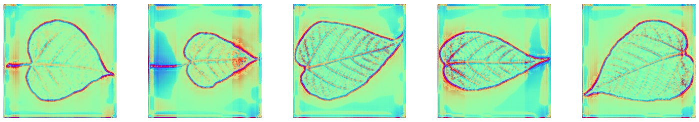
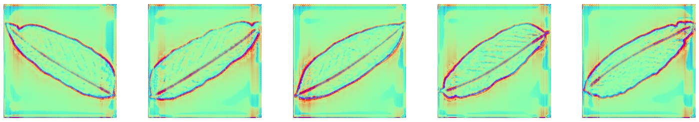
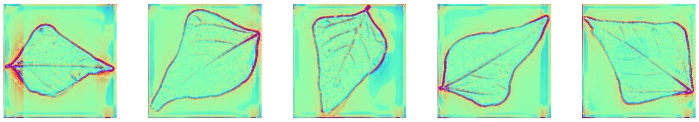
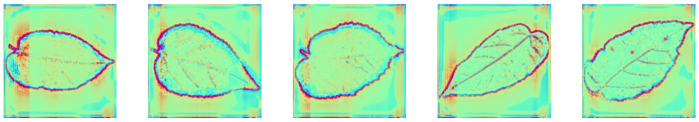
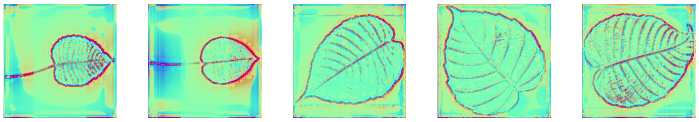
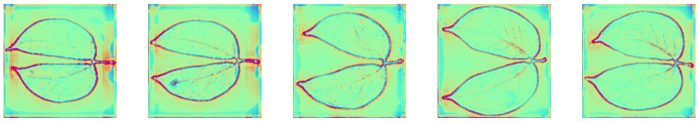
#	Species	Heatmaps
1	Ardisia revoluta	
2	Bauhinia ungulata	
3	Blakea maurofernandeziana	
4	Brosimum alicastrum	
5	Croton draco	
6	Dipteryx panamensis	
7	Erythrina poeppigiana	
8	Guazuma ulmifolia	
9	Hura crepitans	
10	Hymenaea courbaril	

Fig. 2. Heat maps generated with Layer-wise Relevance Propagation (LRP)

where, $R_d > 0$ expresses evidence of the presence of the structure to be classified, while $R_d < 0$ expresses evidence against its presence. To obtain this decomposition, the relevance concentrated at the output of a neural network is iteratively propagated backward through the network, using local propagation rules, until the final propagation maps the relevance back to the input image.

Table 3. Intra-specific variability of explanation vectors

Species	CNN Variability	Expert Variability
Ardisia revoluta	0.34	0.032
Bauhinia ungulata	0.17	0
Blakea maurofernandeziana	0.29	0.073
Brosimum alicastrum	0.19	0.16
Croton draco	0.10	0.085
Dipteryx panamensis	0.11	0
Erythrina poeppigiana	0.22	0.035
Guazuma ulmifolia	0.17	0.078
Hura crepitans	0.37	0.19
Hymenaea courbaril	0.087	0
Average	0.20	0.065
Standard deviation	0.099	0.0044

Table 4. Discrepancy of explanation vectors by species

Species	Discrepancy
Ardisia revoluta	0.62
Bauhinia ungulata	0.67
Blakea maurofernandeziana	0.77
Brosimum alicastrum	0.61
Croton draco	0.46
Dipteryx panamensis	0.68
Erythrina poeppigiana	0.56
Guazuma ulmifolia	0.75
Hura crepitans	0.49
Hymenaea courbaril	0.53
Average	0.61
Standard deviation	0.011

An in-depth overview of LRP can be found in [3, 15]. Figure 1 shows the heat maps generated using the LRP technique for some leaves from the CRLeaves dataset. It indicates the scientific name of the species, along with the probability reached by the model in the identification process. We can deduce that for the species Croton draco, Hura crepitans, and Blakea maurofernandeziana, the shape, and especially the venation, appear to be the most relevant features considered by the CNN in the identification process, while for the species Bauhinia ungulata, the shape seems to be the most relevant feature.

3 Methodology

3.1 Preparatory Activities (Workflow)

The following preparatory activities were carried out before the main experiment:

1. **Selecting a visual explanation technique**. With the goal of implementing a feature visualizer, experiments were conducted with the following visual explanation techniques: Class Activation Map, Saliency Map, and LRP. The expert criteria of two taxonomists was used to compare the heat maps of randomly selected images from the CRLeaves dataset. Section 4 gives more details about the chosen technique.

2. **Selecting the species for the experiment**. The CRLeaves dataset is composed of approximately 7262 images that correspond to 255 plant species from Costa Rica. It includes images of the adaxial (upper) and abaxial (lower) sides of the leaf.

   Discussions with the two taxonomists in the working team were conducted to select a subset of species and all of their corresponding specimen images (CRLeavesSubset dataset) that show notable interspecific variability and those the taxonomists were more familiar with according to their expertise area.

