© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other work.

Applying d-XChoquet integrals in classification problems

Jonata Wieczynski
Dept. de Estadística, Informática y Matemáticas
Universidad Pública de Navarra
Pamplona, Spain
jonata.wieczynski@unavarra.es

Giancarlo Lucca, Eduardo Borges
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
{giancarlo.lucca,eduardoborges}@furg.br

Leonardo Emmendorfer
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
leonardoemmendorfer@furg.br

Mikel Ferrero-Jaurrieta
Dept. de Estadística, Informática y Matemáticas
Universidad Pública de Navarra
Pamplona, Spain
mikel.ferrero@unavarra.es

Graçaliz Dimuro
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
gracalizdimuro@furg.br

Humberto Bustince
Dept. de Estadística, Informática y Matemáticas
Universidad Pública de Navarra
Pamplona, Spain
bustince@unavarra.es

***Abstract*—Several generalizations of the Choquet integral have been applied in the Fuzzy Reasoning Method (FRM) of Fuzzy Rule-Based Classification Systems (FRBCS's) to improve its performance. Additionally, to achieve that goal, researchers have searched for new ways to provide more flexibility to those generalizations, by restricting the requirements of the functions being used in their constructions and relaxing the monotonicity of the integral. This is the case of CT-integrals, CC-integrals, CF-integrals, CF1F2-integrals and dCF-integrals, which obtained good performance in classification algorithms, more specifically, in the fuzzy association rule-based classification method for high-dimensional problems (FARC-HD). Thereafter, with the introduction of Choquet integrals based on restricted dissimilarity functions (RDFs) in place of the standard difference, a new generalization was made possible: the d-XChoquet (d-XC) integrals, which are ordered directional increasing functions and, depending on the adopted RDF, may also be a pre-aggregation function. Those integrals were applied in multi-criteria decision making problems and also in a motor-imagery brain computer interface framework. In the present paper, we introduce a new FRM based on the d-XC integral family, analyzing its performance by applying it to 33 different datasets from the literature.**

***Index Terms*—d-XChoquet integral, pre-aggregation functions, OD-increasing functions, Fuzzy Rule-Based Classification System**

I. Introduction

There are many techniques to solve classification problems [1]. It is possible to mention as examples: Support Vector Machines (SVM) [2], Decision Trees [3], Neural Networks (NNs) [4] and Fuzzy Rule-Based Classification Systems (FRBCS's) [5]. The last are considered in this paper and have been applied in different situations, such as, big data [6], image segmentation [7], health [8], anomaly detection [9] and many others.

An important point of an FRBCS is the Fuzzy Reasoning Method (FRM) [10], [11]. The FRM is composed by four different steps, where one of them is responsible to aggregate the information, per class, of the fired rules of the system. To do so, it is applied an aggregation function [12], [13]. Thus, depending on the considered function, the system will perform the final classification in different ways.

The Choquet integral [14] is a well know operator in the field of aggregation functions [12], which was already applied in FRM, as the work by Barrenechea et al. [15], which presented a new FRM that accounts the usage of all given information by the fired fuzzy rules when classifying a new instance. This was achieved by applying the standard Choquet integral as aggregation operator in the process. Following this approach, Lucca et al. [16] introduced the concept of pre-aggregation functions, which lead to the development of several generalizations of the Choquet integral [17], e.g., C_T-, CC-, C_F-, C_{F_1,F_2}- and gC_{F_1,F_2}-integrals [16], [18]–[23]. These advances in generalizing the standard Choquet integral, aiming at restricting the requirements of the functions being used in their constructions and also relaxing the monotonicity of the integral, were applied with success in the FRM of FRBCS. C_T- and CC-integral were also applied to multi-criteria decision making problems [24], [25] and image processing [26].

Then, Bustince et al. [27] introduced a generalization of the Choquet integral by restricted dissimilarity functions (RDF) [28] in place of the standard difference, in order to provide more flexibility and also to allow its use in cases where the difference may not be the best choice. Then, Wieczynski

Supported by Navarra de Servicios y Tecnologías, S.A. (NASERTIC), CNPq (301618/2019-4, 305805/2021-5), FAPERGS (19/2551-0001660-3), the Spanish Ministry of Science and Technology (TIN2016-77356-P, PID2019-108392GB I00 (MCIN/AEI/10.13039/501100011033)).

et al. [29] introduced the dC_F-integrals, a generalization of the Choquet integral by means of two functions, F and RDFs. More recently, Wieczynski et al. [30] have proposed a generalization of the expanded form of the standard Choquet integral using RDFs, called d-XC integrals, with application to multi-criteria decision making problems and also in a motor-imagery brain computer interface framework.

In this work, we aim to study the behavior of d-XC integrals when applied in the FRM of FRBCS, by using 6 different d-XC integrals, analyzing their performance and making a comparison of the results. This is done to improve the performance of the FRBCS when compared to it using CC-integrals.

This work is organized as follows. Section II presents the background theory in respect to the following sections. Section III reviews the definition of the d-XC integral and its main properties. In Section IV, we present the new framework of FRBCS. Then, Section V presents and discuss the results of using 6 different d-XC integrals in the process. Lastly, Section VI is the conclusion thoughts of the work.

II. Background Theory

This section is dedicated to present the preliminary concepts and notations necessary to develop the paper.

A fuzzy set (FS) [31] F on a universe U is given by a membership function $\mu_F : U \to [0,1]$, as $F = \{\langle x, \mu_F(x)\rangle \mid x \in U\}$.

