Rule-Ranking-Based Approximate Knowledge Interpolation With Directional Monotonicity

Zhou, Mou; Shang, Changjing; Shen, Qiang

Published in:
IEEE Transactions on Cybernetics

DOI:
10.1109/TCYB.2023.3335472

Publication date:
2023

Citation for published version (APA):
Zhou, M., Shang, C., & Shen, Q. (2023). Rule-Ranking-Based Approximate Knowledge Interpolation With Directional Monotonicity. *IEEE Transactions on Cybernetics*, *54*(8). Advance online publication. https://doi.org/10.1109/TCYB.2023.3335472

Rule-Ranking-Based Approximate Knowledge Interpolation with Directional Monotonicity

Mou Zhou, *Student Member, IEEE,* Changjing Shang and Qiang Shen

***Abstract*—Fuzzy rule interpolation (FRI) empowers fuzzy rule-based systems (FRBS) with the ability to infer, even when presented with a sparse rule base where no direct rules are applicable to a given observation. The core principle lies in creating an intermediate fuzzy rule—either interpolated or extrapolated—derived from rules neighbouring the observation. Conventionally, the selection of these rules hinges upon distance metrics. While this approach is easy to grasp and has been instrumental in the evolution of various FRI methods, it's burdened by the necessity of extensive distance calculations. This becomes particularly cumbersome when swift responses are imperative or when dealing with large datasets. This paper introduces a groundbreaking rule-ranking-based FRI method, termed RT-FRI, which overcomes the constraints of the longstanding distance-centric FRI approach. Instead of relying on distances, RT-FRI harnesses ranking scores for rules and unmatched observation. These scores are produced by amalgamating the antecedent attributes using aggregation functions, thus streamlining the rule selection procedure. Recognising the rigid monotonicity demands of aggregation functions, a variant—DMRT-FRI—has been introduced to ensure directional monotonicity. Experimental results indicate that RT-FRI emerges as a highly efficient technique, with DMRT-FRI exemplifying a notable balance of accuracy and efficiency.**

***Index Terms*—Approximate inference, fuzzy rule interpolation, rule ranking, directional monotonicity.**

I. Introduction

RULE-BASED systems enriched with semantics offer insightful representations of human intelligence, as outlined by Zadeh [1]. The IF-THEN rule structure is extensively employed to model real-world complexities. Fuzzy rule-based systems (FRBS) stand out, providing a versatile framework for handling inferences in situations marked by ambiguity and imprecision. Their efficacy is well-demonstrated across diverse applications, from image processing [2] and geophysical to biological analyses [3], as well as in risk assessment [4]. Nevertheless, traditional FRBS face a limitation: they function optimally only when inputs align with the specified domain of their rule base. If the rule base doesn't cover the entire task domain comprehensively, certain observations may not activate any rules, resulting in the breakdown of the FRBS. Addressing this pivotal shortcoming is Fuzzy rule interpolation (FRI) [5], [6], which extends the capabilities of FRBS.

Over the past two decades, numerous techniques have been introduced to enhance FRI, particularly when dealing with sparse rule bases (e.g., [7]–[13]). Broadly, these techniques can be bifurcated into two categories: non-transformation-based and transformation-based FRI. Despite their categorial differences, both approaches typically commence their reasoning by selecting the nearest neighbouring rules to the given unmatched observation. This rule selection forms the bedrock for the ensuing interpolation. Historically, the rule selection process predominantly relies on distance metrics. For instance, the classical scale and move transformation-based FRI (T-FRI) [8] employs the Euclidean distance metric to gauge distances between unmatched observations and established rules. Weighted fuzzy rule interpolation (WT-FRI) [13] integrates feature selection [14] into distance computations by attributing unique weights to conditional variables during Euclidean distance calculations. Taking into account attribute correlations, a recently proposed method, the Mahalanobis distance-based FRI [15], computes the Euclidean distance within a specially transformed coordinate system. Here, distances between instances of the same class are minimised, whereas those between distinct classes are maximised. While these methods offer intuitive interpretations and have found success in numerous applications, their efficiency is often compromised by the extensive distance calculations they necessitate. Each new observation mandates a repeat of this intensive computational process, further dwindling efficiency— a significant drawback in scenarios demanding swift responses or big data contexts.

To address the longstanding limitations of the distance-based mechanism inherent in FRI, this paper introduces a rule-ranking-based FRI method, termed RT-FRI. Rather than computing distances between an unmatched observation and existing rules, this approach leverages a pre-established rule-ranking list to identify neighbouring rules. This list is generated by ranking all given rules based on scores derived from aggregating their pertinent antecedent variables using aggregation functions (AFs) [16]. Notably, the process of determining these ranking scores, and subsequently the rule-ranking list, is carried out once during the offline training phase. This ensures the list's availability prior to any online observation input. Consequently, when an unmatched observation is introduced, only its respective score is computed. This score's position within the rule-ranking list then facilitates the selection of the nearest rules to the observation, bypassing direct interactions with the full rule base. The subsequent reasoning aligns with conventional FRI procedures. By sidestepping extensive distance computations, this method enhances the efficiency of FRI fundamentally.

The proposed score-based FRI mechanism draws its reliability from the inherent monotonicity property of AFs.

M. Zhou, C. Shang and Q. Shen are with the Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, Aberystwyth SY23 3DB, U.K. (e-mail: {moz3, cns, qqs}@aber.ac.uk).

M. Zhou is also with the School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China.

Corresponding author: Changjing Shang.

This essential characteristic ensures a monotonic relationship between the inputs and outputs of an AF: closer antecedent attributes yield proximate aggregated scores. Consequently, for any unmatched observation, rules characterised by distant antecedent variables will be inherently excluded based on their aggregated scores.

While the monotonicity property of AFs is theoretically robust, its stringent constraints can pose challenges in practical applications. This is due to its inherent requirement: for one input's output to be larger or equal to another's, all antecedent values of the former input must not be smaller than those of the latter. If even a single attribute of the former input is not greater than its counterpart, the outputs cease to maintain their monotonic relationship. Consider the context of decision-making tasks. Here, multiple criteria are aggregated into a singular score using an AF. If every partial evaluation of a candidate is positive, an increase in the candidate's overall score is readily justifiable, as this candidate surpasses all criteria, not just a subset [17]. However, such ideal scenarios, where full monotonicity constraints are met, are infrequent in real-world applications. Therefore, a challenge arises when implementing this efficient score-based FRI method: how can one rationalise the selection of neighbouring rules? Especially when these rules, though they have the closest ranking scores to the observation, do not exhibit complete monotonicity across all antecedent attributes, rendering the entire monotonicity principle inapplicable.

The issue of monotonicity within FRBS was initially examined in [18], setting the stage for a series of advancements. For instance, to ensure monotone input-output dynamics in Mamdani systems, the properties of mathematical operators have been scrutinised [19]. Parametric conditions of monotonic FRBS in the development of TSK systems are detailed in [20], with a unique focus on a monotone fuzzy rule relabelling method for zero-order TSK systems [21]. More recently, a monotone FRI modelling scheme for TSK systems was introduced [22]. However, the common thread across these techniques has been their emphasis on constructing ideally monotonic models by delineating constraints tailored to various fuzzy model structures. Drawing inspiration from these studies and recognising the potential of directional monotonicity in fusion functions [23], this paper unveils a modified RT-FRI approach. Instead of striving for a perfect monotonic model, our approach offers greater flexibility by potentially relaxing the stringent monotonicity criterion. Both theoretical assessments and empirical evaluations validate that our proposed score-based FRI with directional monotonicity is both interpretable and efficacious.

The remainder of this paper is structured as follows. Section II introduces the foundational concepts of T-FRI, aggregation functions and directional monotone functions. Section III elucidates the rule-ranking-based FRI methodology and its subsequent extension leveraging directional monotonicity. Systematic experimental comparisons are presented in Section IV. Finally, Section V offers concluding remarks and hints at potential avenues for future enhancements.

II. Preliminaries

As mentioned earlier, this proposal employs T-FRI as the foundational FRI platform, replacing the distance metric within T-FRI with aggregation functions (AFs). To ensure a comprehensive understanding, this section delineates the primary steps involved in implementing T-FRI and offers an overview of the fundamental concepts and characteristics of AFs.

A. Scale and Move Transformation-based FRI (T-FRI)

Prior to going through the procedure of the T-FRI technique, basic concepts are introduced first.

1) Underlying Notions: To maintain generality, let a fuzzy rule base with multiple multi-antecedents be represented as a collection of production rules of the following format:

$$\begin{aligned}&\text{Rule } R_i:\\ &\text{If } x_1 \text{ is } X_{i1} \text{ and } x_2 \text{ is } X_{i2} \text{ and } \cdots \text{ and } x_m \text{ is } X_{im}, \quad (1)\\ &\qquad\qquad\qquad\qquad\qquad \text{then } y \text{ is } Y_i\end{aligned}$$

where R_i is the ith rule; $i = 1, 2, \ldots, N$ with N being the number of rules for this rule base; m is the number of antecedent attributes; X_{ij} and Y_i represent the jth ($j \in \{1, \ldots, m\}$) antecedent variable and the consequent part in R_i, each taking a fuzzy set as its value, respectively.

An observation (or input) is represented by

$$\begin{aligned}&\text{Observation } O^*:\\ &x_1 \text{ is } X_1^* \text{ and } x_2 \text{ is } X_2^* \text{ and } \cdots \text{ and } x_m \text{ is } X_m^*\end{aligned} \quad (2)$$

where X_j^* denotes the fuzzy set taken by the jth antecedent attribute.