   Additionally, a randomly selected set of 10 species with 5 randomly selected specimen images (CRLeaves10 dataset) should be chosen to annotate their images, obtain their heat maps, and conduct the main experiment.

3. **Defining and fine-tuning a CNN**. With the images in CRLeavesSubset, a CNN was trained and validated to identify those plant species. The resulting architecture and the training and validation results of the network are briefly described in Section 4.

4. **Designing and implementing a “hot features” visualizer**. Using the technique chosen in preparatory activity 1., a “hot regions” visualizer was implemented. This allowed the generation of heat maps for each of the leaves in the CRLeaves10 dataset.

Table 5. 63 species selected from the CRLeaves dataset for the experiment, based on abaxial leaf images

No.	Species	No.	Species
1	Acnistus arborescens	33	Hyeronima alchorneoides
2	Aegiphila valerioi	34	**Hymenaea courbaril**
3	Anacardium excelsum	35	Manilkara chicle
4	Annona mucosa	36	Muntingia calabura
5	**Ardisia revoluta**	37	Ocotea sinuata
6	Astronium graveolens	38	Pachira quinata
7	Bauhinia purpurea	39	Persea americana
8	**Bauhinia ungulata**	40	Pimenta dioica
9	**Blakea maurofernandeziana**	41	Platymisciumparviflorum
10	**Brosimum alicastrum**	42	Platymiscium pinnatum
11	Calophyllum brasiliense	43	Posoqueria latifolia
12	Calycophyllum candidissimum	44	Psidium guajava
13	Cedrela odorata	45	Quercus corrugata
14	Cestrum tomentosum	46	Quercus insignis
15	Citharexylum donnell-smithii	47	Robinsonella lindeniana var. divergens
16	Clethra costaricensis	48	Samanea saman
17	Clusiacroatii	49	Sapium glandulosum
18	Coccoloba floribunda	50	Sideroxylon capiri
19	Colubrina spinosa	51	Simarouba glauca
20	Cordia eriostigma	52	Solanum rovirosanum
21	**Croton draco**	53	Stemmadenia donnell-smithii
22	Croton niveus	54	Swietenia macrophylla
23	Dalbergia retusa	55	Tabebuia impetiginosa
24	Dendropanax arboreus	56	Tabebuia ochracea
25	**Dipteryx panamensis**	57	Tabebuia rosea
26	**Erythrina poeppigiana**	58	Tabernaemontana litoralis
27	Eugenia hiraeifolia	59	Terminalia amazonia
28	Ficus cotinifolia	60	Terminalia oblonga
29	Genipa americana	61	Trichilia havanensis
30	**Guazuma ulmifolia**	62	Urera caracasana
31	Heliocarpus appendiculatus	63	Vernonia patens
32	**Hura crepitans**		

5. **Identifying features for feature maps**. The goal of this activity is to identify the n features used by taxonomists in the identification of the species in the CRLeavesSubset dataset.

   For this, five images of the leaves of each species were shown to two botanical taxonomists and they were asked to identify the set of features (e.g., apex, base, and main vein) needed to achieve a reliable identification. The resulting n-dimensional vector is called the explanation vector. The decision had to be unanimous.

6. **Annotating leaf images to create feature maps**. The objective of this activity is to capture the expert's knowledge used to identify plants from their leaves. For each of the 50 images in CRLeaves10, a taxonomist used a customized COCO Annotator image annotation tool [5]

to generate a feature map, which is equivalent to a manually produced heat map. The feature map of an image I consists of image I, an additional graphical layer with regions associated with each of the n features, and metadata (an explanation vector) that defines the weight associated with each of the features.

COCO ANNOTATOR is a web-based tool with a customizable interface that allows labeling a region of an image, tracking object instances, labeling disconnected objects, and storing and exporting annotations in COCO format. The COCO format is a JSON-based structure that defines how labels and metadata are stored for a set of images.

3.2 Experiment

The main objective of this experiment is to identify the relative weights (explanation vectors) assigned by the CNN to each of the features in the process of identifying the specimens in the CRLEAVES10 dataset. For this, a "hot pixel" counter was designed and implemented. It counts, for each of the images, the pixels that are above a threshold for each of the regions annotated by the taxonomists, corresponding to each of the considered features (apex, base, main vein, etc).

As an indicator of variability for criteria (explanation vectors) obtained from the CNN, a measure of intra-specific variability was defined and computed. Finally, for each of the 10 species in CRLEAVES10, a measure of discrepancy between the explanation vectors obtained from the heat maps and the (explanation vectors) assigned by the taxonomist was proposed and computed.

3.3 Terminology

Explanation vector: given an image I of a specimen, we define expl(I), the explanation of I, as an n-dimensional vector v where entry v_i is a positive real number that represents the relevance level of feature i in the identification process. For each image I, two explanation vectors are obtained: one is calculated from its heat map, and the other from the corresponding feature map.

For instance, Table 1 displays, for some leaves, the normalized explanation vectors obtained from their heat maps. Table 2 presents the normalized explanation vectors generated by the expert for the same leaves in Table 1.

To quantify the intra-specific variability of the criteria and the discrepancy between the criteria used by the CNN and an expert, two metrics were defined: intra-specific variability and discrepancy.