A function $A : [0,1]^n \to [0,1]$ is an aggregation function (AF) [32] if:

- (A1) A is increasing in each argument: for each $i \in \{1,\dots,n\}$, if $x_i \leq y$, then
$$A(x_1,\dots,x_n) \leq A(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n);$$
- (A2) A satisfies the boundary conditions: (i) $A(0,\dots,0) = 0$ and (ii) $A(1,\dots,1) = 1$.

A function $F : [0,1]^n \to [0,1]$ is averaging if and only if (AV) $\forall (x_1,\dots,x_n) \in [0,1]^n$ it holds

$$\min\{x_1,\dots,x_n\} \leq F(x_1,\dots,x_n) \leq \max\{x_1,\dots,x_n\}.$$

Now, we recall the concept of directional monotonicity [33]. Let $\boldsymbol{r} = (r_1,\dots,r_n)$ be a real n-dimensional vector such that $\boldsymbol{r} \neq \boldsymbol{0} = (0,\dots,0)$. A function $F : [0,1]^n \to [0,1]$ is said to be $\boldsymbol{r}$-increasing if, for all $\boldsymbol{x} = (x_1,\dots,x_n) \in [0,1]^n$ and $c > 0$ such that $\boldsymbol{x} + c\mathbf{r} = (x_1 + cr_1,\dots,x_n + cr_n) \in [0,1]^n$, it holds $F(\boldsymbol{x} + c\boldsymbol{r}) \geq F(\boldsymbol{x})$. Similarly, one defines an $\boldsymbol{r}$-decreasing function.

The idea of directional monotonicity induced Lucca et al. [16] to develop the theory of pre-aggregation functions (see also [33]), that is, a function $PA : [0,1]^n \to [0,1]$ is said to be a pre-aggregation function (PAF) if the following conditions hold:

- (PA1) PA is directional increasing, for some $\boldsymbol{r} = (r_1,\dots,r_n) \in [0,1]^n$, $\boldsymbol{r} \neq \boldsymbol{0}$;
- (PA2) PA satisfies the boundary conditions: (i) $PA(0,\dots,0) = 0$ and (ii) $PA(1,\dots,1) = 1$.

We call F an $\boldsymbol{r}$-PAF whenever it is a PAF with respect to a vector $\boldsymbol{r}$.

Another way of thinking in monotonicity of a function is by using the ordered directionally (OD) functions [34]. Consider a function $F : [0,1]^n \to [0,1]$ and let $\boldsymbol{r} = (r_1,\dots,r_n)$ be a real n-dimensional vector, $\boldsymbol{r} \neq \boldsymbol{0}$. F is said to be ordered directionally (OD) $\boldsymbol{r}$-increasing if, for each $\boldsymbol{x} \in [0,1]^n$, any permutation $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$ with $x_{\sigma(1)} \geq \dots \geq x_{\sigma(n)}$, and $c > 0$ such that $1 \geq x_{\sigma(1)} + cr_1 \geq \dots \geq x_{\sigma(n)} + cr_n$, it holds that $F(\boldsymbol{x} + c\boldsymbol{r}_{\sigma^{-1}}) \geq F(\boldsymbol{x})$, where $\boldsymbol{r}_{\sigma^{-1}} = (r_{\sigma^{-1}(1)},\dots,r_{\sigma^{-1}(n)})$. Similarly, one defines an ordered directionally (OD) $\boldsymbol{r}$-decreasing function.

Given that we will be working with integrals we must have a measure, in this case a fuzzy measure, that is, a function $m : 2^N \to [0,1]$ that, for all $X, Y \subseteq N$, have these two properties:

- (m1) m is increasing, that is, if $X \subseteq Y$, then $m(X) \leq m(Y)$;
- (m2) m satisfies the boundary conditions, $m(\emptyset) = 0$, $m(N) = 1$.

This measure is a non additive measure, that is, it is not required to hold the additive property, only an increasing one [35].

The fuzzy measure is the responsible for assigning a relationship value among the elements in the aggregation process performed by the Choquet integral, definition is written now.

Definition 1: [14] Let $m : 2^N \to [0,1]$ be a FM. The discrete Choquet integral (CI) is the function $\mathfrak{C}_m : [0,1]^n \to [0,1]$, defined, for all of $\boldsymbol{x} \in [0,1]^n$, by

$$\mathfrak{C}_m(\boldsymbol{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot m\left(A_{(i)}\right), \qquad (1)$$

where $\left(x_{(1)},\dots,x_{(n)}\right)$ is an increasing permutation on the input $\boldsymbol{x}$, that is, $0 \leq x_{(1)} \leq \dots \leq x_{(n)} \leq 1$, with $x_{(0)} = 0$, and $A_{(i)} = \{(i),\dots,(n)\}$ is the subset of indices corresponding to the $n - i + 1$ largest components of $\boldsymbol{x}$.

One may notice that when the product operation is distributed in Eq. (1), we have the expanded form of the CI (X-CI), which is given by

$$\mathfrak{C}_m(\boldsymbol{x}) = \sum_{i=1}^{n} \left(x_{(i)} \cdot m\left(A_{(i)}\right) - x_{(i-1)} \cdot m\left(A_{(i)}\right)\right). \qquad (2)$$

Other important functions are the restricted dissimilarity functions, introduced by Bustince et al. [36]. A restricted dissimilarity function (RDF) $\delta : [0,1]^2 \to [0,1]$ is a function such that, for all $x, y \in [0,1]$, the following conditions hold:

- (d1) $\delta(x,y) = \delta(y,x)$;
- (d2) $\delta(x,y) = 1$ if and only if $\{x,y\} = \{0,1\}$;
- (d3) $\delta(x,y) = 0$ if and only if $x = y$;
- (d4) if $x \leq y \leq z$, then $\delta(x,y) \leq \delta(x,z)$ and $\delta(y,z) \leq \delta(x,z)$.

