Interval-valued aggregation functions based on moderate deviations applied to Motor-Imagery-Based Brain Computer Interface

Javier Fumanal-Idocin, Zdenko Takáč, Javier Fernández, Member, IEEE, Jose Antonio Sanz, Harkaitz Goyena, Ching-Teng Lin, Fellow, IEEE, Yu-Kai Wang, Member, IEEE, Humberto Bustince, Senior, IEEE

Abstract—In this work we develop moderate deviation functions to measure similarity and dissimilarity among a set of given interval-valued data to construct interval-valued aggregation functions, and we apply these functions in two Motor-Imagery Brain Computer Interface (MI-BCI) systems to classify electroencephalography signals. To do so, we introduce the notion of interval-valued moderate deviation function and, in particular, we study those interval-valued moderate deviation functions which preserve the width of the input intervals. In order to apply them in a MI-BCI system, we first use fuzzy implication operators to measure the uncertainty linked to the output of each classifier in the ensemble of the system, and then we perform the decision making phase using the new interval-valued aggregation functions. We have tested the goodness of our proposal in two MI-BCI frameworks, obtaining better results than those obtained using other numerical aggregation and interval-valued OWA operators, and obtaining competitive results versus some non-aggregation-based frameworks.

Index Terms—Electroencephalography; Brain-Computer-Interface; Moderate Deviations; Interval-valued aggregation; Motor Imagery; Admissible orders; Classification; Signal Processing;

I. INTRODUCTION

Brain Computer Interface (BCI) is one of the most popular methods for controlling devices using variations in the brain dynamics [1], [2], [3]. One popular BCI method is Motor-Imagery (MI), in which a person imagines a specific body movement, which produces a reaction in the motor areas of the brain [4], [5]. BCI systems are composed of some different components, such as signal detection, feature extraction and command identification, in order to successfully convert a brain signal into a computer command [6].

Usually, BCI systems use different wave transformations to extract useful information from the ElectroEncephaloGrahphy (EEG) data [7], [8], such as the Fast Fourier transform (FFT), to convert the signal in the frequency domain and the Meyer Wavelet transform. It is also very common to use algorithms such as Common Spatial Filtering (CSP), to classify the signals or to use its output as features to feed further classifiers [5], [9]. Some of the most common classifiers used in BCI systems are Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machines (SVM) and K-Nearest Neighbours (KNN) [10], [11].

In the literature, there are many different approaches to EEG-based BCI classification systems. In [12], the authors extended the CSP to regression problems using fuzzy sets, and applied it to measure responsiveness in psychomotor vigilance tasks. In [13], the authors studied the correlation between different EEG channels and the target class, in order to select only meaningful channels to the classification problem, while in [14] the authors purge the outliers from the signal and then use Dempster-Shafer theory to discover features with the highest interclass variability. In [15] the authors studied the effects on visual stimuli, in order to understand how human perceive other people’s emotions in the cocktail party problem [16]. Also, in [17], the authors used Bispectrum analysis [18] to select the optimal channels to perform classification.

One recent approach to BCI research is to focus on the information fusion processes of the system [19], [20], [21]. Due to the high number of components of the BCI, it is necessary to combine the output from different elements into a single numerical value. This process is key to the performance of the system due to the relevance of these components interactions and correlations. One possibility to deal with this problem are aggregation functions [22], [23].

Aggregation functions are used to fuse several input values into one single output value. They have been widely applied in classification systems [24], [25], fuzzy rules-based systems [26], [27] and image processing [28], [29], among others.

In some cases, there is imprecision in the data to aggregate. For the case of the EEG signal, the presence of noise and imprecision in the measurements can significantly affect the performance of the BCI system [7]. One solution to model that uncertainty is to represent each data as an interval, where its width represents the uncertainty associated to each observation.
The use of intervals has shown to be a suitable solution to tackle classification problems \([32], [33], [34]\). For this reason, large efforts have been devoted to the development of mechanisms to fuse information in the interval-valued setting \([35], [36], [37], [38]\).

Taking into account these considerations, the objective of this paper is double:

- To construct a new MI BCI framework to classify EEG signals where the uncertainty in each classifier output is modeled using interval-valued data.
- To determine the best aggregation function to be applied to the set of interval-valued data to obtain the final decision.

The selection of the best aggregation function in an interval-valued setting is still an open problem. In the case of numerical data, several solutions have been proposed for this problem. In particular, in this work we consider the following ones: (i) Penalty-based aggregation functions, which determine the output from a set of inputs by minimizing a disagreement measure between the original set of values and the possible outputs \([39], [40]\). (ii) Deviation-based aggregation functions that were introduced in \([41]\) based on Darczy’s deviation functions \([42]\), which aggregate a set of deviation functions to determine how different is a given value from a set of inputs.

To reach our objective, we first develop the theoretical concept of interval-valued moderate deviation based aggregation function, studying the special case where the width of the input intervals is the same for all of them. Then, using the newly-developed interval-valued aggregation functions, we extend two aggregation-based MI BCI frameworks, namely: the traditional framework described in \([20]\) and the Multimodal Fuzzy Fusion (MFF) framework proposed in \([19]\).

The goodness of our proposal is shown by comparing our results (i) with the outcome obtained by its numerical counterpart using classical aggregation functions, (ii) with the new method using as aggregation function the interval-valued OWA operators proposed in \([43]\) and (iii), with other non aggregation-based MI BCI frameworks \([44], [45], [46]\).

The structure of this paper is described as follows. In Section II, we explain some preliminary concepts related to the developed work. In Section III, we discuss the interval-valued moderate deviations, and, in Section IV we discuss the specific case in which the length of all the interval-valued inputs are the same. Then, in Section V we explain how to apply these functions in a MI BCI framework and, in Section VI we show our experimental results and comparisons with other aggregation functions. Subsequently, in Section VII we show how our system performs compared to non aggregation-based MI BCI frameworks. Finally, in Section VIII we summarize the work done and give the final remarks.

II. PRELIMINARIES

In this section we introduce the concept of aggregation function, fuzzy implication function, OWA operator and their interval-valued version.

A. Aggregation functions

Aggregation functions \([23]\) are used to fuse information from \(n\) sources into one single output. A function \(A: [0, 1]^n \rightarrow [0, 1]\) is said to be an \(n\)-ary aggregation function if the following conditions hold:

- \(A\) is increasing in each argument: \(\forall x_i, y \in [0, 1], i \in \{1, \ldots, n\} : x_i \leq y \implies A(x_1, x_2, \ldots, x_n) \leq A(x_1, x_2, \ldots, x_n)\)
- \(A(0, \ldots, 0) = 0\)
- \(A(1, \ldots, 1) = 1\)

Some examples of classical \(n\)-ary aggregation functions are:

- Arithmetic mean: \(A(x) = \frac{1}{n} \sum_{i=1}^{n} x_i\);
- Median: \(A(x) = x_m\), where for any permutation \(\sigma\) : \(\{1, \ldots, n\}\) such that \(x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}\), \(x_m = x_{\sigma(\frac{n+1}{2})}\), if \(n\) is odd, and \(x_m = \frac{1}{2} (x_{\sigma(\frac{n}{2})} + x_{\sigma(\frac{n}{2} + 1)})\), otherwise.
- Max: \(A(x) = \max(x_1, x_2, \ldots, x_n)\);
- Min: \(A(x) = \min(x_1, x_2, \ldots, x_n)\);

where \(x = (x_1, \ldots, x_n) \in [0, 1]^n\).

B. Interval-valued fuzzy implication functions

An fuzzy implication function is a function \(I: [0, 1]^2 \rightarrow [0, 1]\) that satisfies the following properties, for all \(x, z, y \in [0, 1]\) \(\([47]\)\):

- \(x \leq z\) implies \(I(x, y) \geq I(z, y)\);
- \(y \leq t\) implies \(I(x, y) \leq I(x, t)\);
- \(I(0, x) = 1\) for all \(x \in [0, 1]\);
- \(I(x, 1) = 1\) for all \(x \in [0, 1]\);
- \(I(1, 0) = 0\).

Examples of fuzzy implication functions are:

- Kleene-Dienes: \(I(x, y) = \max(1 - x, y)\)
- Łukasiewicz: \(I(x, y) = \min(1, 1 - x + y)\)
- Reichenbach: \(I(x, y) = 1 - x + xy\)

where \(x, y \in [0, 1]\).