Intra-specific variability: Measures the level of variability between the explanations provided by an identifier (which can be an algorithm or a human) for a set of images of the same species X. Let $A = \{I_1, I_2, \cdots, I_n\}$ be a set of images of a species X, we define var(A), the variability of A, as the average (Euclidean distance), between the explanation vectors (expl(I)) for each of the images in A, that is:

$$\mathrm{var}(A) = \frac{2}{n(n-1)} \sum_{I,J \in A, I \neq J} \| \mathrm{expl}(I) - \mathrm{expl}(J) \|. \quad (2)$$

Discrepancy: Measures the level of discrepancy between the identifications made by the CNN and those made by the taxonomist for a given species X.

Let $A = \{I_1, I_2, \cdots, I_n\}$ be a set of images of a species X, and let $\mathrm{expl}_E(I)$ and $\mathrm{expl}_{\mathrm{CNN}}(I)$ be the explanation vectors for image I, generated from identifications made by the expert and the CNN, respectively.

We define discr(A), the discrepancy between the identifications made by the CNN and the taxonomist for a species X, as the average (Euclidean) distance, between the explanation vectors for each of the images in A:

$$\mathrm{discr}(A) = \frac{1}{n} \sum_{I \in A} \| \mathrm{expl}_{\mathrm{CNN}}(I) - \mathrm{expl}_E(I) \|. \quad (3)$$

Red Intensity: quantifies the importance of a pixel in the decision made by the CNN, calculated from the heat map. Let $X = (R, G, B)$ be a pixel in image I in RGB, then we define the red intensity of pixel X as:

$$RI(X) = \frac{100 \cdot R}{R + G + B}. \quad (4)$$

doi: 10.13053/CyS-28-4-4720

Table 6. Expert and CNN explanation vectors for 50 species

Leaf	Margin	Apex	Main vein	Base	Complement	Secondary veins
Ardisia revoluta_1_1_2	0.53	0.19	0.28	0.76	0.13	0.00
Ardisia revoluta_1_2_2	0.53	0.15	0.29	0.76	0.17	0.00
Ardisia revoluta_1_3_2	0.53	0.15	0.29	0.76	0.17	0
Ardisia revoluta_2_4_2	0.53	0.15	0.29	0.76	0.17	0.00
Ardisia revoluta_X_X_2(11)	0.53	0.19	0.29	0.76	0.135	0.00
Bauhuinia ungulata_2_3_2	0.34	0.86	0.34	0.17	0.00	0.00
Bauhuinia ungulata_2_4_2	0.34	0.86	0.34	0.17	0.00	0.00
Bauhuinia ungulata_X_X_2(1)	0.34	0.86	0.34	0.175	0.00	0.00
Bauhuinia ungulata_X_X_2(9)	0.34	0.86	0.34	0.17	0.00	0.00
Bauhuinia ungulata_X_X_2	0.34	0.86	0.34	0.17	0.00	0.00
Blakea maurofernadeziana_X_X_2(1)	0.18	0.18	0.88	0.18	0.35	0.00
Blakea maurofernadeziana_X_X_2(11)	0.18	0.18	0.88	0.18	0.35	0.00
Blakea maurofernadeziana_X_X_2(2)	0.17	0.086	0.86	0.17	0.43	0.00
Blakea maurofernadeziana_X_X_2(4)	0.17	0.086	0.86	0.17	0.43	0.00
Blakea maurofernadeziana_X_X_2(5)	0.17	0.086	0.866	0.17	0.43	0.00
Brosimum alicastrum_3_2_2	0.30	0.79	0.20	0.30	0.00	0.40
Brosimum alicastrum_X_X_2(12)	0.29	0.77	0.096	0.29	0.00	0.48
Brosimum alicastrum_X_X_2(18)	0.19	0.75	0.094	0.28	0.00	0.56
Brosimum alicastrum_X_X_2(19)	0.30	0.79	0.20	0.30	0.00	0.40
Brosimum alicastrum_X_X_2	0.31	0.72	0.21	0.41	0.00	0.41
Croton draco_1_2_2	0.22	0.33	0.55	0.65	0.11	0.33
Croton draco_2_2_2	0.22	0.33	0.55	0.65	0.11	0.33
Croton draco_X_X_2(14)	0.30	0.32	0.54	0.64	0.021	0.32
Croton draco_X_X_2(3)	0.26	0.32	0.53	0.64	0.021	0.36
Croton draco_X_X_2(5)	0.23	0.32	0.53	0.64	0.021	0.38
Dipteryx panamensis_X_X_2(1)	0.75	0.19	0.19	0.57	0.00	0.19
Dipteryx panamensis_X_X_2(15)	0.75	0.189	0.19	0.57	0.00	0.19
Dipteryx panamensis_X_X_2(6)	0.75	0.19	0.19	0.57	0.00	0.19
Dipteryx panamensis_X_X_2(9)	0.75	0.19	0.19	0.57	0.00	0.19
Dipteryx panamensis_X_X_2	0.75	0.19	0.19	0.57	0.00	0.19
Erythrina poeppigiana_X_X_2(11)	0.62	0.21	0.62	0.41	0.041	0.16
Erythrina poeppigiana_X_X_2(13)	0.61	0.20	0.61	0.41	0.00	0.20
Erythrina poeppigiana_X_X_2(4)	0.62	0.21	0.62	0.41	0.041	0.16
Erythrina poeppigiana_X_X_2(5)	0.62	0.20	0.61	0.41	0.00	0.20
Erythrina poeppigiana_X_X_2(9)	0.62	0.21	0.62	0.41	0.041	0.16
Guazuma ulmifolia_1_3_2	0.23	0.18	0.35	0.88	0.035	0.088
Guazuma ulmifolia_1_4_2	0.22	0.18	0.46	0.83	0.055	0.092
Guazuma ulmifolia_3_1_2	0.26	0.14	0.35	0.88	0.035	0.088
Guazuma ulmifolia_X_X_2(1)	0.26	0.17	0.35	0.88	0.00	0.088
Guazuma ulmifolia_X_X_2(16)	0.26	0.17	0.35	0.88	0.00	0.088
Hura crepitans_1_2_2	0.45	0.23	0.23	0.57	0.23	0.57
Hura crepitans_3_1_2	0.45	0.23	0.23	0.57	0.23	0.57
Hura crepitans_X_X_2(1)	0.62	0.21	0.21	0.51	0.00	0.51
Hura crepitans_X_X_2(16)	0.61	0.14	0.20	0.50	0.00	0.57
Hura crepitans_X_X_2(4)	0.62	0.23	0.21	0.51	0.00	0.51
Hymenaea courbaril_2_1_2	0.68	0.39	0.19	0.58	0.097	0.00
Hymenaea courbaril_2_3_2	0.68	0.39	0.19	0.58	0.097	0.00
Hymenaea courbaril_2_4_2	0.68	0.39	0.19	0.58	0.097	0.00
Hymenaea courbaril_3_1_2	0.68	0.39	0.19	0.58	0.097	0.00
Hymenaea courbaril_3_3_2	0.68	0.39	0.19	0.58	0.097	0.00