From these functions (RDFs), Bustince et al. [27] recently generalized the standard Choquet integral by substituting the difference operator by an RDF, introducing the d-Choquet

TABLE I
RDFs AND RESPECTIVE d-XCHOQUET INTEGRALS [30].

δ	RDF	d-XC
δ_0	$\|x-y\|$	$x_{(1)}+\sum_{i=2}^{n} m\left(A_{(i)}\right)\left(x_{(i)}-x_{(i-1)}\right)$
δ_1	$(x-y)^2$	$x_{(1)}+\sum_{i=2}^{n} m\left(A_{(i)}\right)^2\left(x_{(i)}-x_{(i-1)}\right)^2$
δ_2	$\sqrt{\|x-y\|}$	$x_{(1)}+\sum_{i=2}^{n} \sqrt{m\left(A_{(i)}\right)}\sqrt{x_{(i)}-x_{(i-1)}}$
δ_3	$\|\sqrt{x}-\sqrt{y}\|$	$x_{(1)}+\sum_{i=2}^{n} \sqrt{m\left(A_{(i)}\right)}\left(\sqrt{x_{(i)}}-\sqrt{x_{(i-1)}}\right)$
δ_4	$\|x^2-y^2\|$	$x_{(1)}+\sum_{i=2}^{n} m\left(A_{(i)}\right)^2\left(x_{(i)}^2-x_{(i-1)}^2\right)$
δ_5	$\left(\sqrt{x}-\sqrt{y}\right)^2$	$x_{(1)}+\sum_{i=2}^{n} m\left(A_{(i)}\right)\left(\sqrt{x_{(i)}}-\sqrt{x_{(i-1)}}\right)^2$

integral (d-CI). The discrete d-Choquet integral with respect to a FM $m: 2^N \to [0,1]$ and an RDF $\delta: [0,1]^2 \to [0,1]$ is a mapping $C_{m,\delta}: [0,1]^n \to [0,n]$, defined, for all $\boldsymbol{x} \in [0,1]^n$, by

$$C_{m,\delta}(\boldsymbol{x}) = \sum_{i=1}^{n} \delta\left(x_{(i)}, x_{(i-1)}\right) \cdot m\left(A_{(i)}\right)$$

where $x_{(i)}$, $A_{(i)}$, with $0 \leq i \leq n$, were stated in Def. 1.

The usage of RDFs to generalize the standard Choquet integral overcomes certain drawbacks that the standard difference has, like not being correctly defined in the application domain [27] or even width estimation "errors" when working with interval-valued data [37], [38], which may result in non meaningful information [39].

Some examples of RDFs are presented in Table I. Those are the ones that will be applied to the FRM in the next half of this paper. For construction methods and properties of the RDFs see [28].

III. d-XCHOQUET INTEGRALS

From the expanded standard Choquet integral and the d-Choquet integral, Wieczynski et al. [30] introduced the d-XChoquet integral as in the following definition.

Definition 2: The generalization of the expanded form of the CI by RDFs $\delta: [0,1]^2 \to [0,1]$ with respect to a FM $m: 2^N \to [0,1]$, named d-XChoquet integral (d-XC), is a mapping $X\mathfrak{C}_{\delta,m}: [0,1]^2 \to [0,n]$, defined, for all $\boldsymbol{x} \in [0,1]^n$, by

$$X\mathfrak{C}_{\delta,m}(\boldsymbol{x}) = x_{(1)} + \sum_{i=2}^{n} \delta\left(x_{(i)} \cdot m\left(A_{(i)}\right),\ x_{(i-1)} \cdot m\left(A_{(i)}\right)\right), \quad (3)$$

where $x_{(i)}$, $m\left(A_{(i)}\right)$, with $0 \leq i \leq n$, were stated in Def. 1.

The d-XChoquet integral properties were studied in [30]. In that work the authors have showed the properties for each of the d-XChoquet integrals in Table I.

Notice from [30] that all studied integrals are OD-increasing and greater than the minimum, although only when using δ_0 and δ_1 they are averaging. This is not a drawback for the development of this work, since, in the literature, it was shown that non averaging functions performed better in FRBCS than the averaging ones [22].

Additionally, all of d-XC-integrals of Table I satisfy the $0,1$-conditions **(PA2)**. Four of them (based on δ_0, δ_1, δ_3 and δ_5) are $\mathbf{1}$-increasing, but only with δ_0 and δ_1 they are pre-aggregation functions, since the other two do not present the range in the unit interval.

Lastly, notice that only the d-XChoquet composed by the δ_0, that the standard difference, is increasing and an aggregation function.

IV. ANALYZING THE BEHAVIOR OF THE D-XCHOQUET INTEGRAL IN AN APPLICATION TO FRBCS'S

The d-XChoquet integral was recently applied, with success, in two decision-making problems [30]. The first was in a multi-criteria decision making method that modifies the Group Modular Choquet Random Technical Order by Preference to Ideal Solution (GMC-RTOPSIS) [40] in the aggregation step. The former was in a Brain Computer Interface (BCI) Motor Imagery-based (MI) framework [41], where the d-XChoquet integral was applied in the decision making phase of the algorithm.

