Study of Explainability Analysis Methods for the LAMDA Family Algorithms in Classification and Clustering Tasks

Carlos Quintero Gull
Dpto de Cs Aplicadas y Humanísticas. Facultad Ingeniería
Universidad de Los Andes
Mérida, Venezuela
carlgull@gmail.com

Jose Aguilar
IMDEA Networks Institute, Madrid, Spain
CEMISID, Universidad de Los Andes, Mérida, Venezuela;
GIDITIC, Universidad EAFIT, Medellín, Colombia
jose.aguilar@ula.ve

Rodrigo García
Universidad del Sinú
Monteria, Colombia
Rodrigogarcia@unisinu.edu.co

***Abstract*—Explainability analysis is a very relevant topic today, due to the interest of allowing the interpretability of machine learning models. In this work, we carry out an in-depth study of explainability analysis for the algorithms of the LAMDA (Learning Algorithm for Multivariate Data Analysis) family that have been used in the context of supervised and unsupervised learning. In particular, for the case of classification the LAMDA-HAD algorithm, and for the case of clustering the LAMDA-RD algorithm. For the explainability analysis, two classic methods from the explainability area were considered, LIME (Local Interpretable Model-Agnostic Explanation) and Feature Importance, and another one developed by us for the LAMDA family. In particular, our explainability method for LAMDA allows measuring the importance of each characteristic in a general way, and for each cluster. In general, the results obtained in both cases (classification and clustering) are satisfactory, especially because our explainability method for LAMDA gives an explainability similar to the traditional ones, but in addition, it can be given by cluster.**

***Keywords*—Explainability analysis, LAMDA, LIME, feature importance, classification, clustering,**

I. INTRODUCTION

The implementation and use of machine learning methods is of increasing importance today, because there are different applications in the real world, for example, in medicine, in industry, in academia, in others [1], [2]. However, interest has also increased in understanding how machine learning models operate to make informed decisions [2]. Generally, many of these methods generate black box models (they are not interpretable), which is why in Artificial Intelligence an area has emerged dedicated to developing methods that allow an explainability and interpretability analysis to be carried out on them.

This area of AI has considered three relevant aspects, transparency, interpretability and explainability of the methods [11], [13]. Transparency occurs when the parameters of each model are easy to determine and justify, interpretability is when the model is easy for humans to understand, and explainability allows us to understand why a specific output is produced for a particular input. In this sense, some of the methods that have been developed are the Feature Importance Method, and LIME (Local Interpretable Model-Agnostic Explanation), among others, which try to analyze the importance of the characteristics in the results obtained [12].

On the other hand, LAMDA is a Machine Learning algorithm based on the theory of fuzzy sets that allows individuals to be assigned to any class (classification tasks) or cluster (clustering task) [7]. Multiple versions of LAMDA have been developed for classification and clustering tasks. In this particular work, the explainability of the LAMDA-HAD algorithm in classification tasks [9], and the LAMDA-RD algorithm in clusterin tasks [10], will be analyzed, as they are algorithms of the LAMDA family with which satisfactory results have been obtained in previous works [8].

Thus, this work performs an in-depth analysis of the explainability of the LAMDA algorithms, both in the classification context (LAMDA-HAD algorithm), as well as in the clustering context (LAMDA-RD algorithm). For the study of explainability in both cases, traditional explainability methods are considered, but at the same time, one of our own is proposed for the LAMDA family. The most relevant contributions are:

- An explainability analysis for several algorithms of the LAMDA family, both for the classification context (LAMDA-RD) and for the clustering context (LAMDA-RD), until now not previously studied.
- The definition of a new method of explainability analysis by class/cluster for algorithms of the LAMDA family, based on the degrees of membership of the variables to the classes/cluster

This work is organized as follows: Section 2 presents the related works. Section 3 introduces the fundamentals of LAMDA-HAD and LAMDA-RD algorithms. Section 4 presents the explainability methods, especially our explainability method for LAMDA. Section 5 shows the experiments, and finally, Section 7 presents the conclusions and future works.

II. RELATED WORKS

In this section, we present some recent works about the explainability and interpretability in Machine Learning. In particular, in this section we refer to works that carry out

literature reviews on articles that talk about explainability or interpretability methods in the area of artificial intelligence.

Lisboa et al. [3] made a review of different interpretability and explainability methods. In this sense, they mention that interpretable Machine learning models is about models that are inherently interpretable by the human mind, whereas explainable Machine Learning "tries to provide post hoc explanations for existing black box models". Heuillet et al. [4] made a review about Explainable Reinforcement Learning (XRL) as a new subfield of Explainable Artificial Intelligence. They evaluate the XRL into two categories Transparent algorithms and post-hoc explainable, and they conclude that despite that different methods for XRL exist, it is not clear that these methods serve for all purposes since they are specifically designed for a particular task.

Bücker et al. [1] developed a framework to describe machine learning models in the categories of transparent, auditable and explainable. In this framework, two aspects are considered, the first delves into the factors that can influence the performance of the model, the second has to do with the explainability of the model, global or local. The global is referred to when reviewing the factors that influenced the discrimination of the model, and the local facilitates the explanation of the model for specific cases of input. Goodwin et al., 2020 [5] defined a taxonomy on explainability in the context of machine learning to try to understand the context of the problem. That allowed them to identify gaps and potential solutions, to implement explainable machine learning.

Finally, Linardatos et al. [6] conducted a literature review on the interpretability of machine learning methods. In this sense, the authors classify interpretability methods taking into account various factors such as local and global explainability, the type of data (discrete, continuous), the type of model (specific or agnostic model), among others. According to the reviewed literature, there are no previous works that analyze the explainability for the specific case of the LAMDA family algorithms.