C. Interval-valued aggregation functions \([48]\)

We consider closed subintervals of the unit interval \([0, 1]\):

\[
L([0, 1]) = \{X = [\underline{X}, \overline{X}] \mid 0 \leq \underline{X} \leq \overline{X} \leq 1\}. 
\]

The width of the interval \(X \in L([0, 1])\), denoted by \(w(X)\), is given by \(w(X) = \overline{X} - \underline{X}\). An interval-valued function \(f: L([0, 1])^n \rightarrow L([0, 1])\) is called \(w\)-preserving if \(w(f(X_1, \ldots,X_n)) = w(f(X_1, \ldots,X_n))\) for all \(X_1, \ldots, X_n \in L([0, 1]).\)

An order relation on \(L([0, 1])\) is a binary relation \(\leq\) on \(L([0, 1])\) such that, for all \(X, Y, Z \in L([0, 1]),\)

\(L(0, 1)\) \(\leq\) \(X\), (reflexivity), \(L(1) \leq X\), (antisymmetry), \(L(2) \leq Y\) and \(Y \leq X\) imply \(X = Y\), (transitivity), \(L(3) \leq Y\) and \(Y \leq Z\) imply \(X \leq Z\), (transitivity).

An order relation on \(L([0, 1])\) is called total or linear if any two elements of \(L([0, 1])\) are comparable, i.e., if for every \(X, Y \in L([0, 1]),\), \(X \leq Y\) or \(Y \leq X\).

We denote by \(\leq L\) any order in \(L([0, 1])\) (which can be partial or total) with \(0_L = [0, 0]\) as its minimal element and \(1_L = [1, 1]\) as its maximal element.
The $K_a$ operator is defined, for all $X \in L([0,1])$ and $a \in [0,1]$, by:

$$K_a(X) = (1-a)X + aX. \quad (2)$$

For $\alpha, \beta \in [0,1]$ with $\alpha \neq \beta$, the total order $\leq_{\alpha, \beta}$, induced by $K_\alpha$ and $K_\beta$, is defined, for all $X, Y \in L([0,1])$, as:

$$X \leq_{\alpha, \beta} Y \quad \text{if} \quad \left\{ \begin{array}{l}
K_\alpha(X) < K_\beta(Y) \\
K_\alpha(X) = K_\beta(Y) \text{ and } K_\beta(X) \leq K_\beta(Y)
\end{array} \right. \quad (3)$$

A total order on $L([0,1])$ is called an admissible order [48], [49] if it generalizes the standard product order [49], [50] between intervals, which is a partial order.

**Definition 1:** Consider $n \geq 2$. An $n$-dimensional interval-valued aggregation function in $L([0,1])$ with respect to $\leq_L$ is a mapping $\mathcal{A} : (L([0,1]))^n \rightarrow L([0,1])$ which verifies:

- (A1) $\mathcal{A}([0,0], \ldots, [0,0]) = 0_L$.
- (A2) $\mathcal{A}([1,1], \ldots, [1,1]) = 1_L$.
- (A3) $\mathcal{A}$ is a non-decreasing function in each variable with respect to $\leq_L$.

**D. Interval-valued Ordered Weighted Averaging (OWA) operators**

Let $\preceq$ be an admissible order on $L([0,1])$, $[a, b] \in L([0,1])$ and $w = (w_1, \ldots, w_n) \in [0,1]^n, w_1 + \cdots + w_n = 1$ a weighting vector. An interval-valued OWA operator associated with $w$ and $\preceq$ is a mapping $L([0,1])^n \rightarrow L([0,1])$ defined by [43]:

$$\text{OWA}([a_1, b_1], \ldots, [a_n, b_n]) = \sum_{i=1}^{n} w_i \cdot [a_{\sigma(i)}, b_{\sigma(i)}] \quad (4)$$

where $\sigma$ is a permutation such that, for all $i \in \{1, \ldots, n-1\}$, it holds that $[a_{\sigma(i)}, b_{\sigma(i)}] \preceq [a_{\sigma(i+1)}, b_{\sigma(i+1)}]$.

The weighting vector can be computed using a quantifier function $Q : [0,1] \rightarrow [0,1]$, defined here, for all $x \in [0,1]$ and $a, b \in [0,1]$, by:

$$Q_{a,b}(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } x > b \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

We then define, for $i \in \{1 \ldots n\}$:

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \quad (6)$$

Depending on the value of the parameters $a$ and $b$, different weighting vectors can be obtained. For this study, we have used the following configurations:

- OWA1: $a = 0.1, b = 0.5$
- OWA2: $a = 0.5, b = 1$
- OWA3: $a = 0.3, b = 0.8$

**E. Moderate deviation functions**

Regarding the notion of moderate deviation function, we follow the approach given in [51].

**Definition 2:** A function $\mathcal{D} : [0,1]^2 \rightarrow \mathbb{R}$ is called a moderate deviation function, if, for all $x, y \in [0,1]$, it satisfies:

- (MD1) $\mathcal{D}$ is non-decreasing in the first component;
- (MD2) $\mathcal{D}(x,y) = 0$ if and only if $x = y$.

The set of all moderate deviation functions is denoted by $\mathcal{MD}$.

The notion of moderate deviation function is closely related to those of restricted equivalence function [52] that we recall now.

**Definition 3:** A function $R : [0,1]^2 \rightarrow [0,1]$ is called a restricted equivalence function, if, for all $x, y, z \in [0,1]$, it satisfies:

- (R1) $R(x,y) = 0$ if and only if $\{x, y\} = \{0, 1\}$;
- (R2) $R(x,y) = 1$ if and only if $x = y$;
- (R3) $R(x,y) = R(y,x)$;
- (R4) If $x \leq y \leq z$, then $R(x,z) \leq R(x,y)$ and $R(x,z) \leq R(y,z)$.

***III. INTERVAL-VALUED MODERATE DEVIATION FUNCTIONS***

A moderate deviation function was introduced and corresponding deviation-based aggregation functions were studied in [51]. We make a similar study for intervals.

**Definition 4:** Let $\leq_L$ be a total order on $L([0,1])$ and

$$L(\mathbb{R}) = \{A = [A_\underline{A}, A_\overline{A}] | A_\underline{A}, A_\overline{A} \in \mathbb{R}, A_\underline{A} \leq A_\overline{A}\} \quad (7)$$

A function $D : (L([0,1]))^2 \rightarrow L(\mathbb{R})$ is called an interval-valued moderate deviation function w.r.t. $\leq_L$, if, for $X, Y \in L([0,1])$, it satisfies:

- (MD1) $D$ is non-decreasing in the second component w.r.t. $\leq_L$;
- (MD2) $D$ is non-increasing in the first component w.r.t. $\leq_L$;
- (MD3) $D(X, Y) = 0_L$ if and only if $X = Y$.

The set of all interval-valued moderate deviation functions w.r.t. $\leq_L$ is denoted by $\mathcal{IMD}$.

Interval-valued aggregation functions based on a given moderate deviation function can be defined by: $A(X_1, \ldots, X_n) = Y$ if and only if the equation $\sum_{i=1}^{n} D(X_i, Y) = [0,0]$ is satisfied for $X_1, \ldots, X_n \in L([0,1])$. However, from Definition [4] it is clear that the equation may not have a solution, or it may have more than one solution. Hence, we modify the procedure in a similar way as it was done in [51]. We adopt the convention $\sup\{a \in [b, c] | a \notin \Theta\} = b$ and $\inf\{a \in [b, c] | a \notin \Theta\} = c$.

**Definition 5:** Let $n \in \mathbb{N}, \leq_L$ be a total order on $L([0,1])$, $A : (L([0,1]))^2 \rightarrow L([0,1])$ be an idempotent interval-valued aggregation function w.r.t. $\leq_L$ and $D : (L([0,1]))^2 \rightarrow L(\mathbb{R})$ be an interval-valued moderate deviation function w.r.t. $\leq_L$.

Then the function $M_D : (L([0,1]))^n \rightarrow L([0,1])$ defined, for $X_1, \ldots, X_n \in L([0,1])$, by

$$M_D(X_1, \ldots, X_n) = A\left(\sup\{Y \in L([0,1]) | \sum_{i=1}^{n} D(X_i, Y) \leq_L [0,0]\}\right),$$

$$\inf\{Y \in L([0,1]) | \sum_{i=1}^{n} D(X_i, Y) >_L [0,0]\}\right) \quad (8)$$

is called an interval-valued $D$-mean w.r.t. $\leq_L$.

We are going to show that the proposed $D$-mean is a symmetric idempotent interval-valued aggregation function.