The representative value (Rep) of a fuzzy set captures its core information, encompassing both its domain range's global position and its membership function's geometric shape. This makes it a prevalent tool in T-FRI. To formalise, consider an arbitrary polygonal fuzzy set $X = (a_1, a_2, \ldots, a_n)$, where each $a_t, t = 1, 2, \ldots, n$ represents the abscissas of vertices. The ordinates of these vertices signify the membership values. The representative value for this set, $\text{Rep}(X)$, is defined as:

$$\text{Rep}(X) = \sum_{t=1}^{n} w_t a_t \quad (3)$$

where w_t is the weight applied to a_t. This work encodes fuzzy sets using the popular triangle membership functions for computational efficiencies, such as $X_{ij} = (a_{ij}^1, a_{ij}^2, a_{ij}^3), Y^* = (c_1^*, c_2^*, c_3^*)$. Thus, the representative value of an arbitrary fuzzy triangular membership function X with the abscissas of the three vertices (a_1, a_2, a_3) defined as

$$\text{Rep}(X) = \frac{a_1 + a_2 + a_3}{3} \quad (4)$$

with the $w_t, t = 1, 2, 3$ all being set to $\frac{1}{3}$.

Generally, there are four core procedures to perform T-FRI interpolative inference, as summarised below.

2) Neighbouring Rule Selection: The cornerstone of implementing FRI is identifying the nearest neighbouring rules to an observation that remains unmatched by the rule base. This selection is predicated on the notion that these proximate rules bear similarity to the observation in question. Within the framework of T-FRI, the goal is to pinpoint n rules most akin to the unmatched observation, forming the foundation for the construction of an intermediate rule, pivotal for ensuing steps. The distance between a given rule R_i and observation O^* is determined by aggregating the distances across all corresponding attributes.

$$d(R_i, O^*) = \sqrt{\sum_{j=1}^{m} d(X_{ij}, X_j^*)^2} \tag{5}$$

where $d(X_{ij}, X_j^*)$ gives the distance between two fuzzy sets representing relevant antecedent attributes and is formulated as

$$d(X_{ij}, X_j^*) = \frac{|\text{Rep}(X_{ij}) - \text{Rep}(X_j^*)|}{range_j} \tag{6}$$

with $range_j = \max_{X_j} - \min_{X_j}, j = 1, 2, \ldots, m$ denoting the domain range value for the jth antecedent variable, and Rep$(\cdot)$ being defined in Eqn. (3) or (4).

3) Intermediate Rule Construction: From the above, the most similar n rules to the (unmatched) observation are selected for constructing the required single intermediate rule (*IR*):

$$\begin{aligned} &\text{If } x_1 \text{ is } X_1' \text{ and } x_2 \text{ is } X_2' \text{ and } \cdots \text{ and } x_m \text{ is } X_m', \\ &\qquad\qquad \text{then } y \text{ is } Y'. \end{aligned} \tag{7}$$

The foundational tenet of T-FRI's analogical reasoning postulates that if the antecedent value of an artificially generated intermediate rule, denoted X_j', shares a certain degree of similarity with the unmatched input X_j^* across all antecedent attributes $j = 1, 2, \ldots, m$, their respective consequent components, Y' and Y^*, should exhibit a commensurate similarity. This principle underpins not just this phase but extends to all subsequent FRI processes.

For the weighted degree to which the jth attribute of the ith rule informs the construction of the jth antecedent fuzzy set X_j' of the intermediate rule, denoted as *IR*, let it be represented by $w_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$. This weight is inversely proportional to the distance between X_{ij} and X_j^*.

$$w_{ij} = \frac{1}{d(X_{ij}, X_j^*) + 1} \tag{8}$$

where $d(X_{ij}, X_j^*)$ is defined in Eqn. (6). Then, the antecedent variable of *IR* is obtained from

$$X_j' = X_j'' + \delta_j range_j \tag{9}$$

with X_j'' denoting a temporary intermediate antecedent fuzzy set

$$X_j'' = \sum_{i=1,2,\ldots,n} \tilde{w}_{ij} X_{ij} \tag{10}$$

where $\tilde{w}_{ij}$ is the normalised weighted degree of Eqn. (8) and δ_j is the shift factor of X_j, defined respectively by

$$\tilde{w}_{ij} = \frac{w_{ij}}{\sum_{k=1,2,\ldots,n} w_{kj}}, \tag{11}$$

$$\delta_j = \frac{\text{Rep}(X_j^*) - \text{Rep}(X_j'')}{range_j}. \tag{12}$$

The intermediate consequent part Y' is then calculated in a similar way as above, that is

$$Y' = \sum_{i=1,2,\ldots,n} \tilde{w}_{iy} Y_i + \tilde{\delta}_y range_Y \tag{13}$$

where Y_i is the consequent fuzzy set of R_i; $range_Y = \max_Y - \min_Y$; and the two important factors $\tilde{w}_{iy}$ and $\tilde{\delta}_y$ are:

$$\tilde{w}_{iy} = \frac{\sum_{j=1}^{m} \tilde{w}_{ij}}{m}, \tag{14}$$

$$\tilde{\delta}_y = \frac{\sum_{j=1}^{m} \delta_j}{m}. \tag{15}$$

The rest of the interpolative reasoning is controlled by the critical scale and move rate factors, ensuring that the reasoning system attains the similarity degree between X_j' and X_j^*.

4) Scale Transformation: A scaled intermediate rule (*SIR*) is built to facilitate this transformation such that

$$\begin{aligned} &\text{If } x_1 \text{ is } X_1^\dagger \text{ and } x_2 \text{ is } X_2^\dagger \text{ and } \cdots \text{ and } x_m \text{ is } X_m^\dagger, \\ &\qquad\qquad \text{then } y \text{ is } Y^\dagger. \end{aligned} \tag{16}$$

In order to obtain *SIR* from the *IR*, the notion of scale rate s_{X_j} is defined. Given an arbitrary fuzzy term with triangular membership functions $X_j'(a_{j1}', a_{j2}', a_{j3}')$ of the intermediate rule, the scale rate is computed by

$$s_{X_j} = \frac{a_{j3}^* - a_{j1}^*}{a_{j3}' - a_{j1}'} \tag{17}$$

that essentially lengthens or shortens the support length of X_j': $a_{j3}' - a_{j1}'$ so that it matches that of X_j^*. Thus, the corresponding fuzzy set $X_j^\dagger(a_{j1}^\dagger, a_{j2}^\dagger, a_{j3}^\dagger)$ in *SIR* can be obtained by

$$\begin{bmatrix} a_{j1}^\dagger \\ a_{j2}^\dagger \\ a_{j3}^\dagger \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 + 2s_{X_j} & 1 - s_{X_j} & 1 - s_{X_j} \\ 1 - s_{X_j} & 1 + 2s_{X_j} & 1 - s_{X_j} \\ 1 - s_{X_j} & 1 - s_{X_j} & 1 + 2s_{X_j} \end{bmatrix} \begin{bmatrix} a_{j1}' \\ a_{j2}' \\ a_{j3}' \end{bmatrix}. \tag{18}$$

In doing so, the Rep for each antecedent attribute remains consistent throughout the scale transformation. Based on the analogical reasoning principle, intuitively, the consequent part of the *SIR* $Y^\dagger(c_1^\dagger, c_2^\dagger, c_3^\dagger)$ is

$$\begin{bmatrix} c_1^\dagger \\ c_2^\dagger \\ c_3^\dagger \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 + 2\tilde{s}_y & 1 - \tilde{s}_y & 1 - \tilde{s}_y \\ 1 - \tilde{s}_y & 1 + 2\tilde{s}_y & 1 - \tilde{s}_y \\ 1 - \tilde{s}_y & 1 - \tilde{s}_y & 1 + 2\tilde{s}_y \end{bmatrix} \begin{bmatrix} c_1' \\ c_2' \\ c_3' \end{bmatrix} \tag{19}$$

where $\tilde{s}_y$ is the average of s_{X_j}:

$$\tilde{s}_y = \frac{\sum_{j=1}^{m} s_{X_j}}{m}. \tag{20}$$

5) Move Transformation: Subsequent to the previous step, the goal here is to align $X_j^\dagger$ with the position of the jth attribute of the original observation X_j^*. Similarly, $Y^\dagger$ is adjusted to yield the inferred outcome Y^*. This alignment is achieved using the move ratio m_{X_j} applied to $X_j^\dagger$:

$$m_{X_j} = \begin{cases} \frac{3(a^*_{j1}-a^\dagger_{j1})}{a^\dagger_{j3}-a^\dagger_{j2}}, & \text{if } a^*_{j1}<a^\dagger_{j1} \\ \frac{3(a^*_{j1}-a^\dagger_{j1})}{a^\dagger_{j2}-a^\dagger_{j1}}, & \text{otherwise.} \end{cases} \tag{21}$$

Consequently, the resultant interpolative reasoning consequence $Y^*(c_1^*, c_2^*, c_3^*)$ is computed by

$$\begin{bmatrix} c_1^* \\ c_2^* \\ c_3^* \end{bmatrix} = \begin{cases} \frac{1}{3}\begin{bmatrix} 3-\tilde{m}_y & \tilde{m}_y & 0 \\ 2\tilde{m}_y & 3-2\tilde{m}_y & 0 \\ -\tilde{m}_y & \tilde{m}_y & 3 \end{bmatrix}\begin{bmatrix} c_1^\dagger \\ c_2^\dagger \\ c_3^\dagger \end{bmatrix}, & \text{if } \tilde{m}_b \geq 0 \\ \frac{1}{3}\begin{bmatrix} 3 & -\tilde{m}_y & \tilde{m}_y \\ 0 & 3+2\tilde{m}_y & -2\tilde{m}_y \\ 0 & -\tilde{m}_y & 3+\tilde{m}_y \end{bmatrix}\begin{bmatrix} c_1^\dagger \\ c_2^\dagger \\ c_3^\dagger \end{bmatrix}, & \text{otherwise} \end{cases} \tag{22}$$

where the move rate $\tilde{m}_y$ is

$$\tilde{m}_y = \frac{\sum_{j=1}^m m_{X_j}}{m}. \tag{23}$$

B. Aggregation Functions

An aggregation function (AF) mathematically provides a systematic tool to combine and merge multi-dimensional input into a single output [24].