III. Fundamentals of the LAMDA Family Algorithms

In this section, we present the conceptual bases of the LAMDA family algorithms [7]. LAMDA is based on the assignment of individuals to a cluster/class using its membership grade. For that, each individual X is represented by a vector of features:

$$X = [x_1;\ x_2;\ \ldots;\ x_j\ ;\ \ldots;\ x_n]$$

Where $x_j$is the feature j of the individual X.

In general, it is necessary to standardize the vector of features of each individual. In this case, we standardize it using the minimum and maximum values:

$$\bar{x}_j = \frac{x_j - x_{jmin}}{x_{jmax} - x_{jmin}} \quad (1)$$

Where: $\bar{x}_j$ is the standardized feature; $\bar{x}_{jmin}$ is the minimum value of feature j; $\bar{x}_{jmax}$ is the maximum value of feature j.

Below, we present the main definitions of LAMDA.

Definition 1. The Marginal Adequacy Degree (MAD) establishes how similar a feature of an individual is with respect to the same feature in a given cluster/class. We use density functions to calculate MAD, and one of the typical is the Fuzzy Binomial function:

$$MAD_{kj} = MAD(\bar{x}_j/\rho_{kj}) = \rho_{kj}{}^{\bar{x}_j}\ (1-\rho_{kj})^{(1-\bar{x}_j)} \quad (2)$$

Where ρ_{kj} is the average value of the feature j that belongs to the cluster/class k, determined using Eq. (3):

$$\rho_{kj} = \frac{1}{n_{kj}} \sum_{t=1}^{n_{kj}} \bar{x}_j(t) \quad (3)$$

Where n_{kj} is the number of individuals of class/cluster k and feature j.

Definition 2. The Global Adequacy Degree (GAD) defines the adequacy of an individual to each cluster/class. This value is determined using the next Eq.:

$$GAD_{k,\bar{X}} = \alpha T(MAD_{k,1}, \ldots, MAD_{k,n}) + (1 - \alpha) S(MAD_{k,1}, \ldots MAD_{k,n}) \quad (4)$$

Where $\alpha \in [0, 1]$ is the exigency parameter; T and S are linear interpolation functions.

Definition 3. Let $p = \{1, \ldots, m\}$ be the number of existing clusters/classes in a dataset. The object $\bar{X}$ is assigned to the cluster/class with the maximum GAD, where the index corresponds to the number of the cluster/class.

$$index = \max(GAD_{1,\bar{X}}, GAD_{k,\bar{X}}, \ldots, GAD_{m,\bar{X}}, GAD_{NIC,\bar{X}}) \quad (5)$$

NIC is used to create new clusters/classes after the training, when an object is unrecognized (it is sent to the NIC), making the algorithm more adaptive (online learning). It is considered $\rho_{NIC} = 0.5$ because with this value in Eqs. (2), $MAD_{NIC} = 0.5$ for any value of the feature $\bar{x}_j$.

A. LAMDA-HAD

LAMDA-HAD is an extension to LAMDA that defines an adaptable NICs by class to prevent that correctly classified individuals create new classes [9]; and using this value computes the Higher Adequacy Degree (HAD). Below, we present the main definitions of LAMDA-HAD, defined for classification tasks:

Definition 4: Let $MGAD_{k,p}$ be the average of $GAD's$ of the individual in the class p in the class k:

$$MGAD_{k,p} = \frac{1}{n_k} \sum_{t=1}^{t=n_k} GAD_{p,t} \quad (6)$$

Where $MGAD_{k,p}$ is the average of the Global Adequacy Degree of class k in class p; n_k is the number of objects belonging to class k, and $GAD_{p,t}$ is the GAD of the individual t for class p, in class k.

Definition 5: Let GAD_{NIC_p} be the GAD of the NIC for the class p computed as:

$$GAD_{NIC_p} = \frac{1}{m} \sum_{p=1}^{p=m} MGAD_{k,p} \quad (7)$$

Definition 6: Let $AD_{GAD_{k,p,\bar{X}}}$ be the new Global Adequacy Degree (GAD), which is a parameter that allows comparing the similarity between the GAD of an individual and each $MGAD_{k,p}$:

$$AD_{GAD_{k,p,\bar{X}}} = MGAD_{k,p}{}^{GAD_{p,\bar{X}}} (1 - MGAD_{k,p})^{(1-GAD_{p,\bar{X}})} \quad (8)$$

Definition 7: The Highest Degree of Adequacy of an individual to a class $(HAD_{k,\bar{X}})$ is determined by adding all the $AD_{GAD_{k,p,\bar{X}}}$ in class p:

$$HAD_{k,\bar{X}} = \sum_{p=1}^{m} AD_{GAD_{k,p,\bar{X}}} \quad (9)$$

Let E_I be the class to which the individual has the highest probability of belonging:

$$E_{I,\bar{X}} = max(HAD_{1,\bar{X}}, HAD_{2,\bar{X}}, \dots HAD_{k,\bar{X}}, \dots . HAD_{m,\bar{X}}) \quad (10)$$

Definition 8: Let index be the value that identifies the class that an individual will be allocated, which is obtained by comparing the maximum value between E_I and the $GAD_{NIC_{E_I}}$:

$$\boldsymbol{index} = max\left(HAD_{E_I,\bar{X}}, GAD_{NIC_{E_I}}\right) \quad (11)$$

Thus, it is verified if the maximum value of $\mathrm{HAD}_{\mathrm{E_I},\bar{\mathrm{X}}}$ is greater than the $\mathrm{GAD}_{\mathrm{NIC}_{\mathrm{E_I}}}$ (the $\mathrm{GAD}_{\mathrm{NIC}}$ adapted to each class).