Table 7. CNN explanation vectors for 50 leaf specimens in the CRLeaves10 dataset

Leaf	Margin	Apex	Main vein	Base	Complement	Secondary veins
`Ardisia revoluta_1_1_2.jpg`	0.45	0.64	0.39	0.03	0.48	0.00
`Ardisia revoluta_1_2_2.jpg`	0.44	0.44	0.46	0.46	0.43	0.00
`Ardisia revoluta_1_3_2.jpg`	0.46	0.32	0.53	0.36	0.53	0.00
`Ardisia revoluta_2_4_2.jpg`	0.44	0.37	0.49	0.47	0.45	0.00
`Ardisia revoluta_X_X_2(11).jpg`	0.56	0.17	0.40	0.48	0.52	0.00
`Bauhinia ungulata_2_3_2.jpg`	0.43	0.48	0.47	0.39	0.46	0.00
`Bauhinia ungulata_2_4_2.jpg`	0.44	0.48	0.47	0.39	0.46	0.00
`Bauhinia ungulata_X_X_2(1).jpg`	0.48	0.46	0.39	0.51	0.37	0.00
`Bauhinia ungulata_X_X_2(9).jpg`	0.48	0.42	0.37	0.55	0.38	0.00
`Bauhinia ungulata_X_X_2.jpg`	0.54	0.43	0.28	0.57	0.35	0.00
`Blakea maurofernandeziana_X_X_2(1).jpg`	0.43	0.54	0.44	0.39	0.42	0.00
`Blakea maurofernandeziana_X_X_2(11).jpg`	0.42	0.57	0.34	0.41	0.32	0.34
`Blakea maurofernandeziana_X_X_2(2).jpg`	0.44	0.47	0.35	0.48	0.33	0.36
`Blakea maurofernandeziana_X_X_2(4).jpg`	0.43	0.48	0.44	0.46	0.43	0.00
`Blakea maurofernandeziana_X_X_2(5).jpg`	0.49	0.49	0.33	0.55	0.33	0.00
`Brosimum alicastrum_3_2_2.jpg`	0.41	0.50	0.33	0.55	0.32	0.27
`Brosimum alicastrum_X_X_2(12).jpg`	0.47	0.42	0.32	0.45	0.38	0.39
`Brosimum alicastrum_X_X_2(18).jpg`	0.47	0.33	0.29	0.56	0.35	0.39
`Brosimum alicastrum_X_X_2(19).jpg`	0.46	0.41	0.30	0.45	0.45	0.35
`Brosimum alicastrum_X_X_2.jpg`	0.40	0.51	0.42	0.37	0.38	0.35
`Croton draco_1_2_2.jpg`	0.41	0.40	0.41	0.42	0.41	0.40
`Croton draco_2_2_2.jpg`	0.40	0.39	0.42	0.42	0.40	0.42
`Croton draco_X_X_2(14).jpg`	0.45	0.42	0.34	0.48	0.37	0.38
`Croton draco_X_X_2(3).jpg`	0.42	0.39	0.40	0.51	0.37	0.33
`Croton draco_X_X_2(5).jpg`	0.38	0.40	0.47	0.42	0.41	0.37
`Dipteryx panamensis_X_X_2(1).jpg`	0.40	0.37	0.43	0.35	0.41	0.48
`Dipteryx panamensis_X_X_2(15).jpg`	0.42	0.38	0.41	0.48	0.40	0.36
`Dipteryx panamensis_X_X_2(6).jpg`	0.40	0.40	0.43	0.33	0.42	0.46
`Dipteryx panamensis_X_X_2(9).jpg`	0.41	0.42	0.39	0.40	0.40	0.43
`Dipteryx panamensis_X_X_2.jpg`	0.42	0.41	0.39	0.44	0.41	0.39
`Erythrina poeppigiana_X_X_2(11).jpg`	0.43	0.50	0.38	0.47	0.37	0.25
`Erythrina poeppigiana_X_X_2(13).jpg`	0.49	0.48	0.32	0.52	0.32	0.22
`Erythrina poeppigiana_X_X_2(4).jpg`	0.41	0.48	0.40	0.49	0.35	0.29
`Erythrina poeppigiana_X_X_2(5).jpg`	0.46	0.58	0.32	0.39	0.36	0.25
`Erythrina poeppigiana_X_X_2(9).jpg`	0.40	0.23	0.44	0.57	0.37	0.36
`Guazuma ulmifolia_1_3_2.jpg`	0.38	0.41	0.43	0.35	0.42	0.45
`Guazuma ulmifolia_1_4_2.jpg`	0.42	0.39	0.46	0.46	0.40	0.31
`Guazuma ulmifolia_3_1_2.jpg`	0.37	0.45	0.39	0.35	0.42	0.45
`Guazuma ulmifolia_X_X_2(1).jpg`	0.41	0.47	0.39	0.39	0.39	0.40
`Guazuma ulmifolia_X_X_2(16).jpg`	0.47	0.54	0.32	0.41	0.36	0.29
`Hura crepitans_1_2_2.jpg`	0.41	0.39	0.40	0.37	0.43	0.44
`Hura crepitans_3_1_2.jpg`	0.27	0.10	0.51	0.38	0.47	0.54
`Hura crepitans_X_X_2(1).jpg`	0.50	0.45	0.20	0.64	0.26	0.20
`Hura crepitans_X_X_2(16).jpg`	0.43	0.47	0.36	0.42	0.38	0.37
`Hura crepitans_X_X_2(4).jpg`	0.50	0.51	0.26	0.46	0.32	0.33
`Hymenaea courbaril_2_1_2.jpg`	0.47	0.51	0.38	0.42	0.45	0.00
`Hymenaea courbaril_2_3_2.jpg`	0.46	0.47	0.44	0.43	0.44	0.00
`Hymenaea courbaril_2_4_2.jpg`	0.45	0.44	0.48	0.40	0.46	0.00
`Hymenaea courbaril_3_1_2.jpg`	0.41	0.51	0.45	0.38	0.46	0.00
`Hymenaea courbaril_3_3_2.jpg`	0.45	0.43	0.45	0.46	0.45	0.00