Definition 1. *A function that maps n ($n > 1$) components from an n-dimensional unit cube onto a unit interval (denoted as $f : [0,1]^n \to [0,1]$), satisfying the following main constraints (bounds preservation and monotonicity), is termed an n-ary aggregation function [24]:*

(i) $f(0,\ldots,0) = 0$ *and* $f(1,\ldots,1) = 1$.
(ii) $\boldsymbol{x} \leq \boldsymbol{y}$ *implies* $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$.

One of the most fundamental and earliest aggregation functions (AFs) is the arithmetic mean. Over time, various alternative specifications have emerged to cater to different applications. Below, two widely adopted AFs are introduced, both of which can be effectively integrated into the proposed rule-ranking-based FRI framework.

1) Weighted Arithmetic Mean: Consider an aggregation function that allocates a weight w_i in the interval $[0,1]$ to each component of the input vector $\boldsymbol{x}$. Typically, there's a constraint such that $\sum_{i=1}^n w_i = 1$. These weight values can be interpreted as representing the relative significance or importance of the corresponding attributes or features. The function is defined as:

$$F_{\boldsymbol{w}}(\boldsymbol{x}) = w_1x_1 + w_2x_2 + \cdots + w_nx_n = \sum_{i=1}^n w_ix_i. \tag{24}$$

2) Choquet Integral: Choquet integral is defined with respect to fuzzy measure [25], which enables different degrees of signification to be assigned to not just individual features but also to subsets of them, thereby offering a sophisticated mathematical function to fuse variables [26]. A discrete Choquet integral defined on finite spaces is formulated by

$$C_\mu(\boldsymbol{x}) = \sum_{i=1}^n (x_{\sigma(i)} - x_{\sigma(i-1)})\mu(\mathcal{S}_{\sigma(i)}) \tag{25}$$

or

$$C_\mu(\boldsymbol{x}) = \sum_{i=1}^n x_{\sigma(i)}(\mu(\mathcal{S}_{\sigma(i)}) - \mu(\mathcal{S}_{\sigma(i+1)})) \tag{26}$$

where $(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$ is a non-decreasing permutation of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, $\mu(\cdot)$ is a fuzzy measure, and $\mathcal{S}_{\sigma(i)} = \{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\}$, with the convention that $x_{\sigma(0)} = 0$ and $\mathcal{S}_{\sigma(n+1)} = \emptyset$. The concept of a fuzzy measure, denoted by a set function $\mu : 2^{\mathcal{N}} \to [0,1]$ where $\mathcal{N} = \{1,\ldots,n\}$, serves as a foundational tool in constructing the Choquet integral [25]. While powerful, the computation of aggregation functions (AFs) utilising a fuzzy measure can be intricate, especially when dealing with 2^n elements — a challenge that amplifies as n increases. To mitigate this computational complexity, various methods have been proposed. Notable among these are the λ-fuzzy measure [27], the k-additive fuzzy measure [28], and the power measure [29]. This research primarily adopts the latter approach, which is defined as:

$$\mu(\mathcal{S}) = (\frac{|\mathcal{S}|}{n})^q, \ with \ q > 0 \tag{27}$$

where $\mathcal{S} \subseteq \mathcal{N}$ with $|\mathcal{S}|$ representing the cardinality of $\mathcal{S}$. It is observed that $\mu(\mathcal{S}_1) = \mu(\mathcal{S}_2)$ when subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ possess identical cardinality. Therefore, the Choquet integral informed by this specific power measure exhibits symmetry. Typically, with q designated as a consistent real value, the element count essential for computing the Choquet integral is $m-1$, presenting a significant reduction compared to 2^m. In this paper, q is set to 2 for simplicity (as conventionally done).

C. Directional Monotonicity

To accommodate functions exhibiting specific non-monotonic properties, weakly monotonic averaging functions were introduced as a means to temper the stringent monotonicity prerequisites associated with AFs [17]. Subsequently, this concept evolved into what is now recognised as directional monotone functions [23].

Definition 2. *A function $f : [0,1]^n \to [0,1]$ is $\vec{r}$-increasing ($\vec{r}$ is a real n-dimensional vector, $\vec{r} \neq \vec{0}$), if for $\boldsymbol{x} = (x_1,\ldots,x_n) \in [0,1]^n$ and for all $c > 0$ such that $\boldsymbol{x} + c\vec{r} \in [0,1]^n$, it holds that*

$$f(\boldsymbol{x} + c\vec{r}) \geq f(\boldsymbol{x}).$$

A function displaying $\vec{r}$-decreasing characteristics can be discerned by inverting the aforementioned inequality. This dual nature permits a fusion function, denoted as $f : [0,1]^n \to [0,1]$, to exhibit a specific direction of monotonicity, either increasing or decreasing, as dictated by vector $\vec{r}$. Therefore,

one can perceive fusion functions as expansive variants of AFs. More precisely, AFs can be classified as specialised fusion functions when appended with two specific criteria: bounds preservation and an inherent monotonic nature (refer to Definition 1, parts (i) and (ii)). In a formal context, AFs deviate from fusion functions in that they are $\vec{e_i}$-increasing for every $i \in \{1, \ldots, n\}$, while concurrently adhering to boundary limitations. Here, $\vec{e_i}$ is represented as $(\alpha_1, \ldots, \alpha_n)$, where $\alpha_i = 1$ and $\alpha_j = 0$ for every $j \neq i$.

III. Rule-ranking-based FRI and Extension

This section delineates the proposed rule-ranking-based FRI methodology and its extension, incorporating directional monotone functions.

A. Framework for Rule-ranking-based FRI

As alluded to earlier, if an observation fails to align with any rule within the established rule base, FRI's initial step is to identify the proximate rules prior to executing interpolative inference. A natural inquiry would be: how does one ascertain whether an observation corresponds to a particular rule without resorting to resource-intensive distance or similarity calculations? Bypassing the conventional approach, which typically relies on the Euclidean distance metric or similar alternatives, this method assesses the simple overlap between the support set of an observed attribute and its rule counterpart. If all antecedent attributes of a specific rule exhibit non-empty intersections, that rule is considered to align with the observation, irrespective of the partiality of this alignment. Under these circumstances, a standard inference mechanism, such as the compositional rule of inference [1], is applied, obviating the need for rule interpolation; in situations where this isn't the case, FRI steps in to facilitate the interpolation by selecting suitable rules.

Based on the above preliminaries, the proposed rule-ranking-based FRI approach is described as follows.

1) Proposed Generic Algorithm: Rather than relying on distance measures as defined in Eqn. (5), the proposed methodology employs AFs to amalgamate antecedent variables. This procedure yields a ranking score for each rule in the predefined rule base, as expressed by:

$$\text{Score}(R_i) = \text{AF}(\text{Rep}(X_{i1}), \ldots, \text{Rep}(X_{im})). \tag{28}$$

Here, $\text{AF}(\cdot)$ can be realised using any prominent aggregation function found in scholarly literature. The methodologies to consolidate elements can vary significantly across different AFs. For instance, the arithmetic mean is computed using summation and averaging. On the other hand, the weighted arithmetic mean, as illustrated in Eqn. (24), necessitates prior parameter learning through a dedicated training phase. Such parameters can be derived from contemporary feature selection methods. For classification tasks, methods like information gain [30] and relief-F [31] are pertinent; for predictive tasks, techniques such as local learning-based clustering [32] and the Laplacian score [33] prove to be effective.

To facilitate the identification of neighbouring rules, a list ranking these rules is formed by organising the ranking scores of every rule provided. Crucially, this procedure is conducted offline, well before the system's practical deployment. For the sorting process, the use of classical sorting algorithms, such as bubble sort, insertion sort, or cocktail sort, is suitable. This results in an ordered list of rules, arranged according to their respective scores. In the unusual event where multiple rules share an identical score, their relative positions in the list are designated randomly, ensuring the overall order of other distinct scores remains unchanged.

When encountering an unmatched observation during online operations, its score is derived using the same AF. This ensures the parameters remain consistent with prior uses, leading to:

$$\text{Score}(O^*) = \text{AF}(\text{Rep}(X_1^*), \ldots, \text{Rep}(X_m^*)). \tag{29}$$

It's worth noting that a new observation is typically represented as a crisp value. Similar to traditional fuzzy systems, there's a standard practice to fuzzify this crisp value. The method of fuzzification, which translates a crisp value into a fuzzy set representation, can vary based on the specific fuzzy membership function used. To ensure consistent fuzzification throughout the inference process and to streamline computational demands, this study solely employs the triangular fuzzy memberships, recognised for their simplicity.

Upon calculating the scores for the rules, rule selection is streamlined to a procedure that contrasts the observation's score with the ranked list, pinpointing the n rules whose scores are closest to that of the observation. This is achieved by computing and subsequently ordering the numerical disparities between the observation's score and those of individual rules. In instances where multiple rules share identical scores, the following approach is adopted: 1) If m rules, where $m < n$, have already been selected, a random subset of $(n - m)$ rules is chosen to complete the desired set of n rules; and 2) if the score difference of these rules is the smallest in comparison to that of the observation, and there are at least n such rules, a random subset of n rules is selected. However, if there are fewer than n such rules (let's say l, where $l < n$), all l rules are incorporated along with the next $n - l$ rules having the closest score differences.

Algorithm 1 provides a detailed procedure of the reasoning process, once both the rule-ranking list and the unmatched observation's score are available. The traditional T-FRI methodology comprises four stages, as discussed in Section II-A. In the present study, the initial stage is simplified as previously described. The function ConstInterRu(nR, O^*), spanning lines 2 to 7 in Algorithm 1, carries out the subsequent stage. To commence this phase, preliminary computations determine the requisite weighted degree $\tilde{w}_{ij}$ and the shift factor δ_j, as illustrated in Eqn. (11) and (12).