Once the LAMDA-HAD algorithm is finished, the result will be the number of the class to which the individual is assigned; otherwise, the individual is sent to the non-informative class (NIC).

B. LAMDA-RD

The LAMDA-RD algorithm was developed to improve the LAMDA algorithm for clustering tasks because LAMDA tends to create more clusters than necessary [10]. In this sense, LAMDA-RD includes a robust metric of clustering, and also, an automatic cluster fusion process. Next, the conceptual basis of this algorithm:

Definition 9. The Cauchy Marginal Adequacy Degree (CMAD) corresponds to the Marginal Adequacy Degree, but using the Fuzzy Cauchy Function:

$$CMAD = \frac{1}{1+dist(\bar{x}_j,\rho_{kj})} \quad (12)$$

Where: $\mathrm{dist}(\bar{\mathrm{x}}_\mathrm{j}, \rho_{\mathrm{kj}})$ is the distance between the individual xj and the average ρ_{kj}.

Definition 10. The Robust Marginal Adequacy Degree (RMAD) corresponds to the Marginal Adequacy Degree, but now accompanied by a factor that penalizes each cluster. It is defined as:

$$RMAD = k_{\bar{x}k} * CMAD \quad (13)$$

To calculate $k_{\bar{x}k}$ is required the average distance of the individual between the clusters:

$$d_{k,\bar{X}_r} = dist(\bar{x}_j, \rho_{k,j}) = \frac{1}{n}\sum_{j=1}^{n}\left|\bar{x}_j - \rho_{k,j}\right| \quad (14)$$

And the average distance between neighbor clusters $(d_{n,b}) \in [0,1]$ that describes the average distance between clusters, and is obtained through calibration.

Definition 11. The Density of a cluster. Let $dt \in [0,1]$ be a threshold of the density of a cluster, which is obtained through a calibration process.

Definition 12. The Penalty factor. If the average distance $d_{k,\bar{X}}$ is greater than d_{nb}, then the penalty factor is:

$$k_{\bar{x}k} = \frac{d_{n,b}}{d_{n,b}+dist\left(d_{k,\bar{X}_r}, d_{n,b}\right)} \quad (15)$$

On the other hand, Global Adequacy Degree is a linear combination of the Robust Marginal Adequacy Degree (RMAD), as shown in Eq. (16):

$$\mathrm{GAD}_{k,\bar{X}} = \alpha T(RMAD_{k,1}, \dots, RMAD_{k,n},) + (1-\alpha)S(RMAD_{k,1,} \dots RMAD_{k,n}) \quad (16)$$

In addition, LAMDA-RD has an automatic merge algorithm, which determines the compactness of neighboring clusters, calculates the distance between the individuals of the neighboring clusters to establish individuals in the overlap zone, and if the ratio of individuals in the overlap area with respect to the total of individuals between the two neighboring clusters is greater than a threshold, then it proceeds with the merge process.

IV. Explainability methods

In this section, we explain the explainability methods used in this work, and propose an explainability method for LAMDA.

A. Local Interpretable Model Agnostic Explications (LIME)

LIME is a method developed by Ribeiro et al (2009) for explaining the prediction of a model, using a local model surrogate for each individual prediction. Mathematically, it can be expressed as follows

$$\varepsilon(x) = argmin_g\left(L(f,g,\pi_x) + \theta(g)\right) \quad (17)$$

Where g is the family of possible explanations, L is the loss function and it measures how close g (surrogate model) is to the prediction of f (original model) in its vicinity π_x, and θ(g) measures the complexity of the surrogate model.

B. Feature Importance (FI)

FI is an algorithm that allows determining the importance of the features given a dataset providing a way of classifying features according to their importance in the performance of the method. The conceptual basis of this method is next: it executes the model modifying the values of one feature and leaving the rest fixed and evaluates the quality of the results. It does this for each of the features, and establishes a ranking according to the results obtained with each feature. Thus, the sensitivity of the model can be established with each feature, such that the most sensitive are the most relevant/important.

C. Explainability based on LAMDA Algorithm

The explainability in the LAMDA-HAD algorithm for classification tasks is based on the importance of each feature j in each defined class k. Table 1 shows the matrix for the determination of the explainability according to the LAMDA algorithm, considering p features and k classes, where $MAD^n{}_{k,p}$ represents the MAD of the feature p of the individual n in class k.

TABLE 1 MATRIX FOR THE CALCULATION OF THE FEATURE IMPORTANCE FOR LAMDA.

Index	$MAD_{1,1}$	$MAD_{1,2}$	...	$MAD_{1,p}$	$MAD_{2,1}$	...	$MAD_{k,p}$
1	$MAD^1{}_{1,1}$	$MAD^1{}_{1,2}$	...	$MAD^1{}_{1,p}$	$MAD^1{}_{2,1}$	...	$MAD^1{}_{k,p}$
2	$MAD^2{}_{1,1}$	$MAD^2{}_{1,2}$		$MAD^2{}_{1p}$	$MAD^2{}_{2,1}$	...	$MAD^2{}_{kp}$
3	$MAD^3{}_{1,1}$	$MAD^3{}_{1,2}$		$MAD^3{}_{1,p}$	$MAD^3{}_{2,1}$	...	$MAD^3{}_{k,p}$
.	.	.	...	.	.	...	.
.	.	.	...	.	.	...	.
.	.	.	...	.	.	...	.
n	$MAD^n{}_{1,1}$	$MAD^n{}_{1,2}$		$MAD^n{}_{1,p}$	$MAD^n{}_{2,1}$	...	$MAD^n{}_{k,p}$