In this section, we present the application of the d-XC integral in a classification problem in a Fuzzy Rule-Based Classification System [5]. We start presenting the new Fuzzy Reasoning Method that considers the usage of the new operator. After that, we describe the experimental framework. At the end, the obtained results are described.

A. The new Fuzzy Reasoning Method

In this paper, we consider as fuzzy classifier the Fuzzy Association Rule-based Classification model for High Dimensional Problems (FARC-HD) [42]. This fuzzy classifier is considered a state-of-the-art FRBCS.

The rules used by FARC-HD follows this structure:

$$\text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ then Class is } C_j \text{ with } RW_j,$$

where R_j is the label of the j-th rule, A_{ji} is a fuzzy set representing a linguistic term modeled by a triangular shaped membership function, C_j is the class label, and $RW_j \in [0,1]$ is the rule weight [43], which in this case is computed as the confidence of the fuzzy rule (also known as Certainty Factor) [11] defined by Eq. 4.

$$RW_j = CF_j = \frac{\sum_{x_p \in \ Class \ C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{N} \mu_{A_j}(x_p)}, \quad (4)$$

where N is the number of training patterns $x_p = (x_{p1}, \dots, x_{pm}), p = 1, 2, \dots, N$..

Once the fuzzy rules composing the system have been created, the FRM is responsible for classifying new examples. Specifically, let $x_p = (x_{p1}, \dots, x_{pn})$ be a new example to be classified, L being the number of rules in the rule base, and M being the number of classes of the problem. The new FRM, where the generalizations of the extended Choquet integral are used, consist of 4 different steps:

1) To compute the *matching degree*, that is, the strength of the activation of the if-part of the rules for the example x_p, which is computed using a t-norm $T' : [0,1]^n \to [0,1]$ with the equation

$$\mu_{A_j}(x_p) = T'(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})),$$
$$\text{with } j = 1, \ldots, L.$$

2) *Association degree computation*, that is, for the class of each rule the matching degree is weighted with the corresponding rule weight, given by

$$b_j^k(x_p) = \mu_{A_j}(x_p) \cdot RW_j^k,$$
$$\text{with } k = Class(R_j),\ j = 1, \ldots, L.$$

3) The *example classification soundness degree for all classes* in this step that the aggregation functions are applied to combine the association degrees obtained in the previous step, as follows

$$Y_k(x_p) = X\mathfrak{C}_{\delta,m}\left(b_1^k(x_p), \ldots, b_L^k(x_p)\right), \quad (5)$$
$$\text{with } k = 1, \ldots, M,$$

where $X\mathfrak{C}_{\delta,m}$ is the obtained d-XC integral.
Since, whenever $b_j^k(x_p) = 0$, it holds that

$$X\mathfrak{C}_{\delta,m}\left(b_1^k(x_p), \ldots, b_L^k(x_p)\right) = X\mathfrak{C}_{\delta,m}\left(b_1^k(x_p), \ldots, b_{j-1}^k(x_p), b_{j+1}^k(x_p), \ldots, b_L^k(x_p)\right)$$

then, for practical reasons, only those $b_j^k > 0$ are considered in Equation (5). Moreover, in relation to the fuzzy measure, *m*, we consider the usage of the Power Measure that is defined, for a set A, as follows

$$\mathfrak{m}(A) = \left(\frac{|A|}{n}\right)^q, \text{ with } q > 0. \quad (6)$$

4) A *Classification* decision function $F : [0,1]^M \to \{1, \ldots, M\}$ is applied over the example classification soundness degrees of all classes and thus, the class corresponding to the maximum soundness degree is determined.

$$F(Y_1, \ldots, Y_M) = \min_{k=1,\ldots,M} k \text{ s.t. } Y_k = \max_{w=1,\ldots,M}(Y_w).$$

In practical applications, it is sufficient to consider

$$F(Y_1, \ldots, Y_M) = \arg \max_{k=1,\ldots,M}(Y_k).$$

As it can be observed, in the third step of the FRM we propose to use d-XC integrals, which are associated with a fuzzy measure. According to the results obtained in [15] and [44] we have selected the power measure (PM) where the exponent q is genetically learned.[1]

[1]For more information about the learning process consider [16], [18] and [22].

TABLE II
SUMMARY OF THE DATASETS USED IN THE STUDY.

Id.	Dataset	#Inst.	#Atts.	#Class
App	Appendicitis	106	7	2
Bal	Balance	625	4	3
Ban	Banana	5,300	2	2
Bnd	Bands	365	19	2
Bup	Bupa	345	6	2
Cle	Cleveland	297	13	5
Con	Contraceptive	1,473	9	3
Eco	Ecoli	336	7	8
Gla	Glass	214	9	6
Hab	Haberman	306	3	2
Hay	Hayes-Roth	160	4	3
Ion	Ionosphere	351	33	2
Iri	Iris	150	4	3
Led	led7digit	500	7	10
Mag	Magic	1,902	10	2
New	Newthyroid	215	5	3
Pag	Pageblocks	5,472	10	5
Pen	Penbased	10,992	16	10
Pho	Phoneme	5,404	5	2
Pim	Pima	768	8	2
Rin	Ring	740	20	2
Sah	Saheart	462	9	2
Sat	Satimage	6,435	36	7
Seg	Segment	2,310	19	7
Shu	Shuttle	58,000	9	7
Son	Sonar	208	60	2
Spe	Spectfheart	267	44	2
Tit	Titanic	2,201	3	2
Two	Twonorm	740	20	2
Veh	Vehicle	846	18	4
Win	Wine	178	13	3
Wis	Wisconsin	683	11	2
Yea	Yeast	1,484	8	10

B. Experimental Framework

To demonstrate the efficiency and the quality of the proposal, we have conducted a study using 33 different datasets. We highlight that these datasets were selected from KEEL dataset repository [45]. Also, these datasets are the same used in previous studies (see [20], [22] and [44]).