Having obtained $\tilde{w}_{ij}$ and δ_j, the intermediate antecedent variable X_j' is computed by

$$X_j' = f_1(\tilde{w}_{ij}, X_{ij}, \delta_j) = X_j'' + \delta_j range_j. \tag{30}$$

Following these calculations, the two significant factors $\tilde{w}_{iy}$ and $\tilde{\delta}_y$ that jointly reflect the contributions of all the conditional attributes in deriving the intermediate consequent B'

can be obtained, by applying the same AF as used previously (line 4). This leads to:

$$Y' = f_2(\tilde{w}_{iy}, Y_i, \tilde{\delta}_y) = \sum_{i=1,2,\ldots,n} \tilde{w}_{iy} Y_i + \tilde{\delta}_y range_Y \quad (31)$$

where $range_Y = \max_Y - \min_Y$.

Then, the functions ScaleTrans(IR, O^*) and MoveTrans(SIR, O^*), corresponding to lines 8 to 19 in Algorithm 1, adapt the third and fourth steps, respectively. Within these functions: 1) the parameters $\tilde{s}_y$ and $\tilde{m}_y$ are determined using a methodology akin to the previously detailed aggregation procedure; 2) the consequent scaled intermediate fuzzy set, denoted as $Y^\dagger(c_1^\dagger, c_2^\dagger, c_3^\dagger)$, is computed via the function $f_3(Y', \tilde{s}_y)$, which aligns with the expression given in Eqn. (19); 3) the final outcome of interpolative reasoning, represented as $Y^*(c_1^*, c_2^*, c_3^*)$, is derived from $f_4(Y^\dagger, \tilde{m}_y)$, consistent with Eqn. (22). Cumulatively, these refined steps embody the methodology introduced in this study.

2) Properties of Generic Algorithm: From the discussion above, this novel score-based FRI algorithm offers a more efficient means to perform rule interpolation than the original T-FRI. A pivotal enhancement in this methodology lies in substituting the intensive distance computation traditionally used in T-FRI's primary step with a simple score-based mechanism. Subsequent stages largely adhere to the original design.

This revamped methodology boasts several notable attributes:

1) Integrity of Inference: The fidelity of inference mirrors that of the original T-FRI, given that the foundational inference mechanism remains unaltered.

2) Autonomy of Information: The algorithm operates free from external data dependencies. Rule ranking scores emerge from an amalgamation of both antecedent and consequent information, which are intrinsic to the pre-existing rules.

3) Consistent Rule Ranking: For unmatched observations, the ranking remains consistent owing to the consistent use of the aggregation function (AF), which is also employed to determine the ranking scores of established rules.

4) Reliable Rule Selection: The choice of neighbouring rules benefits from enhanced reliability, a result of the inherent monotonicity constraints of the AFs.

B. Rule-ranking-based FRI with Directional Monotonicity

1) Problems in Rule-ranking Based FRI: As previously highlighted, the monotonicity constraint imposed on aggregation functions, as per Definition $1(ii)$, can sometimes pose undue restrictions in practical applications. It requires that only if all the elements being fused in $\boldsymbol{y}$ are at least not decreasing when compared with the corresponding ones in $\boldsymbol{x}$, like

$$\begin{array}{lccccccc} \boldsymbol{x}: & x_1, & x_2 & \ldots & x_p & \ldots & x_m \\ & |\wedge & |\wedge & & |\wedge & & |\wedge \\ \boldsymbol{y}: & y_1, & y_2 & \ldots & y_p & \ldots & y_m \end{array}$$

then the aggregated result of $\boldsymbol{y}$ can be not decreasing relative to $\boldsymbol{x}$, such that $\text{AF}(\boldsymbol{x}) \leq \text{AF}(\boldsymbol{y})$. In the context of decision-making, this might translate to an evaluator favouring a candidate who outperforms others across all evaluation criteria. However, such universally superior candidates are often elusive in real-world scenarios. It's more common to encounter candidates who excel only in specific criteria. As a result, when chosen rules and the input observation don't adhere to the monotonicity constraint, the rationale behind the proposed score-based rule selection method becomes challenging to justify.

Algorithm 1 Framework of Rule-ranking-based FRI.

Input:

- Rule-ranking list $RL = \{\{R_1, S_1\}, \ldots, \{R_N, S_N\}\}$ sorted with respect to ranking scores of rules in sparse rule base;
- Observation and coorresponding score $\{O^*, S_{O^*}\}$;

Output:

- Reasoning result Y^*;

Initialise number of fired rules to construct new interpolative rule n;

1: Select n rules $nR = \{R_1, \ldots, R_n\}$ from RL whose scores $S_i, i \in \{1, \ldots, n\}$ are nearest to S_{O^*};
2: **function** CONSTINTERRU(nR, O^*)
3: Obtain weighted degree $\tilde{w}_{ij}$ and shift factor δ_j, $i = 1, \ldots, n, j = 1, \ldots, m$, and compute new intermediate fuzzy antecedents X'_j by $f_1(\tilde{w}_{ij}, X_{ij}, \delta_j)$;
4: Compute $\tilde{w}_{iy}$ and $\tilde{\delta}_y$ for building intermediate consequence Y':

$$\tilde{w}_{iy} = \text{AF}(\tilde{w}_{i1}, \ldots, \tilde{w}_{im})$$

$$\tilde{\delta}_y = \text{AF}(\delta_1, \ldots, \delta_m);$$

5: Obtain intermediate consequence Y' through $f_2(\tilde{w}_{iy}, Y_i, \tilde{\delta}_y)$;
6: **return** intermediate rule $IR = (X'_1, \ldots, X'_m, Y')$
7: **end function**
8: **function** SCALETRANS(IR, O^*)
9: Obtain scale rate s_{X_j} and transform X'_j to $X_j^\dagger$, maintaining equal scale as with corresponding variable in O^*;
10: Compute $\tilde{s}_y$ to construct $Y^\dagger$

$$\tilde{s}_y = \text{AF}(s_{X_1}, \ldots, s_{X_m})$$

11: Obtain $Y^\dagger$ by $f_3(Y', \tilde{s}_y)$
12: **return** scale transformed rule $STR = (X_1^\dagger, \ldots, X_m^\dagger, Y^\dagger)$
13: **end function**
14: **function** MOVETRANS(STR, O^*)
15: Obtain move rate m_{Y_j} and transform $Y_j^\dagger$ to Y_j^*, maintaining equal position as with corresponding variable in O^*;
16: Compute $\tilde{m}_y$ to construct reasoning consequence Y^*

$$\tilde{m}_y = \text{AF}(m_{X_1}, \ldots, m_{X_m})$$

17: Obtain Y^* by $f_4(Y^\dagger, \tilde{m}_y)$
18: **return** interpolative reasoning result Y^*
19: **end function**
20: **return** Y^*;

Table I
SIMPLE RULE BASE AND OBSERVATION

	Rep(X_1)	Rep(X_2)	Rep(X_3)	Rep(Y)	Score
R_1	5.0	6.4	1.6	0	6.0
R_2	5.3	6.3	2.1	0	6.1
R_3	5.4	6.2	2.5	1	6.2
O^*	5.38	6.22	2.16	Y^*?	6.17

To better illustrate this problem, an artificial rule base and an observation, including their antecedent variables and ranking scores acquired by a certain AF, are shown in Table I. For the sake of clarity, the antecedent attributes of the rules and the observation are denoted using their representative values, as defined in Eqn. (10). This rule base is evidently non-monotonic. Let's consider a scenario where the goal is to select only the two rules most proximate in ranking score to the observation. Prior research has demonstrated that such a pair is usually sufficient for numerous FRI tasks [34]. Now, should rules R_2 and R_3 be chosen due to their scores most closely aligning with that of the observation O^*, an inconsistency emerges. Specifically, for the second antecedent attribute, X_2, the representative value of O^* falls between those of R_2 and R_3, but the overall sequence of ranking scores contradicts this order. This discrepancy is illustrated as follows:

$$\begin{array}{cccccc} & \text{Rep}(X_1) & & \text{Rep}(X_2) & \text{Rep}(X_3) & \text{Score} \\ R_2: & 5.3 & & 6.3 & 2.1 & 6.1 \\ & \wedge & & \vee & \wedge & \wedge \\ O^*: & 5.38 & & 6.22 & 2.16 & 6.17 \\ & \wedge & & \vee & \wedge & \wedge \\ R_3: & 5.4 & & 6.2 & 2.5 & 6.2 \end{array}$$

Considering another scenario, if R_1 and R_2 are selected for extrapolation (which is mathematically analogous to interpolation, save for the fact that the antecedent attributes of the neighbouring rules are located exclusively on one side of the observation), these rules would adhere to the monotonicity requirement. In such a case, the resultant reasoning, Y^*, would typically differ from the outcome achieved through interpolating with R_2 and R_3. Intuitively, one might argue that the interpolated result derived from R_2 and R_3 seems more plausible. This is because the antecedent values of R_3 are proximally closer to the observation than those of R_1.

2) Rank Mutual Information: The scenario described above is not uncommon in real-world applications. It is frequently observed that a decrease in the values of certain conditional attributes can result in an increase in the consequent. Given this, it becomes valuable to establish a method that discerns whether a specific rule base predominantly adheres to rank monotonicity. Rank mutual information (RMI) [35] offers a mathematical tool that can assess the underlying (aka., intrinsic) monotonicity of a given real dataset.