In order to determine the explainability for the LAMDA algorithm, we will describe some definitions:

Definition 13: Importance of the Feature p in the class k ($\psi_{k,p}$). This parameter represents the relevance of the feature p for the class k in Table 1, calculated by the next equation:

$$\psi_{k,p} = \frac{\sum_{i=1}^{n} MAD_{k,p}^{i}}{N} \qquad (18)$$

Definition 14: *Global Importance of the Feature p (ψ_p).* This parameter represents the global average of the feature p for all the classes/clusters in Table 1, calculated by the next equation:

$$\psi_{p} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} MAD_{j,p}^{i}}{n*k} \qquad (19)$$

For the calculation of the feature importance in clustering tasks, the value of $MAD^n{}_{k,p}$ is replaced for $RMAD^n{}_{k,p}$ that represents the Robust Marginal Adequation Degree of the descriptor p of the individual n in the cluster k.

Specifically, explainability in LAMDA-HAD allows determining the importance of each feature for each class in the dataset. In the context of LAMDA-RD, it allows determining the importance of each characteristic in each cluster. In both cases, it is novel in the context of explainability to be able to do it at the level of each class/cluster.

V. EXPERIMENTS

A. Experimental Protocol

In the next, the datasets used are described below. For this work, we have used four datasets, two for classification tasks (iris and Dengue datasets) and two for clustering tasks (Brest Cancer and wine datasets). Table 2 shows the dataset for evaluating the performance of our proposal.

TABLE 2. DATASET FOR EVALUATING

Dataset	Size	features	Characteristics
iris	150	4	Four features length and width of sepals and petals 50 records for each of three species of iris. In this dataset, each species is considered one class
dengue	32559	22	The dataset contains the records of patients with Dengue. The features are according to the symptoms of Dengue. This dataset contains three classes according to the result of the Dengue Test
Breast Cancer	684	9	Contains records for patients with Cancer. The classes are about whether the patient has cancer or not

The Iris dataset is balanced and has low dimensionality. This dataset can be downloaded at https://archive.ics.uci.edu/dataset/53/iris. Table 3 shows the description of each feature in the dataset

TABLE 3. DESCRIPTION OF FEATURES FOR IRIS DATASET

Name	Description
Sepal_L	Sepal Length
Sepal_W	Sepal width
Petal_L	Petal Length
Petal_W	Petal width
Species	Classes: Setosa, Versicolor Virgínica

The Dengue dataset is unbalanced and has high dimensionality. This dataset can be downloaded at https://medata.gov.co/dataset/dengue. Table 4 shows the description for each feature.

TABLE 4. DESCRIPTION OF FEATURES FOR DENGUE DATASET

Name	Description
Age	Time elapsed since the birth of an individual
Fever	Increase in body temperature
Cephalea	Pain and discomfort located in any part of the head
Pain BE	Pain behind eyes
Myalgias	Muscle aches
Arthralgias	Joint pain
Rash	Skin exanthema
Abdominal pain	Intense pain, located in the epigastrium and/or right hypochondrium
Vomit	Violent expulsion by the mouth of what is contained in the stomach.
Lethargy	State of tiredness and deep and prolonged sleep
Hypotension	Excessively low-blood pressure on the artery wall
Hepatomegaly	Condition of having an enlarged liver
Mucosal bleeding	Manifestations of mild to severe bleeding in the nasal mucosa, gums, female genital tract, brain, lungs, digestive tract and hematuria skin,
Hypothermia	Decrease of body temperature
High hematocrit	Indirect increase in hematocrit test
Low platelets	Decrease of platelet levels in the blood
Edema	Swelling caused by excess fluid trapped in body tissues
Extravasation	It is characterized by serous spills at the level of various cavities

Bleeding	Blood leaks from the arteries, veins or capillaries through which it circulates, especially when it is produced in very large quantities
Shock	Manifestation of severity evidenced by cold skin, thready pulse, tachycardia and Hypotension
Organ failure	Affectation of several organs due to the extravasation of liquids
Severity	Dengue severity

Finally, the Breast Cancer dataset has low records and low dimensionality. This dataset can be downloaded from https://archive.ics.uci.edu/dataset/14/breast+cancer. Table 5 shows the description of each feature of this dataset.

TABLE 5. DESCRIPTION OF FEATURES FOR BREAST CANCER DATASET

Name	Description
age	Time elapsed since the birth of an individual
menopause	Categorical feature according to level of menopause at the moment
tumor_S	Size of the tumor in mm
inv-nodes	Metric about the presence
node-caps	Presence of the cancer cells
deg-malig	Grade of the Histological tumor
breast	Side of the Breast affected
beast-quad	Breast Quadrant affected
irradiat	Radiotherapy Applied

For carrying out our experiments in the context of classification, we partitioned each dataset into 80% for training and 20% for testing the model.

B. Performance Metrics

In this section, we present the metric for the evaluation of our proposal. The metrics used for the classification tasks are:

- *Accuracy (Acc):* Proportion of individuals correctly classified,
- *Precision (P):* it is the proportion of correct predictions among all predictions of a certain class.
- F1-score (L): it can be interpreted as a harmonic mean of the precision and recall,

The metrics used for the clustering tasks are:

- *Silhouette coefficient (**SC**):* The range of this metric is between $[-1, 1]$, 1 for well clustering (dense al well separated) and -1 for not well clustering.
- *Sum of Square Within of Cluster (SSW):* This metric allows determining the compactness of the cluster.
- *Sum of Square Between Cluster (SSB):* This metric allows determining the separation in the clusters formed (intecluster distance.