The datasets used in the study are summarized in Table II. Per dataset, we present the corresponding identification (Id), the number of instances (#Inst), attributes (#Atts), and classes (#Class).

Following the idea of previous generalizations, the results are presented taking into account a 5-fold cross-validation procedure [46]. To analyze the classifier performance, we use the accuracy [46]. Consequently, the results presented in this study are related to the average accuracy obtained in the five different folds.

As mentioned before, the fuzzy classifier used in this paper is the FARC-HD, therefore, the configuration of this classifier follows the original author's suggestion. That is, the product t-norm as conjunction operator, the certainty factor is the RW, with 0.05 as minimum support, the threshold for the confidence as 0.8, the depth of the tree is 3, and k_t is 2.

In relation to the parameters used by the genetic algorithm to the q exponent (used by the Power Measure), we have a population composed by 50 individuals, the gray codification consider 30 bits per gen and the maximum number of evaluations is 20.000.

V. Experimental Results

In this section, we present the obtained results in training and test in relation to the extended standard Choquet integral (with respect to δ_0) and the d-XC integrals (see Table I). After that, a statistical study is conducted to perform a deeper analysis on the achieved results.

In Table III, the results obtained by the FRM in which the generalizations were applied are presented. In this table, the rows are related with different datasets, the columns show the results obtained in Training (Tra.) and Testing (Tst.) related to different generalizations, and each cell of this table is related to the accuracy mean.

In order to ease the comprehension of the obtained results, we highlight the highest and lowest achieved results for each set (Tra. and Tst). Precisely, considering the results obtained exclusively in training, we highlight with $^+$ the cases that the accuracy is superior among all training results for a determined dataset. On the other hand, we highlight with $^-$ the cases where the accuracy presents the lowest performance. Similarly, we perform this analysis to the test set. If the generalization achieves the highest mean for a dataset, we highlight it in **boldface**, and we underline it if it has the lowest mean.

At the end of the table, we provide for training and test the obtained mean (among the 33 datasets) for each dissimilarity. Moreover, for training and test, the count of the highest achieved results is presented in #nMax, while the count of lowest performance is shown in #nMin.

Analyzing the results of training in Table III, its noticeable that the model obtained a similar mean (around 89% of accuracy) for the dissimilarities δ_0, δ_2, δ_3 and δ_5. Furthermore, the generalizations considering the δ_1 and δ_4 achieved the lowest mean.

The largest performance mean is obtained by the δ_3 approach, but the number of datasets where the generalizations achieved the highest performance is bigger when is considered the δ_5. On the other hand, considering the cases where the trained model had a low performance, the δ_4 have just 4 datasets, while δ_5 have 6.

Taking into consideration the results obtained in the test, it is noticeable that the generalization presenting the highest performance is the δ_5. In fact, this approach obtained the highest general mean and the largest number of datasets (12 cases). Only in 4 datasets this model did not achieve a good performance. Thus, it is noticeable that results for this generalization can be considered as a good alternative to other approaches, like the CC-integrals [18], since it presents a good number of high performance, just a few bad cases, and also presented stable results, neither high nor low.

A similar analysis can be made with δ_2. By using it, the method achieved the second-highest mean and also good results in 9 different datasets, presenting the same number of bad results as δ_2. The extended Choquet integral (δ_0) and the generalization δ_3 present the same number of cases where the results are superior (5 times), however the δ_3 have a bigger mean and in just 2 cases presented a low performance (3 times in the δ_0). The remaining generalizations, δ_1 and δ_4 were the ones presenting an unsatisfactory performance in testing.

Just analyzing the obtained mean of the different generalizations and the cases where these functions present good or bad performance may not be enough to conclude if the generalizations could be an alternative to the extended Choquet integral. Thus, we have conducted a statistical test to directly compare the δ_0 against the generalizations and provide a more robust analysis.

Knowing that the conditions of parametric tests are not satisfied, we consider the usage of non-parametric tests [47]–[49]. Precisely, we have performed a pair-wise comparison using the Wilcoxon signed-rank test [50].

The results of the statistical test are available in Table IV. In this table, R^+ represents the rank obtained by the method δ_0 (which forms the expanded standard Choquet integral), R^- is the rank related with the five different generalizations by dissimilarities. In the last row, we present the obtained p-value. When considering a confidence level of 90%, that is, the methods are considered statistically different if the obtained p-value is lower than 0.10.

The result obtained in the statistical test demonstrates that for three different generalizations of the extended Choquet integral, δ_2, δ_3 and δ_5 the results are statistically superior to the usage of the original approach. This reinforces that the usage of these generalizations can be considered as a good alternative to aggregate the information in the FRM and deal with classification problems.

VI. Conclusion

In this paper we presented an application of the d-XChoquet integral, which is a new generalization of the expanded standard Choquet integral by using restricted dissimilarity functions, to the Fuzzy Reasoning Method of the Fuzzy Rule-Based Classification System.

The new FRM was applied in 33 different datasets from the literature, that resulted in the d-XChoquet integral with better performance being the one composed by the RDF δ_5. Thereafter, the statistical test have shown that 3 out of the 6 d-XChoquet integrals, namely the ones using δ_2, δ_3 and δ_5 RDFs are statically superior to using the standard Choquet integral (with δ_0).