**Theorem 1:** Let $n \in \mathbb{N}, \leq_L$ be a total order on $L([0,1])$ and $D : (L([0,1]))^2 \rightarrow L(\mathbb{R})$ be an interval-valued moderate
deviation function w.r.t. $\leq_L$. Then the interval-valued $D$-mean $M_D : (L([0, 1]))^n \rightarrow L([0, 1])$ given in Definition 5 is an $n$-ary symmetric idempotent interval-valued aggregation function.

Proof: The symmetry is obvious. Regarding idempotency, let $X_1 = \ldots = X_n$. Then, since $D(X_1, Y) \leq_L [0, 0]$ if and only if $Y \leq_L X_1$, we have

$$\sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) \leq_L [0, 0] \right\} = X_1$$

(9)

and similarly

$$\inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) >_L [0, 0] \right\} = X_1,$$

(10)

hence $M_D(X_1, \ldots, X_1) = A(X_1, X_1) = X_1$.

It suffices to prove that $M_D$ is an interval-valued aggregation function. The boundary conditions follow from the idempotency, so it only remains to show the monotonicity. Let $X_1, \ldots, X_n \in L([0, 1])$ and let there exists $k \in \{1, \ldots, n\}$ such that $X_k \leq_L X_1$. Let $Z \geq_L X_k$. Then, since $D(Z, Y) \leq_L D(X_k, Y)$ for all $Y \in L([0, 1])$, we have

$$\sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) - D(X_k, Y) + D(Z, Y) \leq_L [0, 0] \right\} \geq_L \sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) \leq_L [0, 0] \right\}$$

(11)

and

$$\inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) - D(X_k, Y) + D(Z, Y) >_L [0, 0] \right\} \geq_L \inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D(X_i, Y) >_L [0, 0] \right\}.$$

(12)

Having in mind the monotonicity of $A$, the monotonicity of $M_D$ is proved.

Example 1: Let $D : (L([0, 1]))^2 \rightarrow L(\mathbb{R})$ be given, for $X, Y \in L([0, 1])$, by

$$D(X, Y) = \begin{cases} Z, & \text{if } Y >_L X; \\ -Z, & \text{if } Y <_L X; \\ [0, 0], & \text{if } Y = X, \end{cases}$$

for some $Z \in L([0, 1])$ such that $Z \geq_L [0, 0]$. Then, $D \in IMD$ and the corresponding interval-valued $D$-mean is median if $A$ is the median if $A$ is the arithmetic mean.

**Definition 6**: Let $n \in \mathbb{N}$, $\leq_L$ be a total order on $L([0, 1])$, $A : (L([0, 1]))^2 \rightarrow L([0, 1])$ be an idempotent interval-valued aggregation function w.r.t. $\leq_L$ and $D = (D_1, \ldots, D_n) \in IMD^n$. Let $w \in [0, \infty]^n$ be a non-zero weighting vector. Then

- The function $M_D : (L([0, 1]))^n \rightarrow L([0, 1])$ defined by
  $$M_D(X_1, \ldots, X_n) =$$
  $$= A \left( \sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D_i(X_i, Y) <_L [0, 0] \right\} ,$$
  $$\inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} D_i(X_i, Y) >_L [0, 0] \right\} \right).$$

(13)

is called an interval-valued $D$-mean w.r.t. $\leq_L$.

- The function $M_{D,w} : (L([0, 1]))^n \rightarrow L([0, 1])$ defined by
  $$M_{D,w}(X_1, \ldots, X_n) =$$
  $$= A \left( \sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} w_i D_i(X_i, Y) <_L [0, 0] \right\} ,$$
  $$\inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} w_i D_i(X_i, Y) >_L [0, 0] \right\} \right),$$

(14)

is called an interval-valued weighted $D$-mean w.r.t. $\leq_L$.

- The function $OM_{D,w} : (L([0, 1]))^n \rightarrow L([0, 1])$ defined by
  $$OM_{D,w}(X_1, \ldots, X_n) =$$
  $$= A \left( \sup \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} w_i D_i(X_i, Y) <_L [0, 0] \right\} ,$$
  $$\inf \left\{ Y \in L([0, 1]) \mid \sum_{i=1}^{n} w_i D_i(X_i, Y) >_L [0, 0] \right\} \right),$$

(15)

where $X_1 \geq_L \ldots \geq_L X_n$, is called an interval-valued ordered weighted $D$-mean w.r.t. $\leq_L$.

**Corollary 1**: Under the conditions of Definition 6, the interval-valued $D$-mean $M_D$, interval-valued weighted $D$-mean $M_{D,w}$ and interval-valued ordered weighted $D$-mean $OM_{D,w}$ are idempotent interval-valued aggregation functions. Moreover, $OM_{D,w}$ is symmetric.

**IV. w-PRESERVING INTERVAL-VALUED MODERATE DEVIATION FUNCTION**

In this section, a special class of interval-valued moderate deviation functions is studied, in particular, the functions satisfying the property: if all the input intervals have the same width, then the output interval has also the same width. We only consider $K_{a,b}$ orders in this section.

**Definition 7**: Let $\alpha, \beta \in [0, 1]$ with $\beta \neq \alpha$. A function $D : (L([0, 1]))^2 \rightarrow L(\mathbb{R})$ is called a $w$-preserving interval-valued moderate deviation function w.r.t. $\leq_{a,b}$ if, for $X, Y \in L([0, 1])$, it satisfies:

- $D$ is non-decreasing in the second component w.r.t. $\leq_{a,b}$;
- $D$ is non-increasing in the first component w.r.t. $\leq_{a,b}$;
- $K_{a}(D(X, Y)) = 0$ if and only if $K_{a}(X) = K_{a}(Y)$;
- $w(D(X, Y)) = w(X)$, then $w(D(X, Y)) = w(Y)$.

The set of all $w$-preserving interval-valued moderate deviation functions w.r.t. $\leq_{a,b}$ is denoted by $wIMD$.

A weaker form of monotone interval-valued functions will be of use later, so we define the so-called $w$-monotonicity.

**Definition 8**: An interval-valued function $A : (L([0, 1]))^n \rightarrow L([0, 1])$ is said to be $w$-monotone function w.r.t. order $\leq_L$, if it satisfies:
Proof: Unlike exists, let \( w \in X \) defined by \( L(D) = 0 \), ... , \( D \)-preserving, is an \( n \) \( \{ 0 \} \rightarrow [0, 1] \) that \( (MD4) \) immediately follows from the idempotency of \( B \) and the fact that \( A \) is \( w \)-preserving.

A construction method of \( w \)-preserving interval-valued moderate deviation functions is given in the following Theorem.

**Theorem 3:** Let \( \alpha, \beta \in [0, 1] \) with \( \beta \neq \alpha \). \( D : [0, 1]^2 \rightarrow \mathbb{R} \) be a strictly monotone moderate deviation function and \( C : [0, 1] \rightarrow [0, 1] \) be an idempotent function non-decreasing in the second component and non-increasing in the first component. Then the function \( D : (L([0, 1]))^2 \rightarrow L(\mathbb{R}) \) given by:

\[
\begin{align*}
K_\alpha(D(X,Y)) &= D(K_\alpha(X), K_\alpha(Y)), \\
w(D(X,Y)) &= C(w(X), w(Y))
\end{align*}
\]

is a \( w \)-preserving interval-valued moderate deviation function w.r.t. \( \leq_{\alpha, \beta} \).

**Proof:** Let \( Y \leq_{\alpha, \beta} Z \). There are two possibilities:

1) \( K_\alpha(Y) < K_\alpha(Z) \), then, since \( D \) is strictly monotone, we have \( K_\alpha(D(X,Y)) < K_\alpha(D(X,Z)) \), hence \( D(X,Y) \leq_{\alpha, \beta} D(X,Z) \), or
2) \( K_\alpha(Y) = K_\alpha(Z) \) and \( K_\beta(Y) \leq_{\beta, \beta} K_\beta(Z) \), in which case \( K_\beta(D(X,Y)) = K_\beta(D(X,Z)) \) and

\[
\begin{align*}
w(D(X,Y)) &= C(w(X), w(Y)) \leq C(w(X), w(Z)) = w(D(X,Z)), & \text{for } \beta > \alpha, \\
w(D(X,Y)) &= C(w(X), w(Y)) \geq C(w(X), w(Z)) = w(D(X,Z)), & \text{for } \beta < \alpha
\end{align*}
\]

so, in both cases \( K_\beta(D(X,Y)) \leq_{\beta, \beta} K_\beta(D(X,Z)) \) and finally \( D(X,Y) \leq_{\alpha, \beta} D(X,Z) \).