C. Explainability for Classification tasks

Table 7 shows the performance results for the LAMDA-HAD Algorithm. We can note that in (average), in the Iris dataset, all the metrics are above of 95%. In the Dengue dataset, we obtained an Accuracy of 78%. In the other metrics, we can observe that the average is above 78%, which determines that the performance of LAMDA-HAD is satisfactory.

TABLE 7. RESULTS FOR CLASSIFICATION TASKS FOR EACH DATASET

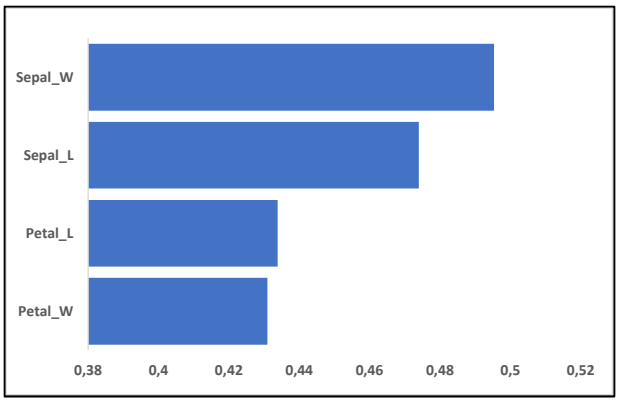
Datasets	Classes	F1	Precision	Recall	Accuracy
Iris	1	1	1	1	0,96
	2	0,96	0,92	1	
	3	0,91	1	0,83	
Avg.		0,97	0,97	0,97	
Dengue	1	0,74	1	0,69	0,78
	2	0,8	0,79	0,95	
	3	0,75	1	0,58	
Avg.		0,73	0,89	0,78	

Figure 1 shows the importance of each feature according to the explainability method for LAMDA algorithms (a), LIME (B) and FI (c). In this figure, we can observe that the feature Sepal Width is the most important in each method, with at least 50% of importance; so, we can conclude that this feature is important for the classification task of the Iris.

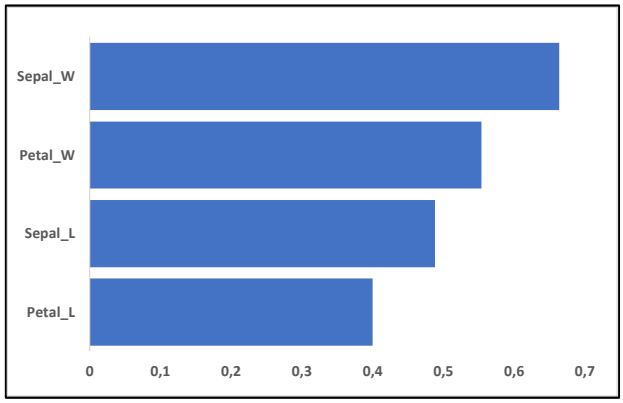
Figure 2 shows the importance of each feature using our explainability method. In this case, we can analyze each class in the iris dataset. We observe that for the Setosa class, the most important feature is Sepal Width, with 48%, and the second place is for Sepal Length, with 45%. For the Versicolor class, the most important features are Sepal Length and Petal Width, with 49% and 48%, respectively. Finally, for the Virginica class, the most important features are Sepal Width and Sepal Length, with 50% and 46%, respectively.

In Figure 3, we show the importance of features for the first 5 features, for the Dengue dataset, according to the explainability method for LAMDA algorithms (a), LIME (B) and feature importance (c). We observe that for LAMDA Algorithm, the first 5 most important features have more than 80%, leading to the conclusion that these features have a great influence in the context of classifications. Additionally, we can see that the two most important features are Hepatomegaly and Hypotension with 95% and 92%, respectively. The most important features according to LIME methods are Hypotension and Low platelets, with 73% and 69%, respectively. Finally, according to Feature importance methods, the two most important features are Hypotension and Fever, with 73% and 68%, respectively. It is to remarkable the fact the Hypotension, Low platelets and Vomit features are there in each method; in this sense, we can conclude that these features have importance in the classification context for the Dengue dataset.

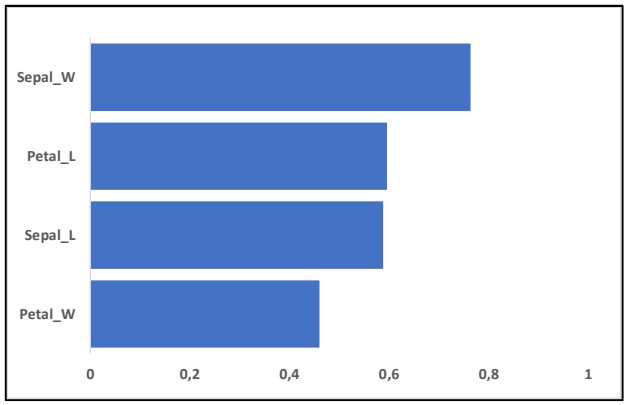
Figure 4 shows the importance of each feature for each class, for the Dengue dataset. In this case, we see the importance of the first five features for this dataset, so, we observe that for the all classes in this dataset, the two most important features are Mucosal bleeding and Hypotension, with a percentage above 80%.