Therefore, the d-XChoquet integral can be a good aggregation function to use in the FRM and with classification problems.

Lastly, for future works we intend to comparison to other generalizations of the Choquet integral.

References

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[2] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[3] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, p. 81–106, mar 1986. [Online]. Available: https://doi.org/10.1023/A:1022643204877

TABLE III

ACCURACY MEAN OBTAINED IN TRAINING (TRA.) AND TESTING (TST.) BY THE EXTENDED CHOQUET INTEGRAL (δ_0) AND ITS GENERALIZATIONS

	δ_0		δ_1		δ_2		δ_3		δ_4		δ_5	
Dataset	**Tra.**	**Tst.**	**Tra.**	**Tst.**	**Tra.**	**Tst.**	**Tra.**	**Tst.**	**Tra.**	**Tst.**	**Tra.**	**Tst.**
App	$^{+}$96.70	80.13	96.46	83.98	$^{-}$95.75	82.08	96.22	84.89	96.46	81.13	96.22	**85.84**
Bal	89.32	82.40	$^{-}$85.48	76.16	$^{+}$92.16	**88.48**	91.64	86.40	89.04	76.80	92.12	87.04
Ban	$^{+}$87.52	**86.32**	85.46	83.87	86.59	85.28	87.43	86.13	86.83	85.25	$^{-}$84.48	83.23
Bnd	89.79	68.56	89.73	67.61	89.45	71.30	89.73	68.59	$^{+}$89.80	68.49	$^{-}$88.56	**71.82**
Bup	81.52	**66.96**	81.23	66.67	$^{-}$80.51	61.16	81.52	64.93	$^{+}$81.67	63.48	80.72	65.51
Cle	89.98	55.58	90.07	55.20	89.73	57.23	$^{+}$90.15	56.56	89.90	55.21	$^{-}$89.31	**58.92**
Com	61.64	51.26	$^{-}$60.32	48.95	62.22	**53.77**	$^{+}$63.03	53.09	60.98	50.92	62.88	52.95
Eco	91.52	76.51	$^{-}$88.99	75.92	92.63	**81.26**	92.63	79.17	89.88	76.80	$^{+}$92.93	77.99
Gla	82.95	64.02	$^{-}$82.71	64.01	84.58	65.44	85.75	**67.76**	82.71	61.69	$^{+}$86.10	65.44
Hab	83.17	72.52	$^{+}$83.50	**76.78**	$^{-}$82.43	71.54	82.60	72.85	83.33	72.54	83.09	70.56
Hay	$^{-}$91.10	**79.49**	$^{-}$91.10	78.01	$^{+}$91.28	**79.49**	$^{-}$91.10	78.75	$^{-}$91.10	78.75	$^{+}$91.28	78.69
Ion	99.00	90.04	$^{+}$99.36	87.47	99.00	**90.89**	99.22	87.76	99.29	88.61	98.86	90.03
Iri	99.00	91.33	99.00	92.00	99.00	**94.67**	$^{-}$98.83	93.33	99.00	93.33	$^{+}$99.17	93.33
Led	75.25	68.20	$^{-}$74.60	68.60	$^{+}$75.90	68.00	74.85	67.80	75.25	**69.00**	75.65	68.60
Mag	84.08	78.86	$^{-}$83.20	76.76	$^{+}$85.28	**80.02**	84.91	79.44	84.48	79.60	84.74	79.44
New	98.95	94.88	$^{-}$97.91	92.56	$^{+}$99.77	**97.67**	98.84	94.88	98.26	91.63	$^{+}$99.77	96.28
Pag	97.03	94.16	97.08	93.97	97.54	94.34	$^{-}$96.99	93.97	97.17	94.16	$^{+}$97.67	**95.07**
Pen	97.73	90.55	$^{-}$95.57	86.91	98.57	92.18	98.16	90.36	96.32	89.64	$^{+}$98.61	**92.45**
Pho	$^{+}$84.96	**82.98**	$^{-}$83.63	81.99	84.11	82.07	84.93	82.59	84.35	81.31	83.74	81.50
Pim	84.99	73.95	84.41	71.74	$^{-}$84.28	75.13	$^{+}$85.84	73.44	85.22	73.18	85.19	**76.43**
Rin	96.39	90.95	$^{-}$92.26	82.97	$^{+}$97.20	90.95	96.72	90.27	96.22	89.46	97.03	**91.08**
Sah	85.12	69.69	85.39	67.76	83.98	69.04	85.33	69.26	$^{+}$85.44	66.66	$^{-}$83.77	**70.54**
Sat	85.34	79.47	83.86	76.98	85.73	79.01	$^{+}$86.55	**79.78**	$^{-}$83.67	78.69	85.34	79.63
Seg	95.16	**93.46**	$^{-}$92.97	90.39	95.27	**93.46**	$^{+}$95.55	92.47	93.63	90.65	95.45	92.51
Shu	97.71	97.61	97.51	97.33	$^{-}$97.17	97.06	$^{+}$99.11	**98.99**	97.89	97.66	97.66	97.06
Son	99.76	77.43	$^{-}$99.28	77.39	$^{+}$99.76	82.73	$^{+}$99.76	78.40	99.64	75.99	99.52	**84.16**
Spe	$^{-}$94.01	77.88	$^{+}$95.04	80.14	94.29	77.51	94.57	79.39	94.38	**81.99**	94.29	80.52
Tit	$^{-}$79.07	78.87	$^{-}$79.07	78.87	$^{-}$79.07	78.87	$^{-}$79.07	78.87	$^{-}$79.07	78.87	$^{-}$79.07	78.87
Two	96.52	84.46	$^{-}$94.46	78.24	98.18	89.19	97.60	84.73	94.97	80.81	$^{+}$98.28	**91.08**
Veh	$^{+}$81.03	68.44	$^{-}$78.55	66.55	79.99	67.97	80.67	**69.38**	79.08	65.02	79.17	66.79
Win	99.86	93.79	$^{-}$99.58	93.83	$^{+}$100.00	96.08	99.86	94.37	99.72	92.10	$^{+}$100.00	**97.19**
Wis	$^{+}$99.08	97.22	$^{-}$98.39	94.88	98.83	96.93	99.01	96.49	99.05	95.90	98.76	**97.36**
Yea	63.54	55.73	$^{-}$63.21	55.12	64.37	56.94	64.12	**57.01**	63.53	55.73	$^{+}$65.09	56.74
Mean	89.05	79.20	$^{-}$88.16	77.87	89.23	80.23	$^{+}$89.46	79.76	88.71	78.21	89.23	**80.44**
#nMax	5	5	3	1	8	9	7	5	3	2	10	12
#nMin	3	3	19	18	6	4	4	2	3	8	6	4