(MD2) can be proved similarly to (MD1). (MD3') follows from the fact that \( D \) is a moderate deviation function and (MD4) immediately follows from the idempotency of \( C \).

**Example 2:**

(i) Taking \( D : [0, 1]^2 \rightarrow L(\mathbb{R}) \) defined by \( D(x,y) = x - y \) and \( C : [0, 1]^2 \rightarrow [0, 1] \) defined by \( C(x,y) = \max(0, \text{min}(1, 2y - x)) \), by Theorem 3 one obtains a \( w \)-preserving interval-valued moderate deviation function w.r.t. \( \leq_{\alpha, \beta} \) for any \( \alpha, \beta \).

(ii) A class of \( w \)-preserving interval-valued moderate deviation functions w.r.t. \( \leq_{\alpha, \beta} \) for any \( \alpha, \beta \) can be obtained considering \( D_{\varepsilon, \delta} : [0, 1]^2 \rightarrow L(\mathbb{R}) \) defined for positive constants \( \varepsilon, \delta \) by (see Example 3.3 in [51]):

\[
D_{\varepsilon, \delta}(x,y) = \begin{cases} 
y - x + \varepsilon, & \text{if } y > x, \\
0, & \text{if } y = x, 
y - x - \delta, & \text{if } y < x
\end{cases}
\]

(20)
and \( C : [0,1]^2 \rightarrow [0,1] \) defined by \( C(x,y) = \max(0, \min(1, f(y) - f(x) + y)) \), where \( f : [0,1] \rightarrow \mathbb{R} \) is any non-decreasing function.

Note that item (i) is a special case of item (ii) for \( \epsilon = \delta = 0 \) and \( f = \mathbf{id} \).

**Example 3:** In [53] (Theorem 6) a construction of a moderate deviation function \( \mathcal{D} : [0,1]^2 \rightarrow [-M_n, M_p] \) was introduced in the following way:

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p - M_p R_1(x,y), & \text{if } x \leq y, \\
M_p R_2(x,y) - M_n, & \text{if } x > y,
\end{cases}
\]

for all \( x,y \in [0,1] \), where \( M_n, M_p \in [0,\infty] \). For different choices of restricted dissimilarity functions \( R_1, R_2 \) we obtain different moderate deviation functions. In particular, we give five examples, in each of them the choice of \( M_p, M_n \) impact results where the ratio between \( M_p \) and \( M_n \) is important since it expresses the emphasis we put on the positive \( (M_p) \) or negative \( (M_n) \) deviation:

(i) If \( R_1(x,y) = R_2(x,y) = 1 - |y - x| \), then

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p(y-x), & \text{if } x \leq y, \\
M_n(y-x), & \text{if } x > y.
\end{cases}
\]  

(ii) If \( R_1(x,y) = R_2(x,y) = 1 - |y^2 - x^2| \), then

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p(y^2 - x^2), & \text{if } x \leq y, \\
M_n(y^2 - x^2), & \text{if } x > y.
\end{cases}
\]  

(iii) If \( R_1(x,y) = R_2(x,y) = 1 - (y^2 - x^2) \), then

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p(y-x^2), & \text{if } x \leq y, \\
-M_n(y-x^2), & \text{if } x > y.
\end{cases}
\]  

(iv) If \( R_1(x,y) = 1 - |y^2 - x^2| \) and \( R_2(x,y) = 1 - (y - x)^2 \), then

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p(y^2 - x^2), & \text{if } x \leq y, \\
-M_n(y^2 - x^2), & \text{if } x > y.
\end{cases}
\]  

(v) If \( R_1(x,y) = 1 - (y - x)^2 \) and \( R_2(x,y) = 1 - |y^2 - x^2| \), then

\[
\mathcal{D}(x,y) = \begin{cases} 
M_p(y-x^2), & \text{if } x \leq y, \\
M_n(y-x^2), & \text{if } x > y.
\end{cases}
\]  

Based on the approach given in Theorem 3, we can build a \( w \)-preserving interval-valued moderate deviation function \( D : L([0,1])^2 \rightarrow L(\mathbb{R}) \) in such a way that we combine one of the five restricted dissimilarity functions \( \mathcal{D} \) from items (i)-(v) with a function \( C : [0,1]^2 \rightarrow [0,1] \) defined by

\[
C(x,y) = \max(0, \min(1, f(y) - f(x) + y))
\]

for some \( f : [0,1] \rightarrow \mathbb{R} \) being a non-decreasing function (for example \( Id, \ldots \)).

It is worth to point out that \( wD \)-mean is based on the idea to use moderate deviation functions in a similar way as penalty functions are used to measure the similarity or dissimilarity between a given set of data [54, 55]. The main idea is, given a set of intervals, to determine another interval which represents all of them and which is the most similar to all of them in the sense determined by the moderate deviation function. That is, we are looking for the interval \( Y \) which makes the sum \( D(X_1,Y) + \ldots + D(X_n,Y) \) to be as close to \([0,0]\) as possible.

In what follows, we use our construction given by Theorem 3 to avoid the computation of sup and inf while obtaining \( wD \)-mean.

**Theorem 4:** Let \( \alpha \in [0,1] \), \( n \in \mathbb{N} \), let \( M_p, M_n \) be positive real numbers and \( \mathcal{D} : [0,1]^2 \rightarrow [-M_n, M_p] \) be a moderate deviation function. Let \( F : L([0,1])^{n+1} \rightarrow \mathbb{R} \) be the function given, for all \( X_1, \ldots, X_n, Y \in L([0,1]) \) such that \( w(Y) = \min(w(X_1), \ldots, w(X_n)) \), by:

\[
F(X_1, \ldots, X_n, Y) = \mathcal{D}(K_\alpha(X_1), K_\alpha(Y)) + \ldots + \mathcal{D}(K_\alpha(X_n), K_\alpha(Y)).
\]

Then

(i) If \( \mathcal{D} \) is continuous, then, for each \( n \)-tuple \( (X_1, \ldots, X_n) \in L([0,1])^n \), there exists \( Y \in L([0,1]) \) such that \( w(Y) = \min(w(X_1), \ldots, w(X_n)) \) and \( F(X_1, \ldots, X_n, Y) = 0 \).

(ii) If \( \mathcal{D} \) is strictly increasing in the second component, then, for each \( n \)-tuple \( (X_1, \ldots, X_n) \in L([0,1])^n \), there exists at most one \( Y \in L([0,1]) \) such that \( w(Y) = \min(w(X_1), \ldots, w(X_n)) \) and \( F(X_1, \ldots, X_n, Y) = 0 \).

Proof: (i) Since the continuity of \( \mathcal{D} \) implies the continuity of the function \( F(X_1, \ldots, X_n, \cdot) \), the proof follows from the observation that \( F(X_1, \ldots, X_n, Z_1) \leq 0 \) and \( F(X_1, \ldots, X_n, Z_2) \geq 0 \) if \( K_\alpha(Z_1) = \min(K_\alpha(X_1), \ldots, K_\alpha(X_n)) \) and \( K_\alpha(Z_2) = \max(K_\alpha(X_1), \ldots, K_\alpha(X_n)) \). Note that since \( w(Y) \) is fully determined by fixed \( n \)-tuple \( (X_1, \ldots, X_n) \), the continuity of the function \( F(X_1, \ldots, X_n, \cdot) \) is considered in the sense of the standard continuity of a real function with the variable \( K_\alpha(Y) \).

(ii) Observe that the strict monotonicity of \( D \) implies the strict monotonicity (increasingness) of the function \( F(X_1, \ldots, X_n, \cdot) \). Again in the sense of a real function with the variable \( K_\alpha(Y) \).

The following corollary gives us a method of constructing \( wD \)-means based on Theorem 3, Theorem 4 and Example 3.