(a) LAMDA



(b) LIME



(c) Feature Importance

FIGURE 1. FEATURE IMPORTANCE FOR THE IRIS DATASET ACCORDING TO DIFFERENT METHODS OF EXPLAINABILITY

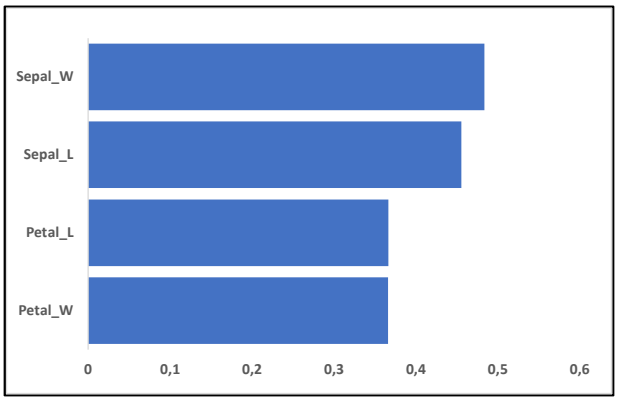
D. Explainability for Clustering tasks

In the clustering context, Table 8 shows the metrics for clustering tasks. We observe that SC is positive and is closer to 0.5. According to the SSW values, we can observe that the clusters formed with LAMDA-RD are compact. Finally, according to the SSB values, we can conclude that the clusters have a good separation.

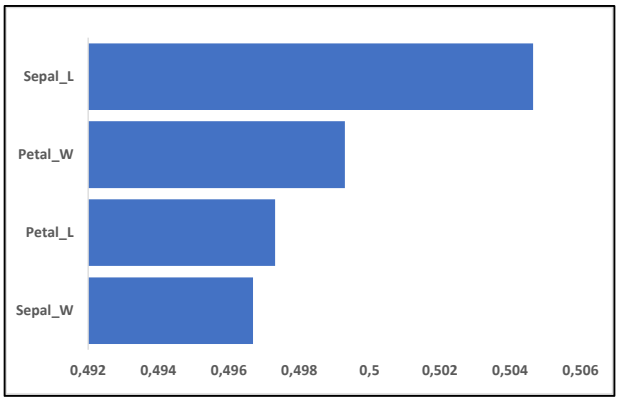
TABLE 8. RESULTS FOR CLUSTERING TASKS FOR EACH DATASET

Dataset/ Metrics	SC	SSW	SSB
Brest Cancer	0,4569	0,03456	0,5789

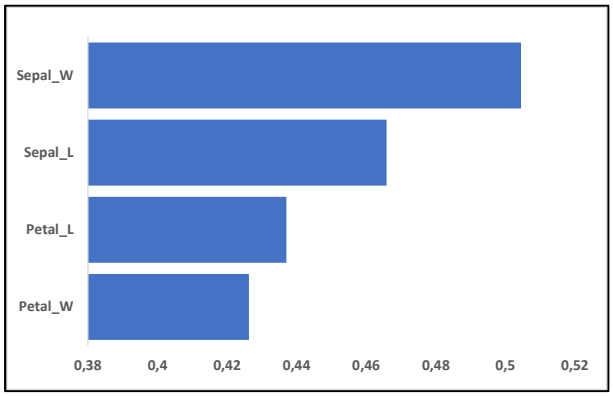
In the context of the clustering tasks, the explainability analysis was only done using our explainability method for the LAMDA algorithms. Figure 5 shows the feature importance for the Cancer dataset. In this case, we show the 5 first most important features. In this sense, the most important features are the presence of the cancer cells and a metric about the presence, with 64% and 62%, respectively.



(a)



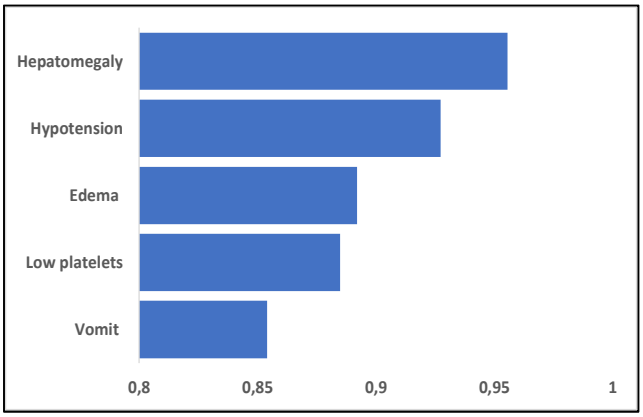
(b)



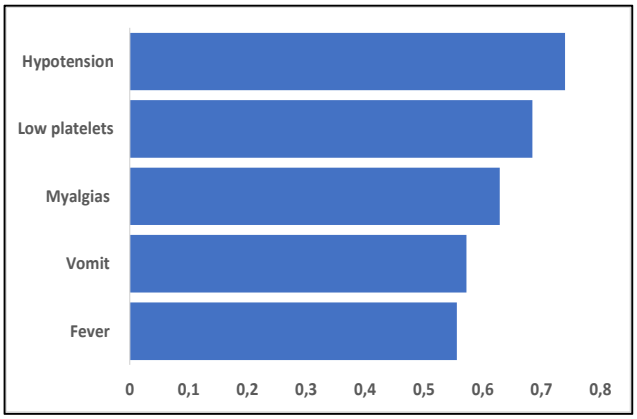
(c)