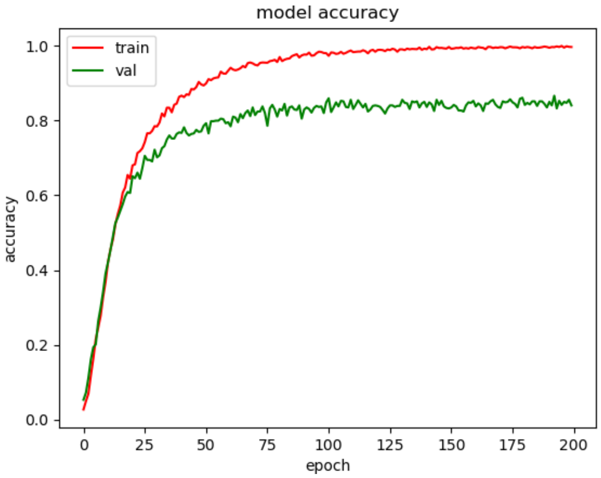


Fig. 3. CNN training and validation accuracy

Computing the explanation vectors: as indicated before, for each image, two explanation vectors are generated. One is explicitly determined by the expert when they annotate the specimen image; the other is obtained from the heat maps and the regions defined by the expert for each feature. Let R_k be the region annotated by the expert for feature k. Entry $v(k)$ of the explanation vector v is obtained by traversing pixel by pixel through region R_k and counting those pixels x whose red intensity satisfies $RI(x) \geq \alpha$, divided by $A(R_k)$, the total number of pixels in region R_k, i.e.:

$$v(k) = \frac{1}{A(R_k)} \sum_{x \in R_k} (RI(x) \geq \alpha), \qquad (5)$$

where an appropriate value for α should be determined experimentally. Since the explanation vectors are at different scales, they were normalized by dividing each of their entries by their magnitude. Finally, to calculate the discrepancy between the criteria used by the taxonomist and the CNN, Equation 3 was applied to each of the images in CRLeaves10.

4 Analysis of Results

Selecting a visual explanation technique. The LRP technique was chosen to generate the heat maps. This decision was based on the criteria of the expert taxonomists who were part of the working team. In their unanimous subjective assessment, LRP produced clearer images then CAM and SM.