**Corollary 2:** Under the assumptions of Theorem 4 where \( \mathcal{D} : [0,1]^2 \rightarrow \mathbb{R} \) is given by Equation (21) with \( R_1, R_2 \) being continuously strictly monotone restricted equivalence functions and \( D : (L([0,1]))^2 \rightarrow L(\mathbb{R}) \) is given by Theorem 3 the following statements are equivalent:

(i) \( F(X_1, \ldots, X_n, Y) = 0 \);

(ii) \( M_D(X_1, \ldots, X_n) = Y \), where the interval-valued \( wD \)-mean \( M_D \) is given by Equation (16) with \( B = \min \);

(iii)

\[
\sum_{i=1}^{k} \left( M_p - M_p R_1 \left( K_\alpha(X_{\sigma(i)}), K_\alpha(Y) \right) \right) + \sum_{i=k+1}^{n} \left( M_n R_2 \left( K_\alpha(X_{\sigma(i)}), K_\alpha(Y) \right) - M_n \right) = 0
\]

where \( \sigma : \{1, \ldots, n\} \rightarrow \{1, \ldots, n\} \) is a permutation such that \( X_{\sigma(1)} \leq \alpha, \beta \leq \cdots \leq \alpha, \beta X_{\sigma(n)} \) and \( k \) is the greatest number from \( \{1, \ldots, n\} \) satisfying

\[
\sum_{i=1}^{n} \mathcal{D}(K_\alpha(X_{\sigma(i)}), K_\alpha(X_{\sigma(k)})) \leq 0.
\]  

Moreover, \( K_\alpha(Y) \in [K_\alpha(X_{\sigma(1)}), K_\alpha(X_{\sigma(k+1)})] \) whenever \( k < n \) and \( K_\alpha(Y) = K_\alpha(X_{\sigma(1)}) = \cdots = K_\alpha(X_{\sigma(n)}) \) whenever \( k = n \).
**Proof:** First observe that, due to (MD1) and (MD2), we have: if \( k \) satisfies Equation (30), then for all \( p \in \{1,\ldots,k\} \) and all \( q \in \{k+1,\ldots,n\} \) it holds
\[
\sum_{i=1}^{n} \varrho(K_{\alpha}(X_{\sigma(i)}), K_{\alpha}(X_{\sigma(p)})) \leq 0
\]
and
\[
\sum_{i=1}^{n} \varrho(K_{\alpha}(X_{\sigma(i)}), K_{\alpha}(X_{\sigma(q)})) > 0. \tag{31}
\]
Then the equivalence of (i) and (ii) follows from Theorem 4 and the equivalence of (i) and (iii) follows from the observation:
\[
F(X_{1}, \ldots, X_{n}, Y) = \sum_{i=1}^{k} \left( M_{p} - M_{p} R_{1} \left( K_{\alpha}(X_{\sigma(i)}), K_{\alpha}(Y) \right) \right) + \sum_{i=k+1}^{n} \left( M_{n} R_{2} \left( K_{\alpha}(X_{\sigma(i)}), K_{\alpha}(Y) \right) - M_{n} \right). \tag{32}
\]

**Example 4:** For each particular choice of \( \varrho \) (or \( R_{1}, R_{2} \)) in cases (i)-(v) of Example 3, the Equation (29) has the following form:

(i)
\[
\sum_{i=1}^{k} \left( M_{p} \left( K_{\alpha}(Y) - K_{\alpha}(X_{\sigma(i)}) \right) \right) + \sum_{i=k+1}^{n} \left( M_{n} \left( K_{\alpha}(Y) - K_{\alpha}(X_{\sigma(i)}) \right) \right) = 0
\]
and the solution is:
\[
K_{\alpha}(Y) = \frac{\sum_{i=1}^{k} M_{p} K_{\alpha}(X_{\sigma(i)}) + \sum_{i=k+1}^{n} M_{n} K_{\alpha}(X_{\sigma(i)})}{k M_{p} + (n-k)M_{n}}.
\]

(ii)
\[
\left( K_{\alpha}(Y) \right)^{2} \left( k M_{p} + (n-k)M_{n} \right) - M_{p} \sum_{i=1}^{k} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2} - M_{n} \sum_{i=k+1}^{n} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2} = 0
\]
and the solution is:
\[
K_{\alpha}(Y) = \sqrt{\frac{M_{p} \sum_{i=1}^{k} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2} + M_{n} \sum_{i=k+1}^{n} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2}}{k M_{p} + (n-k)M_{n}}}
\]

(iii)
\[
\left( K_{\alpha}(Y) \right)^{2} \left( k M_{p} - (n-k)M_{n} \right) + K_{\alpha}(Y) \left( 2 M_{n} \sum_{i=k+1}^{n} K_{\alpha}(X_{\sigma(i)}) - 2 M_{p} \sum_{i=1}^{k} K_{\alpha}(X_{\sigma(i)}) \right) + M_{p} \sum_{i=1}^{k} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2} - M_{n} \sum_{i=k+1}^{n} \left( K_{\alpha}(X_{\sigma(i)}) \right)^{2} = 0
\]

V. INTERVAL-VALUED AGGREGATION FUNCTIONS AND BCI FRAMEWORKS

In this section we present the two MI BCI frameworks we have used in our experiment: the Traditional Framework in Section V.A and the Multimodal Fuzzy Fusion framework in Section V.B. Then, we explain how we apply the Interval-valued moderate deviations in both cases, in Section V.C.

A. Traditional BCI framework

The traditional BCI system was proposed in [20]. Its structure includes four parts (Fig. 1):

1) The first step is acquiring the EEG data from a EEG device and performing band-pass filtering and artefact removal on the collected EEG signals.

2) The second step is EEG feature transformation and feature extraction. Usually, the FFT is used to rapidly transform the EEG signals into different frequency components [56]. The FFT analysis transforms the time-series EEG signals in each channel into the specified frequency range, which in our case is from 1 to 30 Hz, covering the delta (\( \Delta \)) 1-3 Hz, theta (\( \theta \)) 4-7 Hz, alpha (\( \alpha \)) 8-13 Hz, beta (\( \beta \)) 14-30 Hz and All 1-30Hz bands using a 50-point moving window segment overlapping 25 data points. Although some redundancy is included in the system, the All band is considered in order to study possible interactions among non-adjacent frequencies and to gather additional features for the subsequent CSP and classifiers.

3) Subsequently, the CSP is used for feature extraction to extract the maximum spatial separability from the different EEG signals corresponding to the control commands. The CSP is a well-known supervised mathematical procedure commonly used in EEG signal processing. The CSP is used to transform multivariate EEG signals into well-separated subcomponents with maximum spatial variation using the labels for each example [9], [57], [58].
4) Last, pattern classification is performed on the CSP
features signals using an ensemble of classifiers to dif-
ferentiate the commands. Each base classifier is trained
using a different wave band (for instance, if the base
classifier is the LDA, the ensemble would be com-
posed of: $\delta$–LDA, $\theta$–LDA, $\alpha$–LDA, $\beta$–LDA, and
$All$–LDA) and the final decision is taken combining
all of them. The most common way of obtaining the
final decision is to compute the arithmetic mean of the
outputs of all the base classifiers (each one provides
a probability for each class), and take the class with
highest aggregated value. Some of the most common
classifiers used for this task are the LDA, QDA and
KNN [59]. This part would correspond to the orange
box in Fig. 1 labeled as “Traditional”.

For this work, we have used LDA classifiers in the tradi-
tional framework, as they are very popular in the BCI literature
[5], [60], [61], [62], [63].

B. Multimodal Fuzzy Fusion BCI framework

The Multi-modal Fuzzy Framework (MFF) is proposed in
[19]. It follows a similar structure to the one in the traditional
BCI framework (Fig. 1): it starts with the EEG measurements,
it computes the FFT transformation to the frequency domain
and it uses the CSP transform to obtain a set of features to
train the classifiers.

However, in the MFF it is necessary to train not one, but
three classifiers for each wave band: a LDA, a QDA and
a KNN. We name the classifiers according to their type of
classifier and the wave band used to train it. For instance, for
the $\delta$ band we would have $\delta$–LDA, $\delta$–QDA and $\delta$–KNN.

Then, the decision making phase is performed in two phases
(This part would correspond to the green box in Fig. 1 labeled
as “Multimodal Fusion Framework”):

1) Frequency phase: since we got a LDA, QDA and KNN
for each wave band, the first step is to fuse the outputs
of these classifiers in each wave band. For example, in
the case of the LDA classifiers, we have a $\delta$–LDA,
$\theta$–LDA, $\alpha$–LDA, $\beta$–LDA and $All$–LDA that will
be fused them using an aggregation function to obtain a
vector, $FF$–LDA. That is, the same process explained
for the traditional framework is applied but without
making the final decision. We do the same with the QDA
and KNN classifiers. The result of this phase is a list of
collective vectors (one for each type of classifier).

2) Classifier phase: in this phase, the input is the list of col-
lective vectors given by each different kind of classifier
($FF$–LDA, $FF$–LDA, $FF$–KNN) computed in the
frequency phase. We fuse the three vectors according to
the classes, and the result is a vector containing the score
for each class for the given sample. As in the traditional
framework, the decision is made in favour to the class
associated with the largest value.