FIGURE 2.FEATURE IMPORTANCE OF EACH FEATURE BY CLASS FOR THE IRIS DATASET (A) SETOSA, (B) VERSICOLOR, (C) VIRGINICA

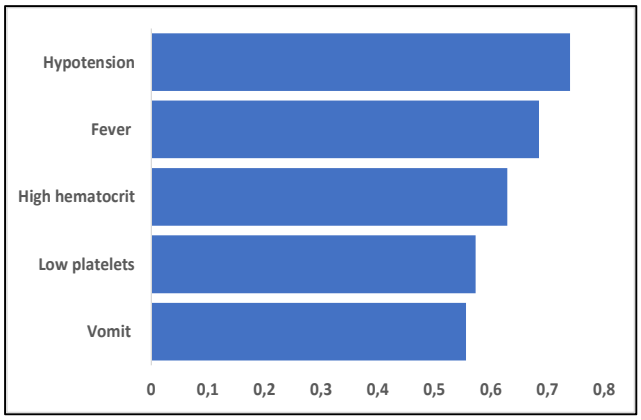
We note that LAMDA-RD formed 4 clusters and Figure 6 shows the importance of features for each cluster. This figure shows the 5 most important features for the Cancer Dataset. For cluster (a), the most important features are Radiotherapy Applied and Presence of the cancer cells, with 87% and 75%, respectively. The rest of the features have above 60%. For cluster (b) are the age and Presence of the cancer cells, with 15% and 14%, respectively. Finally, for clusters (c) and (d), the three most important features are Radiotherapy Applied, Metric about the presence and Presence of the cancer cells.



(a) LAMDA



(b) LIME

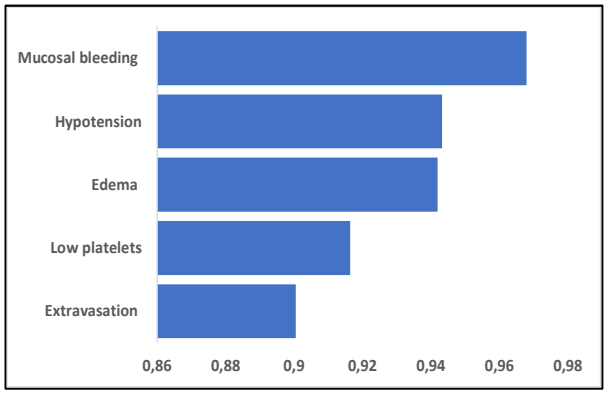


(c) Feature Importance

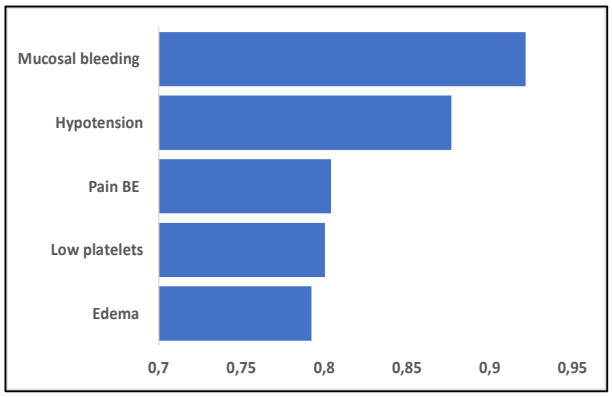
FIGURE 3. FEATURE IMPORTANCE FOR THE DENGUE DATASET ACCORDING TO DIFFERENT METHODS OF EXPLAINABILITY

VI. CONCLUSIONS

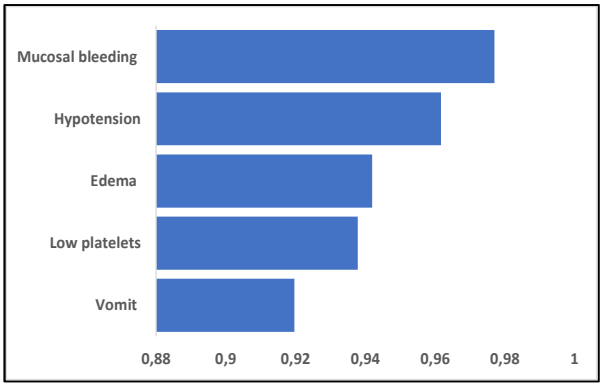
In this work, we have carried out an explainability analysis for the LAMDA family algorithms. For the classification task, we have evaluated the LAMDA-HAD algorithm, and for the clustering task the LAMDA-RD algorithm. We have used two traditional explainability analysis methods, LIME and Feature Importance. Furthermore, we have proposed an explainability method for the LAMDA family algorithms, which is the main contribution of this work. Our explainability method allows measuring the global contribution of each characteristic, but also, the contribution of each characteristic in each cluster/class formed, this being the main difference with respect to the other methods.



(a)



(b)



(c)

FIGURE 4. FEATURE IMPORTANCE OF EACH FEATURE BY CLASS FOR THE DENGUE DATASET

Particularly, since the LAMDA explainability algorithm is based on the degree of membership of each feature in each

class/group, it allows analyzing explainability by class/group, which is a novelty in the context of explainability, since the traditional explainability algorithms typically analyze explainability globally, not by class/group.

Finally, with our method, in general, a ranking of the relevance of characteristics very similar to those obtained with the classic explainability methods was obtained. Future work must make more experiments with other datasets of different dimensions (number of variables) and number of data, and also, more comparisons with other feature-oriented explainability methods.