Definition 3. *Let $U = \{x_{1p}, \ldots, x_{np}\}$ be a set of instances for the pth variable within a dataset, an ascending RMI measured monotonic relationship between the attribute X_p and the aggregated consequence Y is defined as*

$$RMI^{\geq}(X_p, Y) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|[x_i]^{\geq}_{X_p}| \times |[y_i]^{\geq}_{Y}|}{n \times |[x_i]^{\geq}_{X_p} \cap [y_i]^{\geq}_{Y}|} \tag{32}$$

where $[x_i]^{\geq}_{X_p} = \{x_j \in U \mid x_i \geq_{X_p} x_j\}$, $[x_i]^{\geq}_{Y} = \{x_j \in U \mid x_i \geq_{Y} x_j\}$, and $x_i \geq_{X_p} x_j$ and $x_i \geq_{Y} x_j$ denote the circumstance that x_i is no worse than x_j in terms of X_p and Y, respectively.

Consider Table II as an illustrative example. It presents the RMI values for each input variable vis-à-vis the consequent, derived from three arbitrarily selected datasets [36]. Within the table, negative values signify an inverse relationship between the feature and the consequent (i.e., as the feature value rises, the consequent value drops). Conversely, positive values indicate a direct relationship, where both the feature and the consequent increase together. The magnitude of these values serves as an indicator of the strength of the relationship. Specifically, larger absolute values suggest that the corresponding variable leans heavily towards being either entirely inverse or wholly direct.

3) Improvements by Directional Monotone Functions: Directional monotone functions offer a means to relax the stringent monotonicity requirements associated with AFs. In this context, the rule-ranking-based FRI with directional monotonicity, termed DMRT-FRI, emerges. At its core, this method harnesses a directional monotone function. Under this framework, rules that boast ranking scores closely aligned with the score of an unmatched observation become prime candidates for selection. However, there's a caveat: they must conform to a condition. This stipulation demands that their directed paths—colloquially known as vectors—either towards or away from the observation, should be encompassed within the domain of $\mathcal{D}^{\uparrow}(f)$. Here, $\mathcal{D}^{\uparrow}(f)$ includes all real vectors $\vec{r} = (r_1, \ldots, r_n) \in V_n$, for which f shows a $\vec{r}$-increasing trait.

Elaborating further, consider two commonly employed AFs: the weighted arithmetic mean and the Choquet integral. In relation to the relaxed weighted arithmetic mean, the domain $\mathcal{D}^{\uparrow}(F_{\boldsymbol{w}})$ is delineated as

$$\mathcal{D}^{\uparrow}(F_{\boldsymbol{w}}) = \Big\{(r_1, \ldots, r_n) \in V_n \mid \sum_{i=1}^{n} w_i r_i \geq 0\Big\}. \tag{33}$$

It is straightforward to show that if $\sum_{i=1}^{n} w_i r_i \geq 0 (c > 0)$, then

$$\begin{aligned} F_{\boldsymbol{w}}(\boldsymbol{x} + c\vec{r}) &= \sum_{i=1}^{n} w_i (x_i + c r_i) \\ &= \sum_{i=1}^{n} w_i x_i + c \sum_{i=1}^{n} w_i r_i \\ &\geq F_{\boldsymbol{w}}(\boldsymbol{x}). \end{aligned}$$

According to the established theorem in the literature [23], if a function f, defined by a number of linear functions $f_1, \ldots, f_n$, is a continuous piecewise linear fusion function, then $\mathcal{D}^{\uparrow}(f) = \bigcap_{i=1}^{n} \mathcal{D}^{\uparrow}(f_i)$. Thus, for the use of Choquet integral, a special

Table II
RMI MEASURE FOR ALL INPUT FEATURES ON DATASETS

Dataset	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
Iris	0.4017	-0.1409	0.4938	0.4928									
Breast	-0.0176	0.0189	0.0561	0.0980	0.0714	0.1120	-0.0091	-0.0052	0.0545				
Heart	0.0697	0.0765	0.1663	0.0704	0.0623	-0.0028	0.0497	-0.1619	0.1251	0.1595	0.1084	0.1647	0.1718

piecewise linear fusion function depending on the permutation $\sigma(\cdot)$ (as per Eqn. (26)]) can be simplified to

$$C_\mu(\boldsymbol{x}) = \sum_{i=1}^{n} w(\mu, \sigma, i) x_{\sigma(i)} \tag{34}$$

where $w(\mu, \sigma, i) = \mu(\mathcal{S}_{\sigma(i)}) - \mu(\mathcal{S}_{\sigma(i+1)})$. Similar to Eqn. (33), the $\vec{r}$-increasing set of the relaxed Choquet integral consists of all permutations $\sigma(\cdot)$ satisfying $\sum_{i=1}^{n} w(\mu, \sigma, i) r_i \geq 0$, such that

$$\begin{aligned} \mathcal{D}^{\uparrow}(C_\mu) &= \bigcap_{\sigma \in \Omega} \mathcal{D}^{\uparrow}(F_{\sigma(\mu)}) \\ &= \bigcap_{\sigma \in \Omega} \left\{ (r_1, \ldots, r_n) \in V_n \mid \sum_{i=1}^{n} w(\mu, \sigma, i) r_i \geq 0 \right\} \end{aligned} \tag{35}$$

where Ω is the set of all permutations $\sigma(\cdot)$. A fortified variant of the rule-ranking-based FRI framework can be conceived by refining its initial step. Algorithm 2 delineates the specifics to supplant the primary step—precisely, line 1—of Algorithm 1. Notably, rules are systematically reordered according to the disparity between their ranking scores and the score of the unmatched observation (line 2). Thereafter, the rule most congruent in score to the observation's score undergoes an assessment to discern whether its vector $\vec{r}$ adheres to the directional monotonicity conditions as deduced from line 1. The vector $\vec{r}$ is discerned by its score: if the rule's score surpasses the observation's, $\vec{r}$ is perceived directionally from this rule to the observation and contrarily if lesser (lines 4-6). Conformance results in the rule's selection; otherwise, it's discarded. This evaluative sequence persists through the rearranged rule-ranking list Rea_RL until n rules have been selected (lines 7-10). The subsequent steps echo those in Algorithm 1 and hence are omitted for brevity.

It's pivotal to emphasise that this fortified approach singularly alters the initial phase of Algorithm 1. The adjustments hinge on contrasting the ranking scores of a rule subset against the unmatched observation's score. Consequently, the computational load remains largely invariant, showing no significant elevation compared to the original rule-ranking-based FRI paradigm. Typically, scrutiny is confined to a rule subset; the process culminates once the count of chosen rules attains the requisite n for interpolation. Empirically, this count frequently rests at a mere pair [34].

4) Computational Complexity: To underscore the efficacy of the proposed methodology and juxtapose it theoretically with conventional distance-based strategies, an analysis of time complexity is elucidated in this section.

Delving into the traditional T-FRI, one must scrutinise the time complexities inherent to its four cardinal steps, as detailed in Section II-A. Drawing upon the insights from extant literature [13], the complexities corresponding to each phase, contingent upon variables N and m (as expounded in Eqn. (1)), are as follows:

- *Neighbouring Rule Selection*: $N \times O(m) + O(N^2)$, where $N \times O(m)$ denotes the computation time of the ***for*** loop of calculating the distance between a rule R_i from N rules and the observation O^* with each attribute, and $O(N^2)$ represents the time for sorting distances;
- *Intermediate Rule Construction*: $3 \times O(mn) + O(n)$, where $3 \times O(mn)$ is the time for computing w_{ij}, $\tilde{w}_{ij}$ and X'_j (see Eqn. (8), Eqn. (11) and Eqn. (9)), $O(n)$ is for calculating Y' (Eqn. (13)), and $n = 2$ practically [34];
- *Scale Transformation*: $O(m) + O(1)$ denotes the sum of complexities required to compute s_{X_j} and $Y^{\dagger}$ (Eqn. (17) and Eqn. (19));
- *Move Transformation*: $O(m) + O(1)$, which is the same as the last step for calculating the similar factor m_{X_j} and Y^*.

Consequently, the time complexity of T-FRI is $T(\text{T-FRI}) = N \times O(m) + O(N^2) + 3 \times O(mn) + O(n) + 2 \times O(m) + 2 \times O(1) = O(N^2 + mN)$, which is the same to that of WT-FRI [13].

Both RT-FRI and DMRT-FRI adopt the final three steps from the foundational T-FRI. Therefore, the re-estimation of time complexity primarily concerns the initial step. As delineated in Alg. 1, this primary phase is centred around sorting the scores of the rules, a procedure condensed into line 1, with a complexity of $O(N^2)$. Therefore, the composite complexity for RT-FRI becomes $T(\text{RT-FRI}) = O(N^2) + 3 \times O(mn) + O(n) + 2 \times O(m) + 2 \times O(1) = O(N^2)$. For DMRT-FRI, aside from sorting scores, Alg. 2 introduces a *for* loop dedicated to selecting rules in compliance with directional monotonicity conditions. In the worst-case scenario, the complexity could elevate to $O(N^2) + O(N)$ when the terminal rule from the pool of N rules emerges as the nth selected rule. Consequently, the worst-case complexity for DMRT-FRI sums up to $T(\text{DMRT-FRI}) = O(N^2) + O(N) + 3 \times O(mn) + O(n) + 2 \times O(m) + 2 \times O(1) = O(N^2 + N)$. Given that $O(N^2) < O(N^2 + N) < O(N^2 + mN)$ for $m > 1$, it becomes evident that the proposed algorithms not only outpace the state-of-the-art WT-FRI but also surpass the efficiency of the original T-FRI.

It's imperative to understand that this analysis exclusively targets the algorithms and excludes the computational overhead associated with learning attribute weights. Such weights form a pivotal part of the inference process in WT-FRI, RT-

Algorithm 2 Selection of n Neighbouring Rules.