TABLE IV

WILCOXON TEST TO COMPARE THE EXTENDED CHOQUET INTEGRAL (δ_0) AGAINST DIFFERENT GENERALIZATIONS

		δ_1	δ_2	δ_3	δ_4	δ_5
δ_0	R^+:	457.5	145	192.5	445.5	131.5
	R^-:	103.5	416	368.5	115.5	429.5
	p-value:	≤ 0.00	≤ 0.00	0.10	≤ 0.00	≤ 0.00

[4] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. USA: Prentice Hall PTR, 1998.

[5] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining*, ser. Advanced Information Processing. Berlin: Springer, 2005.

[6] P. G. da S. E. Tuy and T. Nogueira Rios, "Summarizer: Fuzzy rule-based classification systems for vertical and horizontal big data," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–8.

[7] H. Leon-Garza, H. Hagras, A. Peña-Rios, A. Conway, and G. Owusu, "A fuzzy rule-based system using a patch-based approach for semantic segmentation in floor plans," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2021, pp. 1–6.

[8] I.-O. Lixandru-Petre, "A fuzzy system approach for diabetes classification," in *2020 International Conference on e-Health and Bioengineering (EHB)*, 2020, pp. 1–4.

[9] L. Decker, D. Leite, L. Giommi, and D. Bonacorsi, "Real-time anomaly detection in data centers for log-based predictive maintenance using an evolving fuzzy-rule-based approach," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–8.

[10] O. Cordon, M. J. del Jesus, and F. Herrera, "Analyzing the reasoning mechanisms in fuzzy rule based classification systems," *Mathware and Soft Computing*, vol. 5, no. 2-3, pp. 321 – 332, 1998.

[11] O. Cordón, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, no. 1, pp. 21 – 45, 1999.

[12] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin: Springer, 2007.

[13] M. Grabisch, J. Marichal, R. Mesiar, and E. Pap, *Aggregation Functions*. Cambridge: Cambridge University Press, 2009.

[14] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1953–1954.

[15] E. Barrenechea, H. Bustince, J. Fernandez, D. Paternain, and J. A. Sanz, "Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems," *Axioms*, vol. 2, no. 2, pp. 208–223, 2013.

[16] G. Lucca, J. Sanz, G. Pereira Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, and H. Bustince Sola, "Pre-aggregation functions: construction and an application," *IEEE Transactions on Fuzzy Systems*,

vol. 24, no. 2, pp. 260–272, April 2016.

[17] G. P. Dimuro, J. Fernández, B. Bedregal, R. Mesiar, J. A. Sanz, G. Lucca, and H. Bustince, "The state-of-art of the generalizations of the Choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions," *Information Fusion*, vol. 57, pp. 27 – 43, 2020.

[18] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, M. J. Asiain, M. Elkano, and H. Bustince, "CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems," *Knowledge-Based Systems*, vol. 119, pp. 32 – 43, 2017.

[19] G. P. Dimuro, G. Lucca, J. A. Sanz, H. Bustince, and B. Bedregal, "Cmin-integral: A Choquet-like aggregation function based on the minimum t-norm for applications to fuzzy rule-based classification systems," in *Aggregation Functions in Theory and in Practice*, V. Torra, R. Mesiar, and B. D. Baets, Eds. Cham: Springer International Publishing, 2018, pp. 83–95.

[20] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince, and R. Mesiar, "CF-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems," *Information Sciences*, vol. 435, pp. 94 – 110, 2018.

[21] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, and H. Bustince, "A proposal for tuning the the α parameter in $C_\alpha C$-integrals for application in fuzzy rule-based classification systems," *Natural Computing*, vol. 19, p. 533–546, 2020.

[22] G. Lucca, G. P. Dimuro, J. Fernandez, H. Bustince, B. Bedregal, and J. A. Sanz, "Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-integrals named $C_{F_1F_2}$-integrals," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 124–134, Jan 2019.

[23] G. P. Dimuro, G. Lucca, B. Bedregal, R. Mesiar, J. A. Sanz, C.-T. Lin, and H. Bustince, "Generalized $C_{F_1F_2}$-integrals: From choquet-like aggregation to ordered directionally monotone functions," *Fuzzy Sets and Systems*, vol. 378, pp. 44 – 67, 2020.