Selecting the species for the experiment. Out of the 255 species in the the CRLeaves dataset, the two taxonomists in the working team selected 63 species, which are listed in Table 5. Only images of the abaxial side of the leaves were used because, for these species, the abaxial side was considered more discriminant. Then, from these 63 species, the following 10 were randomly selected: Ardisia revoluta, Bauhinia ungulata, Blakea maurofernandeziana, Brosimum alicastrum, Croton draco, Dipteryx panamensis, Erythrina poeppigiana, Guazuma ulmifolia, Hura crepitans, and Hymenaea courbaril. These 10 species were used to carry out activity 6 and the main experiment described in the section 3.2.

Defining and fine-tuning a CNN. The CNN used in this research was trained from scratch for 200 epochs with the 63 selected images of plant species. It consists of four convolutional blocks (convolutional layer plus max-pooling layer), a flatten layer, a dense layer, and a softmax layer at the top of the network. While the objective of this research has not been to optimize the performance of a CNN to identify species of plants in Costa Rica from leaf images, we achive a top-1 training accuracy of about 95% and a top-1 validation accuracy of about 80% as shown in Figure 3.

Designing and implementing a "hot features" visualizer. The LRP library developed to generate the heat maps was implemented as described in [15] and in the tutorial: Implementing Layer-Wise Relevance Propagation library[1]. This tutorial explains how to implement LRP using Tensorflow and Keras easily and efficiently. The LRP library enables the creation of heat maps using the LRP technique and the LRP-$\alpha\beta$ relevance propagation rule. For more details about its definition and implementation, refer to [3, 15]. As a result, the CNN includes a visual explanation mechanism based on the LRP technique.

[1] git.tu-berlin.de/gmontavon/lrp-tutorial

Table 8. Refinement of discrepancy of explanatory vectors by species

Species	CNN		Discrepancy	Expert	
Ardisia revoluta	Margin	Base	0.62	Base	Margin
Bauhinia ungulata	Base	Apex	0.67	Apex	Margin
Blakea maurofernandeziana	**Apex**	**Base**	**0.77**	**Main vein**	**Margin**
Brosimum alicastrum	Base	Apex	0.61	Apex	Base
Croton draco	Base	Main vein	0.46	Base	Main vein
Dipteryx panamensis	**Main vein**	**Apex**	**0.68**	**Margin**	**Base**
Erythrina poeppigiana	**Base**	**Apex**	**0.56**	**Main vein**	**Margin**
Guazuma ulmifolia	**Apex**	**Margin**	**0.75**	**Base**	**Main vein**
Hura crepitans	Base	Margin	0.49	Margin	Base
Hymenaea courbaril	Apex	Margin	0.53	Margin	Base

Figure 2 shows the heat maps generated using the LRP library for the 50 images in CRLeaves10. As can be seen, for some species, the apex is more prominent, while for others, it is the base or even the secondary veins.

Identifying features for feature maps. The features defined by both taxonomists are: apex, base, margin, main vein, secondary veins, and complement (the leaf surface minus the union of the regions of the five features). The first five features are generally key for any leaf-based identification, not only for species in the CRLeaves10 dataset, making it relatively easy to achieve unanimity.

It is important to note that there a few features that cannot be characterized by local leaf features beacause of their global qualitative nature, for instance, the shape of the leaf (heart-shaped, sagittate, etc.). This is an inherent limitation to the approach used in this research.

Annotating leaf images to create feature maps. The dataset resulting from annotating the 50 images in CRLeaves10 was called CRLeaves10Annotated. This dataset is available here[2].

The 50 explanation vectors (v_1, v_2, v_3, v_4, v_5, v_6) defined by the expert taxonomist are presented in Table 6.

Experiment. Once the LRP heat maps and annotated images were constructed, it was possible to algorithmically quantify the 50 explanation vectors associated with the decisions made by the CNN. After some experimentation with different red intensity threshold values, $\alpha = 0.35$ was considered a suitable value.

Table 7 presents the 50 explanation vectors obtained. Consequently, we have an explanation of the diagnostic features used by the CNN in the identification process of the specimens in the CRLeaves10 dataset. A deeper analysis of the explanation vectors in Table 7 takes us to address the following questions: How much intra-specific variability is present among the CNN explanation vectors?

How does it compare to the intra-specific variability among the expert explanation vectors? Furthermore, for each species, how similar, on average, are the criteria used by the CNN and the expert? We used Eucliden distance to calculate the distance between explanation vectors.

Since the explanation vectors are at different scales, they were normalized to be compared with each other.

[2] tecnube1-my.sharepoint.com/:f:/g/personal/gfigueroa_itcr_ac_cr/EoBbynu6o9hGhlMW03hmfOMBX5fdjUuKl01hKldRM61Jqw?e=bu1YEh

Additionally, as their inputs are positive, the angle θ between the vectors satisfies $\theta \in [0, \frac{\pi}{2}[$, therefore:

$$\begin{aligned} d(u,v) &= \|u-v\| = \sqrt{\sum_{i=1}^{n}\left(\frac{u_i}{\|u\|}-\frac{v_i}{\|v\|}\right)^2} \\ &= \sqrt{2-2\sum_{i=1}^{n}\frac{u_i v_i}{\|u\|\|v\|}} \qquad (6) \\ &= \sqrt{2-2\frac{u\cdot v}{\|u\|\|v\|}} \\ &= \sqrt{2}\sqrt{1-\cos(\theta)} < \sqrt{2}. \end{aligned}$$

In other words, the distance between two explanation vectors is upper-bounded by $\sqrt{2} \approx 1.41$. Table 3 shows the results of applying Equation 2 to calculate the the intra-specific variability of the explanation vectors for each species. As shown in Table 3, the intra-specific variability of the explanation vectors assigned by the expert is very low (values very close to zero), as expected.