We must point out that the same aggregation is used for both
the frequency phase and the classifier phases.

The aggregation functions tested in the MFF are the Cho-
quet integral, the CF integral using the Hamacher T-norm,
the $CF_{min,min}$ generalizations, the Sugeno integral and the
Hamacher Sugeno integral [19], [26], [64].
C. Interval-valued Moderate Deviations applied in the BCI frameworks

We have tested the interval-valued moderate deviation functions in the two different MI BCI frameworks previously introduced. The idea in both cases is to replace the existing aggregations for the classifier outputs (the arithmetic mean in the traditional and the fuzzy integrals in the MFF) for our new developed ones.

First, we construct the intervals from the probability for each class obtained from each classifier. We use the length of the intervals to measure the inaccuracies or uncertainties related to these classifiers’ outputs. To do so, we have used the mapping in (33):

\[
F(x,y) = \frac{I(x,y) - I(x,y) + y}{2}
\]  

(33)

where \( F \) is an fuzzy implication function, \( x \) is the probability for each class obtained from the classifier and we set \( y \) as 0.3. We crop the values so that they are contained in the \([0,1]\) interval. We have tried three different fuzzy implication functions to construct the intervals: the Łukasiewicz fuzzy implication function, the Reichenbach fuzzy implication function and the Kleene-Dienes fuzzy implication function.

Then, we aggregate the interval-valued logits from the classifiers using an interval-valued moderate deviation based function. We have constructed the deviation function using the Eq. (21). We have named \( MD_1 \) to the interval-valued moderate deviation using \( R_1(x,y) = R_2(x,y) = 1 - |y - x| \) and the \( MD_2 \) setting \( R_1 = 1 - (y - x)^2 \) and \( R_2(x,y) = 1 - |y^2 - x^2| \).

In the traditional BCI framework, we follow this interval construction algorithm and the \( MD_1 \) and \( MD_2 \) respectively to substitute the “Average” block in Figure 1. In the case of the MFF we use them to substitute all the “Integral” blocks in Fig. 1.

To construct the interval-valued moderate deviation functions we need to set the parameters \( \theta_1 \) and \( \theta_2 \). We have optimized them by taking 200 samples of different pairs in the \([1, 100]\) range and testing the accuracy against the training set. We have opted for this method as it seemed to surpass other optimization algorithms in terms of computational time, obtaining similar accuracy results.

VI. EXPERIMENTAL RESULTS IN MOTOR-IMAGERY BASED BCI

In this section we discuss the behaviour of our new approaches in the BCI competition IV dataset 2a (IV-2a) and the BCI competition IV dataset 2b (IV-2b), which are detailed in [66]. The IV-2a dataset consists of four classes of MI tasks: tongue, foot, left-hand and right-hand performed by 9 volunteers. For each task, 22 EEG channels were collected. There is a total of 288 trials for each participant, equally distributed among the 4 classes. The IV-2b dataset consists of three different EEG channels for each subject, who performed two different motor imagery tasks: moving the right hand or moving the left hand.

For our experimental setup, we have used 4 out of the 22 channels available in the IV-2a dataset, (C3, C4, CP3, CP4), as they reported good results in [19]. For the IV-2b dataset, we have used the three channels available. As features, we have used the FFT to obtain the \( \delta \), \( \theta \), \( \alpha \), \( \beta \), SensoriMotor Rhythm (SMR) \((13-15Hz)\) and All, and we have used a CSP filter with 25 components. From each subject, we have generated twenty partitions (50% train and 50% test). So, this produces a total of 180 datasets. The accuracy for each framework is computed as the average for all of them.

For the IV-2a, we have studied the full four classes task, and we have also studied the binary classification task Left hand vs Right hand, as it is common practice in the literature [19], [67], [68]. Results for each individual subject are available at https://github.com/Fuminides/interval_md_bci_results.

A. Left/Right hand task

In Table I we have shown the results obtained using the \( MD_1 \) aggregation in the Left/Right hand task. We have used the three considered fuzzy implication functions and the two possible MI BCI frameworks. We found that the Łukasiewicz fuzzy implication function is the one that works best for both frameworks in the IV-2a dataset. For the IV-2b dataset, we found the Reichenbach operator is the best one in the traditional framework and the three operators gave very similar results for the MFF.

In Table II we have displayed the results using the \( MD_2 \) aggregation function. We found that in this case that in the IV-a dataset the best fuzzy implication function is the Reichenbach fuzzy implication function in the traditional framework and the Łukasiewicz one for the MFF. Results are higher or equal in all cases using this function compared to the \( MD_1 \). The best result is the 0.8251 obtained using the \( MD_2 \) traditional framework with the Reichenbach fuzzy implication function. In the IV-2b dataset, we found very similar performance for any of the implication operators and we also noted that the MFF performed better in all cases than the traditional framework.

B. Four classes task

In Table III we have shown the results obtained using the \( MD_1 \) aggregation in the left hand, right hand, tongue and foot task for the three considered fuzzy implication functions and
D. Comparison against interval-valued OWA operators

To compare the results obtained using the interval-valued moderate deviations with other interval-valued aggregations, we have compared them against the results obtained using interval-valued OWA operators. We have used three OWA operators: OWA1, OWA2 and OWA3, as described in Section III-D. We have computed the Kruskal statistical test and the Anderson post-hoc to look for significant differences among each method, using a level of significance of 0.05.

In Tab. IX we show the comparison among the interval-valued OWAs and the MD2 for the left/right hand task. We show for each aggregation the result for the best configuration (each configuration is composed of one the three different fuzzy implication functions and one of the two different frameworks) for each aggregation. We found that the MD2 is always performed better the the MD1 in this case too.

The best results for this task, 0.6943, were achieved using the MFF, the Łukasiewicz and the MD2.

C. Comparison to the non-interval-valued case

We have compared the best interval-valued MD aggregation, the MD2, with the standard arithmetic average aggregation in the traditional framework and the Choquet integral in the MFF framework, as it was the best aggregation for the MFF in [19].

In Table IV there are the results for the Traditional and MFF frameworks in the Left/Right hand task without using interval-valued aggregations, and the best result found in Sections VI-A and VI-B. In this case, we obtained a moderate increase in accuracy compared to the numerical results for the case of IV-2a dataset, but not in the case of the IV-b, in which the Choquet integral performed best. We also performed an analogous study for the four classes task, reported in Table VI. We found in this case that all the configurations performed almost equally, although the MD2 performed slightly better than the Choquet MFF.

In Tab. VII, we have computed the statistical differences for the left/right hand task among the numerical aggregations and the MD2. We have found significant differences favoring the MD2 aggregation for the case of the IV-2a dataset and the Choquet MFF for the IV-2b dataset using Kruskal test and the Anderson post-hoc to compute the respective P-values. We have performed the analogous study for the four classes task in Tab. VIII. We have found in that case that the difference favouring the MD2 is not significant compared to the Choquet integral in the MFF, but it is in the case of the arithmetic mean in the traditional framework.
the best aggregation for this task, followed by OWA2, OWA1, OWA1. The Kruskal test found statistical differences among the different aggregations, and the results for the Anderson post-hoc are in Tab. X, which shows that the MD2 performs statistically better than the rest of the aggregations.

In Tab. XI we display the comparison among the interval-valued OWAs and the MD2 for the four classes classification task. The results do not change much when it comes to decide which aggregation is the best, and the MD2, is again the best aggregation. The Kruskal test found statistical differences among the different aggregations, and the results for the post-hoc are in Tab. XII which shows that the MD2 performs statistically better than the rest of the aggregations, just as in the Left/Right hand task.