Input:

- Rule-ranking list $RL = \{\{R_1, S_1\}, \ldots, \{R_N, S_N\}\}$ with respect to sparse fuzzy rule base and rule ranking scores;
- Observation and rank score $\{O^*, S_{O^*}\}$;
- Number of fired rules to construct new interpolative rule n;

Output:

- n rules $nR = \{R_1, \ldots, R_n\}$ from RL;

Initialise $nR = \emptyset$;

1: Calculate conditions of $\vec{r}$ belonging to $\vec{r}$-increasing set;
2: Rearrange RL, such that $Rea\_RL = \{\{R_{\tau(1)}, S_{\tau(1)}\}, \ldots, \{R_{\tau(N)}, S_{\tau(N)}\}\}$, where $\tau(\cdot)$ is a permutation arranging $S_{\tau(i)}$ from nearest to farthest against observation;
3: **for** each $i \in [1, N]$ **do**
4: **if** $S_{O^*} \geq S_{\tau(i)}$ **then** $\vec{r} = \Big(\big(\text{Rep}(X_1^*) - \text{Rep}(X_{i1})\big), \ldots, \big(\text{Rep}(X_m^*) - \text{Rep}(X_{im})\big)\Big)$;
5: **else** $\vec{r} = \Big(\big(\text{Rep}(X_{i1}) - \text{Rep}(X_1^*)\big)\big), \ldots, \big(\text{Rep}(X_{im}) - \text{Rep}(X_m^*)\big)\Big)$;
6: **end if**
7: **if** $\vec{r}$ satisfies conditions in line 1 **then** add $R_{\tau(i)}$ to selected list $nR = \{\{R_{\tau(i)}, S_{\tau(i)}\}\}$;
8: **end if**
9: **if** number of nR: $num\_nR = n$ **then** break;
10: **end if**
11: **end for**
12: **return** nR;

Table III
SIMPLE RULE BASE

	Rep(X_1)	Rep(X_2)	Rep(Y)	Score
R_1	0.50	0.66	0	0.540
R_2	0.50	0.70	0	0.550
R_3	0.60	0.58	1	0.585
R_4	0.56	0.68	0	0.590
R_5	0.60	0.65	2	0.612
R_6	0.60	0.70	2	0.625

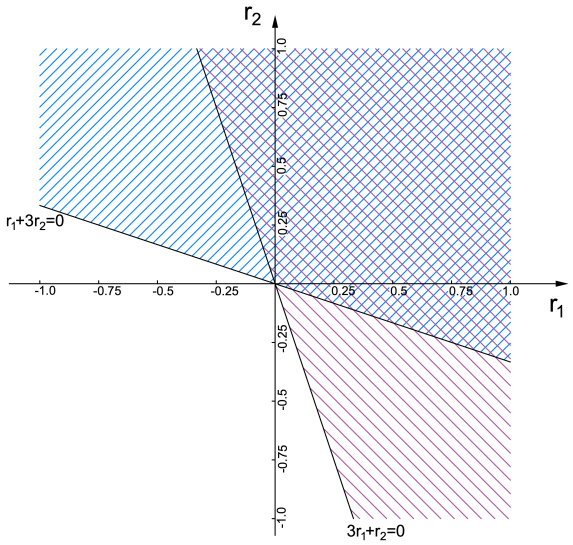

Figure 1. Graphical illustration of set $\mathcal{D}^{\uparrow}(C_\mu(\boldsymbol{x}))$.

FRI, and DMRT-FRI. A comprehensive computational complexity assessment of all the FRI methods under comparison will be undertaken experimentally in subsequent sections. Despite these considerations, the present analysis furnishes theoretical substantiation, underscoring the enhanced efficiency of the proposed rule-ranking-based methodology in contrast to prevailing distance-based techniques.

5) Illustrative Example: To illustrate the proposed approach, this paper employs an artificial fuzzy rule base encompassing two conditional attributes, as presented in Table III. It's essential to highlight that, for ease of representation, the fuzzy terms of the variables are expressed using representative values. Additionally, the rule base has been sequenced based on their ranking scores, which are derived by aggregating antecedent variables using the Choquet integral. The fuzzy measure is formulated using the power measure, as depicted in Eqn. (27). With the parameter q conventionally set to 2, the Choquet integral for this concise dataset can be described as:

$$C_\mu(\boldsymbol{x}) = \begin{cases} \frac{3}{4}x_1 + \frac{1}{4}x_2, & \text{if } x_1 \leq x_2 \\ \frac{1}{4}x_1 + \frac{3}{4}x_2, & \text{if } x_1 \geq x_2. \end{cases} \tag{36}$$

where x_1 and x_2 are the two antecedent variables. This is a typical piecewise linear fusion function. Therefore, the set $\mathcal{D}^{\uparrow}(C_\mu(\boldsymbol{x}))$ is determined by the intersection of $\frac{3}{4}r_1 + \frac{1}{4}r_2 \geq 0$ and $\frac{1}{4}r_1 + \frac{3}{4}r_2 \geq 0$, plotted as Fig. 1.

Consider an observation O^* with two antecedent representative values: 0.55 and 0.65. Its ranking score, determined in a manner consistent with rule-ranking score computation, stands at 0.575. This score is positioned between those of R_2 and R_3. With the conventional rule-ranking methodology, R_2 and R_3 would be identified as the closest rules to the observation for interpolation purposes. However, in the context of DMRT-FRI, the vectors $\overrightarrow{R_2O^*}$ and $\overrightarrow{O^*R_3}$ both fall outside the scope of $\mathcal{D}^{\uparrow}(C_\mu(\boldsymbol{x}))$. This indicates that neither R_2 nor R_3 fulfills the $\vec{r}$-increasing condition in relation to O^*. As a result, these rules are disregarded. Instead, the rules R_1 and R_4 are chosen. This is because their associated vectors, $\overrightarrow{R_1O^*}$ and $\overrightarrow{O^*R_4}$, align within $\mathcal{D}^{\uparrow}(C_\mu(\boldsymbol{x}))$. While the vector relationship between O^* and R_5, described as $\overrightarrow{O^*R_5}$, is $(0.05, 0)$-increasing, the score for R_5 diverges more significantly from that of O^*. This precludes it from being a primary choice. Moreover, R_6, being not in close proximity, results in $\overrightarrow{O^*R_6}$ not aligning with $\mathcal{D}^{\uparrow}(C_\mu(\boldsymbol{x}))$, thereby excluding it from the selection process. This process is visually represented in Fig. 2.

Drawing from the efficiency of RT-FRI, DMRT-FRI harnesses approximate inference in real-world scenarios. Instead

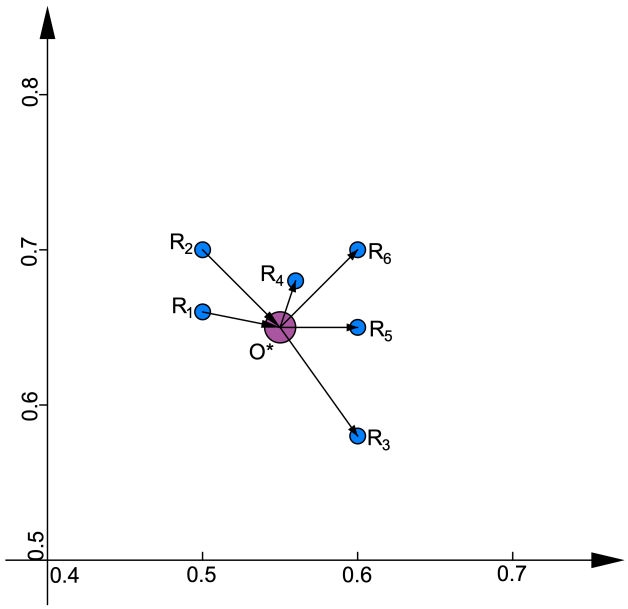

Figure 2. Graphical illustration of selecting rules based on directional monotonicity.

Table IV
DATASETS EMPLOYED

Datasets	#Instances	#Attributes	#Classes
Iris	150	4	3
Appendicitis	106	7	2
Pima	768	8	2
Breast	277	9	2
Wisconsin	683	9	2
Heart	270	13	2
Australian	690	14	2
WDBC	569	30	2
Ionosphere	351	33	2
Spambase	4597	57	2
Caesarian	80	5	2
Cryotherapy	90	6	2
Somerville happiness	143	6	2
Phishing	1353	9	3
Squash-unstored	52	24	3
Mammographic mass	830	5	2
Banknote authentication	1372	4	2
Electrial grid stability	10000	13	2
Skin nonskin	245057	3	2

of exhaustively examining every conceivable condition to build an impeccable monotonic rule base, DMRT-FRI manages to preserve monotonicity by loosening its stringent requirements. In doing so, it overcomes the innate theoretical constraints of the RT-FRI framework. Distinctively, DMRT-FRI diverges from existing monotonic FRBS techniques by simply introducing an additional condition check, retaining efficiency on par with RT-FRI. The upholding of monotonicity ensures the reliability of its reasoning logic.

It's pertinent to mention that the performance of the proposed method in addressing real-world challenges is closely tied to the chosen aggregation function. Although this introduces flexibility in crafting the FRI system, a reasonable selection of the aggregation function becomes paramount to realise effective implementations. Fortunately, this decision can be empirically informed through offline exploration. The ensuing experimental analysis offers valuable insights for such a decision, with specific recommendations on the optimal aggregation function provided in Section IV-C3.

IV. EXPERIMENTAL STUDIES

After the theoretical exposition, this section delves into comprehensive experimental analyses to gauge the efficacy of the proposed method, benchmarking it against T-FRI and WT-FRI—both exemplars of prevalent transformation-based FRI approaches as outlined in [6].

A. Datasets Employed for Experiments

For these evaluations, nineteen benchmark classification problems have been sourced from the KEEL dataset [36] and the UCI machine learning repositories [37]. These datasets span multi-class and multivariate challenges, representing an array of real-world contexts that encompass morphological, medical, sociological, and physical dimensions. To maintain an equitable assessment of systems architected using diverse methodologies, no specific assumptions related to data directional monotonicity have been made, and the datasets have been left unaltered without any specialised data cleaning.