On the other hand, we can observe that the explanation vectors generated from the CNN are more variable but their values are considerably closer to zero than to $\sqrt{2}$, indicating that the relevance values assigned by the CNN to each of the chosen features are similar among leaves of the same species.

Note that the explanation vectors obtained by the CNN for the species Hymenaea courbaril are the least variable among themselves (intra-specific variability value very close to zero), which is also the case for the expert assessment. The species for which the highest CNN intra-specific variability was obtained is Hura crepitans, which is also the one with the highest intra-specific Expert variability.

This high value may be due to the fact that, as shown in Figure 2, the five images have heat maps in which sometimes the apex stands out more (row 9, first and second maps from left to right) and sometimes the secondary veins or other features. This is confirmed in the explanation vectors for this species in Table 6.

By applying Equation 3, we calculated the discrepancy between the explanation vectors annotated by the expert and those generated from the CNN. Table 4 shows the results. As can be observed in Table 4, the discrepancy between the explanation vectors assigned by the expert and those generated from CNN is, on average, 0.61.

Given that the maximum experimental value of discrepancy is 1.41, we can state that both explanation vectors are similar, leading us to believe that the CNN, in its identification process, assigns a relevance similar to that assigned by the expert to the chosen features.

To refine the level of discrepancy, we can observe Table 8, which shows, in descending order, the two most prioritized features, on average, used for the identification of the indicated species, obtained from the explanation vectors of both, the CNN and the expert.

As observed, in 6 out of the 10 species, the CNN and the expert use at least one prioritized feature in common. For the remaining 4 species (highlighted in bold in Table 8), the CNN and the expert use different prioritized features in the identification process, and precisely for these vectors, the discrepancy is high (closer to 1.4, indicating lower similarity between the vectors).

5 Conclusion and Future Work

As a first step in defining and implementing explanation components for a CNN, this research has provided, for the first time, qualitative and quantitative information about the main features used by a CNN in the identification of 10 species of vascular plants from Costa Rica. At the end of the research, the following products have been generated:

- A tabular explanation of the diagnostic features used by the CNN in the identification process of the specimens in the CRLeaves10 dataset.
- A graphical explanation (heat map) that uses the LRP technique for each of the 50 leaf images in the CRLeaves10 dataset.

- Two quantitative measurements of variability in the identification of specimens of the same species, for 10 plant species from Costa Rica, one using the CNN and one using the expert's criteria.
- A quantitative measurement of the discrepancy/similarity between the criteria used by a taxonomist and a CNN in the identification of 10 plant species from Costa Rica.
- A workflow to add explainability to a CNN.
- A CNN that identifies 63 plant species from Costa Rica based on their leaves with a top-1 validation accuracy close to 80%, which can be used as a benchmark for studying other features or training it for a larger number of species as the CRLEAVES dataset grows.
- A relevance region visualizer based on the LRP technique, which can be integrated into other CNNs to generate the respective heat maps.
- A parameterized tool that quantifies the red intensity of each annotated region in images from CRLEAVESANNOTATED and the heat maps.
- A set of 50 annotated leaf images (CRLEAVESANNOTATED), with annotated discriminative features and their respective relative weights for each feature.

With current techniques, it is not possible to automatically generalize the results to any CNN, but it is feasible to do so for a larger number of species of the Costa Rican flora. The essential task is to continue expanding the CRLEAVES dataset and annotating it using the COCO ANNOTATOR tool. It is important to keep the feature visualizer up-to-date with libraries that implement new visualization techniques and include them in the workflow of future experiments.

Scaling up this work for a larger set of species and specimens per species requires streamlining the current workflow. An obvious current bottleneck is the manual annotation of leaf images. It would be ideal to develop and fine-tune an intelligent tool to extract the regions where each feature is present in the leaf image. These annotations plus the corresponding heat map would suffice to automatically compute the corresponding explanation vector.

The ideas developed in this research could be applied to other types of biological samples. For example, the work described in [7] generated a dataset of wood cut samples for 147 tree species from Costa Rica. Additionally, a CNN was developed to identify 75 species for which there are at least five samples.

It would be very important to conduct similar experiments to improve the interpretability of that CNN and to measure intra-specific variability and discrepancy between the criteria used by taxonomists and the CNN.

Acknowledgments

We thank Dr. José Carranza-Rojas for facilitating the use of the CRLEAVES dataset, Engineer Adán Mora-Fallas for his guidance in using the COCO ANNOTATOR tool, and Engineer Enrique Loría Cordero for programming pixel counting tools based on the JSON file generated by COCO ANNOTATOR.