VII. COMPARISON WITH OTHER BCI FRAMEWORKS

In this Section we have tested the results obtained using the interval-valued MD2 with other non aggregation-based MI BCI frameworks. We have compared our proposal with OW A, Riemannian Multi Cov. [44] 0.7328 ± 0.1325 0.6763 ± 0.1102
CSP Multi Cov. [44] 0.7350 ± 0.1507 0.7258 ± 0.1181
Shallow CNN [46] 0.4862 ± 0.1225 0.6996 ± 0.1387
Deep CNN [46] 0.3956 ± 0.0717 0.7045 ± 0.1259
EEG Net [45] 0.5747 ± 0.1063 0.7229 ± 0.1281
Łakasiewicz MD2 MFF 0.6943 ± 0.0338 0.7310 ± 0.0470
Reinchenbach MD2 MFF 0.6905 ± 0.0328 0.7366 ± 0.0708

Table IX: Best accuracy for each different aggregation functions in the Left/Right hand task.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Dataset & Classifier & Accuracy \\
\hline
IV-2a & MD2 & 0.8251 ± 0.0415 \\
& OWA1 & 0.8044 ± 0.0476 \\
& OWA2 & 0.8086 ± 0.0465 \\
& OWA3 & 0.7993 ± 0.0483 \\
\hline
IV-2b & MD2 & 0.7366 ± 0.0708 \\
& OWA1 & 0.7300 ± 0.0800 \\
& OWA2 & 0.7312 ± 0.0285 \\
& OWA3 & 0.7296 ± 0.0786 \\
\hline
\end{tabular}
\caption{Table IX: Best accuracy for each different aggregation functions in the Left/Right hand task.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Dataset & Trad. & MD2 & OWA1 & OWA2 \\
\hline
IV-2a & MD2 & * & * & * \\
& OWA1 & .25 & .25 & \\
& OWA2 & .25 & \\
\hline
MFF & MD2 & OWA1 & OWA2 & OWA3 \\
\hline
MD2 & * & * & * \\
& OWA1 & .25 & .25 & \\
& OWA2 & .13 & \\
\hline
\end{tabular}
\caption{Table XII: P-values for four classes task comparison for the interval-valued aggregations. * marks statistical differences (P-value lesser than 0.05).}
\end{table}

Table X: P-values for the Left/Right hand task comparison for the interval-valued aggregations. * marks statistical differences (P-value lesser than 0.05).

In this work we have presented the interval-valued moderate deviations as a means to aggregate interval-valued data. We have extended the notion of moderate deviation function to the interval-valued setting, we have analyzed different properties and we have proposed different construction methods. We have, in particular, studied the case where the width of all the input interval-valued data is the same, and those interval-valued moderate deviation functions which preserve it. We have applied the interval-valued moderate deviation functions in the decision making phase of two MI BCI frameworks, using fuzzy implication functions to measure the effects of
noise in the EEG measurements in each classifier output. We have studied two different tasks: to discriminate between left hand and right hand classes and among left hand, right hand, tongue and foot classes. We found that the results using interval-valued moderate deviation functions outperform the rest of decision making schemes using other numerical and interval-valued aggregations, except for the case of the Choquet integral in the IV-2b dataset using the MFF.

Regarding non aggregation-based BCI frameworks, we found our proposal to beat CNN approaches, but we found our results not as good as the MI BCI framework that used CSP Multiscale Covariance. Since this method focuses on feature extraction, while ours is devoted to improve the decision making phase, we think that combining both approaches can be studied in order to further improve the current results.

In our future works we intend to develop moderate deviation based-aggregation functions for n-component vectors, and to further explore the combination of aggregation-based MI BCI frameworks with other MI BCI paradigms.

ACKNOWLEDGMENT

Javier Fumana Idocin’s, Jose Antonio Sanz’s, Javier Fernandez’s, Harkaitz Goyena’s and Humberto Bustince’s research has been supported by the project PID2019-108392GB I00 (AEI/10.13039/501100011033).

Z. Takač acknowledges the support of the grant VEGA 1/0545/20.

REFERENCES

[1] C.-T. Lin, C.-Y. Chiu, A. K. Singh, J.-T. King, L.-W. Ko, Y.-C. Lu, and Y.-K. Wang, “A wireless multifunctional sweep-based brain–computer interface assistive system,” IEEE Transactions on Cognitive and Developmental Systems, vol. 11, no. 3, pp. 375–383, 2018.

[2] C.-T. Lin, L.-W. Ko, J.-C. Chiou, J.-R. Duann, R.-S. Huang, S.-F. Liang, T.-W. Chiu, and T.-P. Jung, “Noninvasive neural prosthesis using mobile and wireless eeg,” Proceedings of the IEEE, vol. 96, no. 7, pp. 1167–1183, 2008.

[3] C.-T. Lin, C.-J. Chang, B.-S. Lin, S.-H. Hung, C.-F. Chao, and L.-J. Wang, “A real-time wireless brain–computer interface system for drowsiness detection,” IEEE transactions on biomedical circuits and systems, vol. 4, no. 4, pp. 214–222, 2010.

[4] C. Park, D. Looney, N. ur Rehman, A. Abraham, and D. P. Mandic, “Classification of motor imagery bci using multivariate empirical mode decomposition,” IEEE Transactions on neural systems and rehabilitation engineering, vol. 21, no. 1, pp. 10–22, 2012.

[5] S.-L. Wu, C.-W. Wu, N. R. Pal, C.-Y. Chen, S.-A. Chen, and C.-T. Lin, “Common spatial pattern and linear discriminant analysis for motor imagery classification,” in 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB). IEEE, 2013, pp. 146–151.

[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, “A review of classification algorithms for eeg-based brain–computer interfaces: A 10 year update,” Journal of neural engineering, vol. 15, no. 3, p. 031005, 2018.

[7] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, “Single-trial analysis and classification of erp components—a tutorial,” NeuroImage, vol. 56, no. 2, pp. 814–825, 2011.

[8] S. Vaid, P. Singh, and C. Kaur, “Eeg signal analysis for bci interface: A review,” in 2015 fifth international conference on advanced computing & communication technologies. IEEE, 2015, pp. 143–147.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, “Optimizing spatial filters for robust eeg single-trial analysis,” IEEE Signal processing magazine, vol. 25, no. 1, pp. 41–56, 2007.

[10] M. M. AlSaleh, M. Arvaneh, H. Christensen, and R. K. Moore, “Brain–computer interface technology for speech recognition: A review,” in 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). IEEE, 2016, pp. 1–5.

[11] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K.-R. Müller, “Toward unsupervised adaptation of lda for brain–computer interfaces,” IEEE Transactions on Biomedical Engineering, vol. 58, no. 3, pp. 587–597, 2010.

[12] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, “Spatial filtering for EEG-based regression problems in brain–computer interface (BCI),” IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp. 771–781, 2017.

[13] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, “Correlation-based channel selection and regularized feature optimization for MI-based BCI,” Neural Networks, vol. 118, pp. 262–270, 2019.

[14] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, “Internal feature selection method of csp based on l1-norm and dempster-shafer theory,” IEEE Transactions on Neural Networks and Learning Systems, 2020.

[15] Y. Li, F. Wang, Y. Chen, A. Cichocki, and T. Sejnowski, “The effects of audiovisual inputs on solving the cocktail party problem in the human brain: An fmri study,” Cerebral Cortex, vol. 28, no. 10, pp. 3623–3637, 2018.

[16] E. C. Cherry, “Some experiments on the recognition of speech, with one and with two ears,” The Journal of the acoustical society of America, vol. 25, no. 5, pp. 975–979, 1953.

[17] J. Jin, C. Liu, I. Daly, Y. Miao, S. Li, X. Wang, and A. Cichocki, “Bispectrum-based channel selection for motor imagery based brain-computer interfacing,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 10, pp. 2153–2163, 2020.

[18] N. Kotoyki and S. M. Hazarika, “Bispectrum analysis of eeg for motor imagery classification,” in 2014 International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2014, pp. 581–586.

[19] L.-W. Ko, Y.-C. Lu, H. Bustince, Y.-C. Chang, Y. Chang, J. Fernandez, Y.-K. Wang, J. A. Sanz, G. P. Dimuro, and C.-T. Lin, “Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface,” IEEE Computational Intelligence Magazine, vol. 14, no. 1, pp. 96–106, 2019.

[20] S.-L. Wu, Y.-T. Liu, T.-Y. Hsieh, Y.-Y. Lin, C.-Y. Chen, C.-H. Chuang, and C.-T. Lin, “Fuzzy integral with particle swarm optimization for a motor-imagery-based brain–computer interface,” IEEE Transactions on Fuzzy Systems, vol. 25, no. 1, pp. 21–28, 2016.

[21] J. Fumunal-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernandez, J. A. Sanz, and H. Bustince, “Motor-imagery-based brain-computer interface using signal derivation and aggregation functions,” IEEE Transactions on Cybernetics, 2021.

[22] G. Beliakov, H. Bustince, and T. C. Sánchez, A practical guide to averaging functions. Springer, 2016, vol. 329.

[23] D. Grabisch, M. Mesiar, and E. Pap. Aggregation functions. Cambridge University Press, 2009, vol. 127.

[24] R. Frank, F. Moser, and M. Ester, “A method for multi-relational classification using single and multi-feature aggregation functions,” in European Conference on Principles of Data Mining and Knowledge Discovery. Springer, 2007, pp. 430–437.