B. Experimental Set-up

The experiments were conducted on a MacBook Pro featuring the M1 Chip, running MacOS Monterey, with the analyses executed in Pycharm Professional 2020. Comprehensive details about the datasets employed are delineated in Table IV. As previously mentioned, triangular membership functions serve to represent fuzzy values of all domain variables, unless indicated otherwise. The fuzzy rule base's generation leans on a practical methodology [2], grounded in the foundational Wang-Mendel technique [38]. Here, both the antecedent features and consequent spaces undergo an even segmentation into fuzzy regions. However, should the need arise, alternative rule induction strategies, such as those presented in [39], [40], remain viable options. It's worth noting that, in many instances, a pristine set of monotonic data does not inherently guarantee the generation of a monotonic fuzzy rule base, especially when deploying grid-based rule creation techniques like the Wang-Mendel method [41]. Yet, this research does not fixate on curating an impeccable monotonic fuzzy rule set to underpin interpolation. Instead, the focus shifts toward reinterpreting and tempering the constraints of monotonicity, with the overarching goal of facilitating the selection of proximate rules for interpolation-based reasoning.

To ensure an impartial performance evaluation, every antecedent attribute's domain undergoes normalisation, confining it to the 0 to 1 spectrum, and then uniformly parsed into five fuzzy sets. This strategy aligns with prevalent practices in

academic discourse, sidestepping any excessive emphasis on the inherent variability across datasets. It thereby lays down an equitable foundation for juxtaposing results. Strategically, twenty percent of the induced rules get pruned from the dense rule base, laying the groundwork for a more sparse rule base. This assists in appraising the FRI's performance. In practical scenarios where a dense rule base stands ready at one's disposal, the proclivity would be to leverage a traditional FRBS instead of an FRI. It's essential to underscore that the rules were not artificially trimmed, especially considering scenarios with limited training samples, as highlighted in [42].

For the purposes of evaluating FRI techniques within the context of sparse rule bases, the current experimental configuration entails the deliberate omission of 20% of the generated rules. This aligns with the established norms in academic literature, aiming to facilitate a systematic analysis. This approach probes the fidelity with which interpolated results mirror those inferred by a conventional FRBS [6]. While alternative methodologies for generating sparse rule sets exist—such as constraining an intricate fuzzy rule base through a rule-cutting threshold [43], harnessing curvature values to pinpoint pivotal rules coupled with a Genetic Algorithm (GA) for refinement [44], or devising granular fuzzy models optimised for estimating missing data using information granules [45]—this study maintains its focused scope. The implications of rule pruning on FRI performance have been examined, particularly in scenarios enriched with domain knowledge [46] and varied sparsity levels [47]. To maintain clarity in the comparative analyses, the ramifications of differing rule removal proportions are not explored, exclusively addressing cases where a fifth of the generated rules are omitted.

This empirical exploration contrasts reasoning results with the intrinsic ground truth, leveraging a 10×10-fold cross-validation method to ascertain average accuracy. The outcomes of both RT-FRI and DMRT-FRI are juxtaposed with their counterparts in the vanguard, T-FRI and WT-FRI. To ensure a balanced comparison, a critical parameter—specifically, the n delineated in Algorithm 1, which demarcates the number of rules enlisted for interpolation—is uniformly applied across the examined techniques. As per the insights from [34], a duo of proximate rules typically suffices for interpolation, striking an equilibrium between computational efficiency and precision. Consequently, n is assigned a value of 2 in subsequent trials. For WT-FRI, attribute weights originate from the widely recognised feature selection technique, Relief-F [48], although alternate methodologies remain available. Moreover, the implementation of both RT-FRI and DMRT-FRI invokes a symmetric Choquet integral.

C. Results and Discussion

Interpolative results are systematically investigated here, covering both reasoning effectiveness and efficiency.

1) Accuracy: The experimental outcomes are catalogued in Table V, where metrics such as average classification accuracies and their respective standard deviations (SD) are presented, gleaned from the 10×10 cross-validation procedure. The highest score for each dataset is emphasised in bold. Based on the presented data, several key insights emerge:

- Overall Performance: Predominantly, DMRT-FRI surpasses the other three benchmarked techniques, demonstrating superior performance in over 84% of the assessed datasets.
- Comparison Between Proposed Methods: Both DMRT-FR and RT-FRI, introduced in this study, exhibit analogous results on select datasets—namely, *Wisconsin*, *Cryotherapy*, and *Phishing*. This similarity in performance can be attributed to their shared foundational principle, which emphasises the selection of neighbouring rules corresponding to unmatched observations.
- Noteworthy Improvements: DMRT-FRI registers a marked improvement over T-FRI, particularly evident in datasets like *Wisconsin*, *Australian*, *WDBC*, *Squash-unstored*, and *Electrical grid stability*.
- Performance of WT-FRI: As a refined variant of T-FRI highlighted in prior research, WT-FRI clinches the top spot for three datasets out of the 19 assessed: *Iris*, *Ionosphere*, and *Phishing*.

While the outcomes on specific benchmark datasets, such as the 81% accuracy by DMRT-FRI on the *Iris* dataset, might seem suboptimal when compared to expectations set by established machine learning literature, it's vital to understand the context and objectives of these experiments. The primary focus here is not to achieve the highest absolute performance but to compare the relative performances of various techniques under similar constraints. To entail fair comparison, all methods investigated employ the common fuzzy quantity space specification [49] and the same original sparse rule base, with the latter being learned using a simple rule induction algorithm [38] involving no optimisation. Besides, many rules are deliberately eliminated from the densely induced rule base to obtain a sparse one [46], [47]. Hence, the achieved accuracy can deviate from what might be obtainable using optimised conventional machine learning methods. The key takeaway should be the relative advantage of the introduced approach in comparison to established T-FRI mechanisms.

2) Efficiency: The theoretical premise suggests that RT-FRI and DMRT-FRI can execute FRI interpolation without extensive distance calculations, potentially resulting in significant time savings. This study seeks to empirically validate this presumption by evaluating the time complexity of the involved algorithms. The time consumption statistics for each method across the 19 datasets are presented in Table VI. For algorithms like WT-FRI, RT-FRI, and DMRT-FRI, the time consumption computation encompasses the duration required to generate the attribute weight for every dataset. As per the presented results, RT-FRI consistently registers the least time consumption across all tasks, with its efficiency being particularly evident in larger datasets, such as *Spambase* and *Electrical grid stability*. While DMRT-FRI does not consistently match the efficiency of RT-FRI and occasionally demands slightly more computational resources than T-FRI on datasets like *Appendicitis*, *Pima*, *Wisconsin*, *Australian*, and *Electrical grid stability*, it outperforms WT-FRI in terms of time. This relative inefficiency of DMRT-FRI, when compared to RT-FRI, stems from its need to confirm if rule candidates align with the directional monotonicity conditions, which introduces an addi-

Table V
AVERAGE CLASSIFICATION ACCURACIES WITH SD OVER 10 × 10-FOLD CROSS VALIDATION

Datasets	T-FRI	WT-FRI	RT-FRI	DMRT-FRI
Iris	0.7854 ± 0.0371	**0.8241 ± 0.0263**	0.8055 ± 0.0289	0.8166 ± 0.0252
Appendicitis	0.7527 ± 0.0337	0.7964 ± 0.0263	0.8021 ± 0.0217	**0.8134 ± 0.0219**
Pima	0.6055 ± 0.0125	0.6179 ± 0.0159	0.6134 ± 0.0164	**0.6333 ± 0.0164**
Breast	0.6304 ± 0.0231	0.6183 ± 0.0360	0.6462 ± 0.0201	**0.6738 ± 0.0241**
Wisconsin	0.8116 ± 0.0077	0.8682 ± 0.0056	0.9560 ± 0.0073	**0.9577 ± 0.0033**
Heart	0.6500 ± 0.0193	0.6391 ± 0.0164	0.7322 ± 0.0167	**0.7415 ± 0.0230**
Australian	0.6070 ± 0.0098	0.6635 ± 0.0153	0.7334 ± 0.0124	**0.7654 ± 0.0173**
WDBC	0.7555 ± 0.0101	0.7628 ± 0.0108	0.7281 ± 0.0178	**0.8544 ± 0.0126**
Ionosphere	0.6497 ± 0.0138	**0.6910 ± 0.0106**	0.6775 ± 0.0200	0.6888 ± 0.0217
Spambase	0.5695 ± 0.0315	0.5575 ± 0.0123	0.5688 ± 0.0110	**0.5696 ± 0.0013**
Caesarian	0.4688 ± 0.0353	0.5197 ± 0.0566	0.5475 ± 0.0456	**0.6013 ± 0.0518**
Cryotherapy	0.5649 ±0.0292	0.6234 ± 0.0473	0.6455 ± 0.0199	**0.6479 ± 0.0135**
Somerville happiness	0.5333 ±0.0496	0.4940 ± 0.0382	0.5322 ± 0.0289	**0.5566 ± 0.0279**
Phishing	0.6279 ± 0.0086	**0.6980 ± 0.0124**	0.6444 ± 0.0172	0.6445 ± 0.0168
Squash-unstored	0.5490 ± 0.0669	0.5623 ± 0.0945	0.5748 ± 0.0456	**0.6455 ± 0.0587**
Mammographic mass	0.6378 ± 0.0095	0.6556 ± 0.0630	0.6953 ± 0.0300	**0.7080 ± 0.0317**
Banknote authentication	0.6839 ± 0.0183	0.6748 ± 0.0214	0.7304 ± 0.0097	**0.7554 ± 0.0168**
Electrial grid stability	0.5889 ± 0.0036	0.6523 ± 0.0013	0.7265 ± 0.0062	**0.7358 ± 0.0050**
Skin Nonskin	0.5828 ± 0.0179	0.7275 ± 0.0555	0.7298 ± 0.0138	**0.7488 ± 0.0144**
Average	0.6345 ± 0.0230	0.6656 ± 0.0298	0.6889 ± 0.0205	**0.7136 ± 0.0212**