We also appreciate the collaboration of Engineer Nelson Zamora-Villalobos in selecting the set of 63 species used to train the CNN and defining the features used for identification. Finally, we express our gratitude to the Costa Rica Institute of Technology for the financial support and guidance provided.

References

1. **Aakif, A., Khan, M. (2015).** Automatic classification of plants based on their leaves. Biosystems Engineering, Vol. 139, pp. 66–75. DOI: 10.1016/j.biosystemseng.2015.08.003.

2. **Abdulazeez, A. M., Zeebaree, D. Q., Zebari, D. A., Hameed, T. H. (2021).** Leaf identification based on shape, color, texture and vines using probabilistic neural network. Computación y Sistemas, Vol. 25, No. 3, pp. 1–15. DOI: 10 .13053/cys-25-3-3470.

3. **Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., Samek, W. (2015).** On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE, Vol. 10, No. 7, pp. 1–46. DOI: 10.1371/journal.pone.0130140.

4. **Biran, O., Cotton, C. V. (2017).** Explanation and justification in machine learning: A survey. Proceedings of the International Joint Conference on Artificial Intelligence and Workshop on Explainable AI, pp. 8–13.

5. **Brooks, J. (2019).** COCO annotator. annotator.ait.ac.th.

6. **Doshi-Velez, F., Kim, B. (2017).** Towards a rigorous science of interpretable machine learning. DOI: 10.48550/ARXIV.1702.08608.

7. **Figueroa-Mata, G., Mata-Montero, E., Valverde-Otárola, J. C., Arias-Aguilar, D., Zamora-Villalobos, N. (2022).** Using deep learning to identify Costa Rican native tree species from wood cut images. Frontiers in Plant Science, Vol. 13. DOI: 10.3389/fpls.2022.789227.

8. **Goëau, H., Bonnet, P., Joly, A. (2017).** Plant identification based on noisy web data: The amazing performance of deep learning. Proceedings of the Conference and Labs of the Evaluation Forum, pp. 1–13.

9. **Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (2019).** Un report: Nature's dangerous decline 'unprecedented'; species extinction rates 'acelerating'. www.un.org/sustainabledevelopment/blog/2019/05/nature-decline-unprecedented-report/.

10. **Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., Soares, J. V. B. (2012).** Leafsnap: A computer vision system for automatic plant species identification. Computer Vision – ECCV 2012, pp. 502–516. DOI: 10.1007/978-3-642-33709-3_36.

11. **Lasseck, M. (2017).** Image-based plant species identification with deep convolutional neural networks. Proceedings of the Conference and Labs of the Evaluation Forum, pp. 1–12.

12. **Lee, S. H., Chan, C. S., Mayo, S. J., Remagnino, P. (2017).** How deep learning extracts and learns leaf features for plant classification. Pattern Recognition, Vol. 71, pp. 1–13. DOI: 10.1016/j.patcog.2017.05.015.

13. **Mata-Montero, E., Carranza-Rojas, J. (2015).** A texture and curvature bimodal leaf recognition model for identification of Costa Rican plant species. Proceedings of the Latin American Computing Conference, pp. 1–12. DOI: 10.1109/CLEI.2015.7360026.

14. **Miller, T. (2019).** Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, Vol. 267, pp. 1–38. DOI: 10.1016/j.artint.2018.07.007.

15. **Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K. R. (2019).** Layer-wise relevance propagation: An overview. Explainable AI: Interpreting, explaining and visualizing deep learning, Vol. 11700, pp. 193–209. DOI: 10.1007/978-3-030-28954-6_10.

16. **Ras, G., Xie, N., Van-Gerven, M., Doran, D. (2022).** Explainable deep learning: A field guide for the uninitiated. Journal of Artificial Intelligence Research, Vol. 73, pp. 329–397. DOI: 10.1613/jair.1.13200.

17. **Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Müller, K. R. (2021).** Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, Vol. 109, No. 3, pp. 247–278. DOI: 10.1109/JPROC.2021.3060483.

18. **Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017).** Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

19. **Simonyan, K., Vedaldi, A., Zisserman, A. (2014).** Deep inside convolutional networks:

Visualising image classification models and saliency maps. Workshop at International Conference on Learning Representations, pp. 1–8. DOI: doi.org/10.48550/arXiv.1312.6034.

20. **Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P. (2018).** Automated plant species identification—trends and future directions. PLOS Computational Biology, Vol. 14, No. 4, pp. 1–19. DOI: 10.1371/journal.pcbi.1005993.

21. **Wu, H., Xiang, Y., Liu, J., Wen, Z. (2017).** Automatic leaf recognition based on deep convolutional networks. Neural Information Processing, Springer International Publishing, Vol. 10636, pp. 505–515. DOI: 10.1007/978-3-319-70090-8_52.

22. **Zeiler, M. D., Fergus, R. (2014).** Visualizing and understanding convolutional networks. Proceedings of the European Conference on Computer Vision, Springer International Publishing, Vol. 8689, pp. 818–833. DOI: 10.1007/978-3-319-10590-1_53.

23. **Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A. (2016).** Learning deep features for discriminative localization. IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929. DOI: 10.48550/arXiv.1512.04150.

Article received on 16/10/2023; accepted on 21/08/2024.
**Corresponding author is Geovanni Figueroa-Mata.*