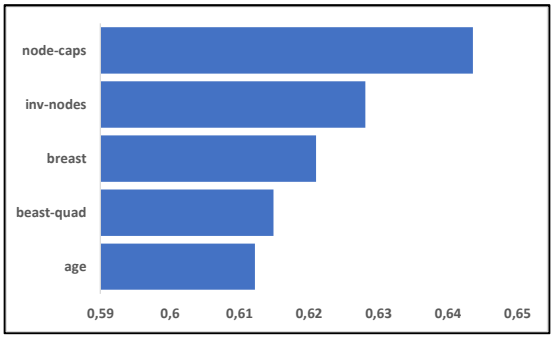
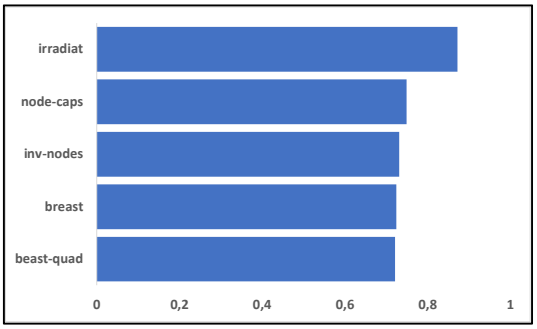
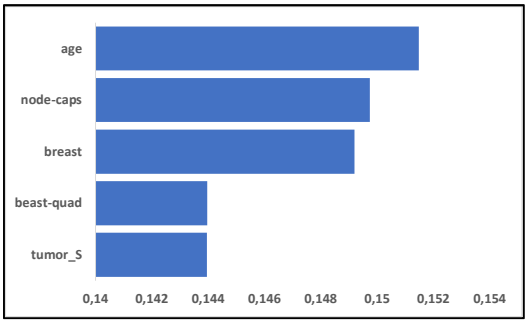


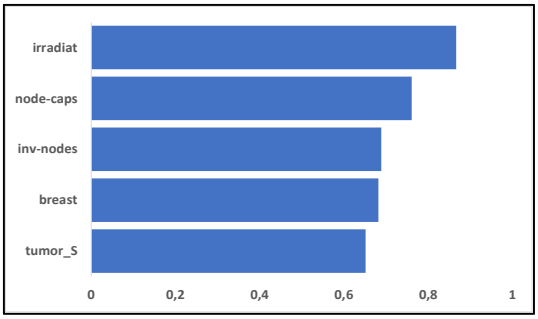
FIGURE 5. FEATURE IMPORTANCE OF EACH FEATURE ACCORDING TO THE LAMDA ALGORITHM FOR THE CANCER DATASET



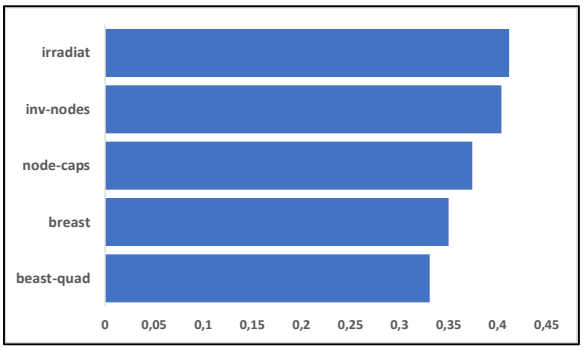
(a)



(b)



(c)



(d)

FIGURE 6.FEATURE IMPORTANCE OF EACH FEATURE ACCORDING TO LAMDA-RD ALGORITHM FOR THE CANCER DATASET

ACKNOWLEDGEMENTS

J. Aguilar work was supported by project TED2021-131264B-I00 (SocialProbing), funded by MICIU/AEI /10.13039/501100011033 and NextGenerationEU/PRTR. The work was supported by project PID2022-140560OB-I00 (DRONAC) funded MICIU/AEI /10.13039/501100011033.

REFERENCES

[1] M. Bücker, G., Szepannek, A.. Gosiewska, P. Biecek. "Transparency, auditability, and explainability of machine learning models in credit scoring". Journal of the Operational Research Society, vol. 73, pp. 70-90, 2022.

[2] R. Roscher, B. Bohn, M. Duarte, J. Garcke. "Explainable machine learning for scientific insights and discoveries". IEEE Access, vol. 8, pp. 42200-42216, 2020.

[3] P.. Lisboa, S. Saralajew, A. Vellido, R. Fernández-Domenech, T. Villmann, "The coming of age of interpretable and explainable machine learning models", Neurocomputing, vol 535, pp. 25-39, 2023.

[4] A. Heuillet, F. Couthouis, F., N. Díaz-Rodríguez, "Explainability in deep reinforcement learning". Knowledge-Based Systems, vol. 214, 2021.

[5] N. Goodwin, S. Nilsson, J. Choong, S. Golden, "Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience", Current Opinion in Neurobiology, vol. 73, 2022.

[6] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods". Entropy, vol. 23, 2020.

[7] J. Waissman, R. Sarrate, T. Escobet, J. Aguilar, B. Dahhou, "Wastewater treatment process supervision by means of a fuzzy automaton model," *Proceedings IEEE International Symposium on Intelligent Control.*, pp. 163-168, 2000.

[8] L. Morales, C Ouedraogo, J Aguilar, J. *et al.* "Experimental comparison of the diagnostic capabilities of classification and clustering algorithms for the QoS management in an autonomic IoT platform" Service Oriented Computing and Applications, vol. 13, pp. 199-219, 2019.

[9] L. Morales, J. Aguilar, D. Chávez, C. Isaza "LAMDA-HAD, an Extension to the LAMDA Classifier in the Context of Supervised Learning", International Journal of Information Technology & Decision Making, vol. 19, pp. 283-316, 2020.

[10] L. Morales, J. Aguilar, "An Automatic Merge Technique to Improve the Clustering Quality Performed by LAMDA," *IEEE Access*, vol. 8, pp. 162917-162944, 2020.

[11] S. Lin, Z. Liang, S. Zhao, S. et al. A comprehensive evaluation of ensemble machine learning in geotechnical stability analysis and explainability. Int J Mech Mater Des, 2023.

[12] J. Brito H. Proença. "A Short Survey on Machine Learning Explainability: An Application to Periocular Recognition". *Electronics*. vol. 10, 2021.

[13] J. Cavaleiro, M. Neves, M. Hewlins A. Jackson 'The photo-oxidation of meso-tetraphenylporphyrins', J. Chem. Soc., Perkin Trans. vol. 1, 1937-1943,1990.