[24] J. C. Wieczynski, G. P. Dimuro, E. N. Borges, H. S. Santos, G. Lucca, R. Lourenzutti, and H. Bustince, "Generalizing the GMC-RTOPSIS method using CT-integral pre-aggregation functions," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Los Alamitos: IEEE, 2020, pp. 1–8.

[25] J. Wieczynski., G. Lucca., E. Borges., G. Dimuro., R. Lourenzutti., and H. Bustince, "CC-separation measure applied in business group decision making," in *Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 1: ICEIS*. SciTePress, 2021, pp. 452–462.

[26] C. Marco-Detchart, G. Lucca, C. Lopez-Molina, L. De Miguel, G. Pereira Dimuro, and H. Bustince, "Neuro-inspired edge feature fusion using Choquet integrals," *Information Sciences*, vol. 581, pp. 740–754, 2021.

[27] H. Bustince, R. Mesiar, J. Fernandez, M. Galar, D. Paternain, A. Altalhi, G. Dimuro, B. Bedregal, and Z. Takáč, "d-Choquet integrals: Choquet integrals based on dissimilarities," *Fuzzy Sets and Systems*, vol. 414, pp. 1–27, 2021.

[28] H. Bustince, E. Barrenechea, and M. Pagola, "Relationship between restricted dissimilarity functions, restricted equivalence functions and normal EN-functions: Image thresholding invariant," *Pattern Recognition Letters*, vol. 29, no. 4, pp. 525 – 536, 2008.

[29] J. C. Wieczynski, G. Lucca, G. P. Dimuro, E. N. Borges, J. A. Sanz, T. C. Asmus, J. Fernández, and H. Bustince, "dC_F-integrals: generalizing C_F-integrals by means of restricted dissimilarity functions," (submitted).

[30] J. C. Wieczynski, J. Fumanal-Idocin, G. Lucca, E. N. Borges, T. C. Asmus, L. C. Emmendorfer, H. Bustince, and G. P. Dimuro, "d-XC integrals: On the generalization of the expanded form of the choquet integral by restricted dissimilarity functions and their applications," *Transactions on Fuzzy Systems*, (In press).

[31] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[32] G. Beliakov, H. Bustince, and T. Calvo, *A Practical Guide to Averaging Functions*. Berlin, New York: Springer, 2016.

[33] H. Bustince, R. Mesiar, A. Kolesárová, G. Dimuro, J. Fernandez, I. Diaz, and S. Montes, "On some classes of directionally monotone functions," *Fuzzy Sets and Systems*, vol. 386, pp. 161 – 178, 2020.

[34] H. Bustince, E. Barrenechea, M. Sesma-Sara, J. Lafuente, G. P. Dimuro, R. Mesiar, and A. Kolesárová, "Ordered directionally monotone functions: Justification and application," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2237–2250, 2017.

[35] T. Murofushi, M. Sugeno, and M. Machida, "Non-monotonic fuzzy measures and the Choquet integral," *Fuzzy Sets and Systems*, vol. 64, no. 1, pp. 73 – 86, 1994.

[36] H. Bustince, A. Jurio, A. Pradera, R. Mesiar, and G. Beliakov, "Generalization of the weighted voting method using penalty functions constructed via faithful restricted dissimilarity functions," *European Journal of Operational Research*, vol. 225, no. 3, pp. 472 – 478, 2013.

[37] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. Philadelphia: SIAM, 2009.

[38] A. Rauh, M. Kletting, H. Aschemann, and E. P. Hofer, "Reduction of overestimation in interval arithmetic simulation of biological wastewater treatment processes," *Journal of Computational and Applied Mathematics*, vol. 199, no. 2, pp. 207–212, 2007, special Issue on Scientific Computing, Computer Arithmetic, and Validated Numerics (SCAN 2004).

[39] T. da Cruz Asmus, G. Pereira Dimuro, B. Bedregal, J. A. Sanz, R. Mesiar, and H. Bustince, "Towards interval uncertainty propagation control in bivariate aggregation processes and the introduction of width-limited interval-valued overlap functions," *Fuzzy Sets and Systems*, 2021, (In press, Corrected Proof, available online 10 September 2021).

[40] R. Lourenzutti, R. A. Krohling, and M. Z. Reformat, "Choquet based TOPSIS and TODIM for dynamic and heterogeneous decision making with criteria interaction," *Information Sciences*, vol. 408, pp. 41 – 69, 2017.

[41] J. Fumanal-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernández, J. A. Sanz, and H. Bustince, "Motor-imagery-based brain-computer interface using signal derivation and aggregation functions," *IEEE Transactions on Cybernetics*, pp. 1–12, 2021, (In press, Early access).

[42] J. Alcala-Fdez, R. Alcala, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 857–872, 2011.

[43] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *Fuzzy Systems, IEEE Transactions on*, vol. 9, no. 4, pp. 506–515, 2001.

[44] G. Lucca, J. A. Sanz, G. P. Dimuro, E. N. Borges, H. Santos, and H. Bustince, "Analyzing the performance of different fuzzy measures with generalizations of the choquet integral in classification problems," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, June 2019, pp. 1–6.

[45] J. Alcalá-Fdez, L. Sánchez, S. García, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, and F. Herrera, "Keel: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.

[46] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[47] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[48] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, 2nd ed. Chapman & Hall/CRC, 2006.

[49] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability," *Soft Computing*, vol. 13, no. 10, pp. 959–977, 2009.

[50] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.