[25] M. Galan, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” Pattern Recognition, vol. 44, no. 8, pp. 1761–1776, 2011.

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

[27] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, M. J. Asiain, M. Elkano, and H. Bustince, “Ci-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.

[28] D. Paternain, J. Fernandez, H. Bustince, R. Mesiar, and G. Beliakov, “Construction of image reduction operators using minimizing penalty functions,” in 12th International Conference on Fuzzy Systems. IEEE, 2012, pp. 1–7.

[29] H. Bustince, E. Barrenechea, M. Pagola, J. Fernandez, Z. Xu, B. Bedregal, J. Montero, H. Hagras, F. Herrera, and B. De Baets, “A historical account of types of fuzzy sets and their relationships,” IEEE Transactions on Fuzzy Systems, vol. 24, no. 1, pp. 179–194, 2015.

[30] J. Sanz, A. Fernandez, H. Bustince, and F. Herrera, “A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position,”
Zdenko Takač received the Graduate degree in teaching mathematics and physics from the Faculty of Mathematics and Physics, Bratislava, Slovakia, in 1998, and the Ph.D. degree in teaching mathematics with the thesis analysis of mathematical proof from Pavol Jozef afrik University, Koice, Slovakia, in 2007. Since 1999, he has been a member of the Department of Mathematics, Faculty of Education, Catholic University in Ruomberok, Ruomberok, Slovakia, and since 2010, he has been a member of the Department of Mathematics, Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia. His research interests include uncertainty modeling, fuzzy sets and fuzzy logic, aggregation operators, interval-valued, and type-2 fuzzy sets.

Javier Fernandez received the M.Sc. and Ph.D. degrees in Mathematics from the University of Zaragoza, Saragossa, and the University of The Basque Country, Spain, in 1999 and 2003, respectively. He is currently an Associate Lecturer with the Department of Statistics, Computer Science and Mathematics Public University of Navarre, Pamplona, Spain. He is the author or coauthor of approximately 45 original articles and is involved with teaching artificial intelligence and computational mathematics for students of the computer sciences and computer engineering. His interests include fuzzy techniques for image processing, fuzzy sets theory, interval-valued fuzzy sets theory, aggregation functions, fuzzy measures, deep learning, stability, evolution equation, and unique continuation.

José Antonio Sanz received the M.Sc. and Ph.D. degrees in computer sciences in 2008 and 2011 respectively, both from the Public University of Navarra, Spain. He is currently an assistant professor at the Department of Statistics, Computer Science and Mathematics, Public University of Navarra, Pamplona, Spain. He is the author of 36 published original articles in international journals and more than 60 contributions to conferences. He is a member of the European Society for Fuzzy Logic and Technology (EUSFLAT) and the Spanish Association of Artificial Intelligence (AEPIA). He received the best paper award in the FLINS 2012 international conference and the Pepe Millá award in 2014. His research interests include fuzzy techniques for classification problems, interval-valued fuzzy sets, genetic fuzzy systems and medical applications using soft computing techniques.

Harkaitz Goyena received the Graduate degree in Computer Science, specializing in Computer Science and Intelligent Systems from the Public University of Navarra, Pamplona, Spain, in 2020. He is currently studying a master’s degree in mathematical investigation and modeling, statistics and computing.

Yu-Kai Wang received the B.S. degree in mathematics education from National Taichung University of Education, Taichung, Taiwan, in 2006, the M.S. degree in biomedical engineering from National Chiao Tung University (NCTU), Hsinchu Taiwan, in 2009, and the Ph.D. degree in computer science from NCTU, Hsinchu Taiwan, in 2015. He is currently a Lecturer of Faculty of Engineering and Information Technology, NCTU, International Faculty of University of California at San-Diego (UCSD), and Honorary Professorship of University of Nottingham. Dr. Lin was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated International Fuzzy Systems Association (IFSA) Fellow in 2012. Dr. Lin received the IEEE Fuzzy Systems Pioneer Awards in 2017. He served as the Editor-in-chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016. He also served on the Board of Governors at IEEE Circuits and Systems (CAS) Society in 2005-2008, IEEE Systems, Man, Cybernetics (SMC) Society in 2003-2005, IEEE Computational Intelligence Society in 2008-2010, and Chair of IEEE Taipei Section in 2009-2010. Dr. Lin was the Distinguished Lecturer of IEEE CAS Society from 2003 to 2005 and CIS Society from 2015-2017. He serves as the Chair of IEEE CIS Distinguished Lecturer Program Committee in 2018-2019. He served as the Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II in 2006-2008. Dr. Lin was the Program Chair of IEEE International Conference on Systems, Man, and Cybernetics in 2005 and General Chair of 2011 IEEE International Conference on Fuzzy Systems. Dr. Lin is the coauthor of Neural Fuzzy Systems (Prentice-Hall), and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific). He has published more than 330 journal papers (Total Citation: 22,418, H-index: 67, i10-index: 284) in the areas of neural networks, fuzzy systems, brain computer interface, multimedia information processing, and cognitive neuro-engineering, including about 120 IEEE journal papers.

Chin-Teng Lin received the B.S. degree from National Chiao-Tung University (NCTU), Taiwan in 1986, and the Master and Ph.D. degree in electrical engineering from Purdue University, USA in 1989 and 1992, respectively. He is currently the Distinguished Professor of Faculty of Engineering and Information Technology, and Co-Director of Center for Artificial Intelligence, University of Technology Sydney, Australia. He is also invited as Honorary Chair Professor of Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego (UCSD), and Honorary Professorship of University of Nottingham. Dr. Lin was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated International Fuzzy Systems Association (IFSA) Fellow in 2012. Dr. Lin received the IEEE Fuzzy Systems Pioneer Awards in 2017. He served as the Editor-in-chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016. He also served on the Board of Governors at IEEE Circuits and Systems (CAS) Society in 2005-2008, IEEE Systems, Man, Cybernetics (SMC) Society in 2003-2005, IEEE Computational Intelligence Society in 2008-2010, and Chair of IEEE Taipei Section in 2009-2010. Dr. Lin was the Distinguished Lecturer of IEEE CAS Society from 2003 to 2005 and CIS Society from 2015-2017. He serves as the Chair of IEEE CIS Distinguished Lecturer Program Committee in 2018-2019. He served as the Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II in 2006-2008. Dr. Lin was the Program Chair of IEEE International Conference on Systems, Man, and Cybernetics in 2005 and General Chair of 2011 IEEE International Conference on Fuzzy Systems. Dr. Lin is the coauthor of Neural Fuzzy Systems (Prentice-Hall), and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific). He has published more than 330 journal papers (Total Citation: 22,418, H-index: 67, i10-index: 284) in the areas of neural networks, fuzzy systems, brain computer interface, multimedia information processing, and cognitive neuro-engineering, including about 120 IEEE journal papers.
Humberto Bustince received the Graduate degree in physics from the University of Salamanca in 1983 and Ph.D. in mathematics from the Public University of Navarra, Pamplona, Spain, in 1994. He is a Full Professor of Computer Science and Artificial Intelligence in the Public University of Navarra, Pamplona, Spain where he is the main researcher of the Artificial Intelligence and Approximate Reasoning group, whose main research lines are both theoretical (aggregation functions, information and comparison measures, fuzzy sets, and extensions) and applied (image processing, classification, machine learning, data mining, and big data). He has led 11 I+D public-funded research projects, at a national and at a regional level. He is currently the main researcher of a project in the Spanish Science Program and of a scientific network about fuzzy logic and soft computing. He has been in charge of research projects collaborating with private companies. He has taken part in two international research projects. He has authored more than 210 works, according to Web of Science, in conferences and international journals, with around 110 of them in journals of the first quartile of JCR. Moreover, five of these works are also among the highly cited papers of the last ten years, according to Science Essential Indicators of Web of Science. Dr. Bustince is the Editor-in-Chief of the online magazine Mathware & Soft Computing of the European Society for Fuzzy Logic and technologies and of the Axioms journal. He is an Associated Editor of the IEEE Transactions on Fuzzy Systems Journal and a member of the editorial board of the Journals Fuzzy Sets and Systems, Information Fusion, International Journal of Computational Intelligence Systems and Journal of Intelligent & Fuzzy Systems. He is the coauthor of a monography about averaging functions and coeditor of several books. He has organized some renowned international conferences such as EUROFUSE 2009 and AGOP. Honorary Professor at the University of Nottingham, National Spanish Computer Science Award in 2019 and EUSFLAT Excellence Research Award in 2019.