Table VI
AVERAGE TIME COST (SECOND) OVER 10 × 10-FOLD CROSS VALIDATION

Datasets	T-FRI	WT-FRI	RT-FRI	DMRT-FRI
Iris	0.502	0.573	**0.366**	0.7054
Appendicitis	0.478	0.730	**0.387**	0.746
Pima	9.818	20.761	**3.234**	17.168
Breast	2.343	4.189	**1.195**	3.587
Wisconsin	7.108	14.971	**2.940**	10.734
Heart	3.081	6.885	**1.464**	3.921
Australian	14.827	34.996	**4.688**	17.985
WDBC	25.825	62.641	**6.899**	19.400
Ionosphere	11.021	24.408	**4.552**	8.676
Spambase	895.640	4108.691	**115.708**	337.172
Caesarian	0.278	0.381	**0.225**	0.441
Cryotherapy	0.359	0.579	**0.297**	0.605
Somerville happiness	0.631	1.110	**0.464**	1.174
Phishing	40.999	57.720	**5.948**	39.898
Squash-unstored	0.477	0.709	**0.417**	0.505
Mammographic mass	3.055	4.815	**2.224**	2.765
Banknote authentication	4.061	6.115	**3.229**	7.764
Electrial grid stability	2944.400	12340.943	**247.970**	4170.711
Skin Nonskin	3206.512	4446.676	**2103.544**	2469.131
Average	377.443	1112.521	**131.882**	374.373

tional computational step. Despite this, the empirical findings reinforce the theoretical prediction: the novel rule selection mechanism within DMRT-FRI offers a proficient method for conducting FRI, especially when emphasising consistent high accuracy.

3) DMRT-FRI with Weighted Arithmetic Mean: While DMRT-FRI excels in accuracy, it does not necessarily provide an unparalleled solution in terms of efficiency when benchmarked against other methods. Though the incorporation of the symmetric Choquet integral in DMRT-FRI simplifies the computational demands associated with conventional fuzzy measures, the constraints introduced by the set $\mathcal{D}^{\uparrow}(C_{\mu}(\boldsymbol{x}))$ can intensify its complexity, consequently affecting its efficiency.

A notable improvement in efficiency emerges when DMRT-FRI is adapted using the weighted arithmetic mean. Table VII delineates the performance metrics across all datasets under this configuration. For this implementation, weights for each antecedent variable are derived from RMI values, as illustrated in Eqn. (32). Within the table, any accuracy figure surpassing its counterpart in Table V is accentuated in bold. These results highlight that DMRT-FRI when implemented with the weighted arithmetic mean, not only amplifies accuracy across most of the datasets but also elevates its efficiency compared to the data in Table VI. Thus, this version of DMRT-FRI is endorsed for those seeking an optimal blend of accuracy and efficiency in real-world applications.

V. CONCLUSION

In the vast realm of Fuzzy Rule Interpolation (FRI), this research carves a distinctive niche by eschewing the well-trodden path of distance-based techniques, a mainstay for decades. The novel rule-ranking-based FRI (RT-FRI), introduced herein, represents a significant paradigm shift. Unlike conventional distance-based methods, it adeptly employs rule-ranking scores for interpolation when an observation cannot (fully or partially) match with any given rules. This innovative

Table VII
AVERAGE CLASSIFICATION ACCURACIES WITH SD AND TIME COST (SECOND) OF DMRT-FRI POWERED BY WEIGHTED ARITHMETIC MEAN OVER 10×10-FOLD CROSS VALIDATION

Datasets	Accuracy with SD	Time Cost
Iris	**0.8494 ± 0.0124**	0.4113
Appendicitis	**0.8285 ± 0.0205**	0.4081
Pima	0.6320 ± 0.0123	4.9269
Breast	**0.6952 ± 0.0140**	1.497
Wisconsin	**0.9582 ± 0.0033**	4.0812
Heart	**0.7544 ± 0.0157**	1.8671
Australian	**0.8027 ± 0.0067**	6.4089
WDBC	**0.9152 ± 0.0098**	8.5682
Ionosphere	**0.7510 ± 0.0155**	5.1788
Spambase	**0.6485 ± 0.0043**	125.4039
Caesarian	0.5776 ± 0.0321	0.2464
Cryotherapy	**0.7610 ± 0.0289**	0.3215
Somerville happiness	**0.5720 ± 0.0196**	0.5372
Phishing	**0.7605 ± 0.0081**	31.4765
Squash-unstored	0.5475 ± 0.0702	0.4217
Mammographic mass	0.6829 ± 0.0229	2.0707
Banknote authentication	**0.7594 ± 0.0074**	3.7476
Electrial grid stability	**0.9310 ± 0.0021**	263.1755
Skin Nonskin	**0.7950 ± 0.0006**	2113.69
Average	**0.7580 ± 0.0161**	135.50

approach, minimising computational demands by merely juxtaposing ranking orders, signifies a major stride in efficiency.

The groundbreaking Directional Monotonicity (DMRT-FRI) variant, another hallmark of this work, further augments the RT-FRI framework. Its inception addresses a realist challenge within RT-FRI: ensuring absolute monotonicity consistency across feature dimensions. By incorporating directional monotonicity conditions, it ushers in a refined, targeted rule selection process. Against the backdrop of conventional FRI methodologies, this research's empirical and theoretical assessments showccase RT-FRI's efficiency, notably outpacing both the revered T-FRI and the avant-garde WT-FRI. The astute fusion of the weighted arithmetic mean into DMRT-FRI elevates accuracy and computational efficiency to unmatched levels.

While this study is tested with Mamdani style fuzzy models [50], it has paved the way for broader horizons. There is an untapped potential in exploring the synergy between the current methodology and TSK-type fuzzy models [51] or the adaptive-network-based fuzzy inference system (ANFIS) [52]. Developing along this idea could amplify the benefits reaped from recent strides in applying approximate reasoning to such models [53], [54]. Another intriguing facet of this work is its reliance on foundational yet simplistic representative values to steer fuzzy interpolative reasoning. Trading with a more expansive inference mechanism that harnesses the full potency of fuzzy sets instead of just representative values beckons further exploration. Recent advancements in the theoretical foundations of aggregation functions [55], [56] have fortified the research framework, facilitating deeper scrutiny of the methodologies proposed in this study. Moreover, there is a compelling need for stringent statistical evaluations of the present work, especially concerning real-world applications, such as mammography risk modelling [47] or the nuanced domain of failure mode and effect analysis (FMEA) [57].

Within the landscape of classical symbolic learning, FRI stands distinctively, spotlighting transparent knowledge representation and championing interpretability, both for models and their use in inference. This commitment sharply diverges from the overarching trends of celebrated deep learning methodologies. A pivotal distinction lies in data demands: FRI thrives on minimalism, addressing scenarios where learned rules fall short of enveloping the full expanse of a problem. Essentially, FRI embraces the challenges of sparse knowledge, aligning with a data-driven learning framework that leverages limited historical data. Conversely, deep learning is voracious, feasting on vast datasets, while classical symbolic learning hinges on data sufficiency. Undeniably, deep learning's prowess in pattern discernment, especially amidst intricate and multilayered data, grants it an edge. However, FRI's adeptness at navigating uncertainty, its finesse with imprecision, and its ability to reason approximately even with a scant rule base accentuates its innate flexibility and latent potential.

REFERENCES

[1] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 1, pp. 28–44, 1973.

[2] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms: with applications to image processing and pattern recognition*. World Scientific, 1996, vol. 10.

[3] L. Duckstein *et al.*, *Fuzzy rule-based modeling with applications to geophysical, biological, and engineering systems*. CRC press, 1995, vol. 8.

[4] H. Chan and X. Wang, *Fuzzy hierarchical model for risk assessment: principles, concepts, and practical applications*. Springer London, 2015.

[5] L. T. Kóczy and K. Hirota, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Information Sciences*, vol. 71, no. 1, pp. 169–201, 1993.

[6] F. Li, C. Shang, Y. Li, J. Yang, and Q. Shen, "Approximate reasoning with fuzzy rule interpolation: background and recent advances," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4543–4590, 2021.

[7] P. Baranyi, L. T. Kóczy, and T. D. Gedeon, "A generalized concept for fuzzy rule interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 6, pp. 820–837, 2004.

[8] Z. Huang and Q. Shen, "Fuzzy interpolative reasoning via scale and move transformations," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 340–359, 2006.

[9] Y. C. Chang, S. M. Chen, and C. J. Liau, "Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1285–1301, 2008.

[10] L. Yang, F. Chao, and Q. Shen, "Generalized adaptive fuzzy rule interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 839–853, 2017.

[11] S.-M. Chen and W.-C. Hsin, "Weighted fuzzy interpolative reasoning based on the slopes of fuzzy sets and particle swarm optimization techniques," *IEEE transactions on cybernetics*, vol. 45, no. 7, pp. 1250–1261, 2014.

[12] C. Chen, N. Mac Parthaláin, Y. Li, C. Price, C. Quek, and Q. Shen, "Rough-fuzzy rule interpolation," *Information Sciences*, vol. 351, pp. 1–17, 2016.

[13] F. Li, Y. Li, C. Shang, and Q. Shen, "Fuzzy knowledge-based prediction through weighted rule interpolation," *IEEE Transactions on Cybernetics*, vol. 50, no. 10, pp. 4508–4517, 2020.

[14] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE transactions on cybernetics*, vol. 44, no. 8, pp. 1259–1268, 2013.

[15] M. Zhou, C. Shang, G. Li, L. Shen, N. Naik, S. Jin, J. Peng, and Q. Shen, "Transformation-based fuzzy rule interpolation with mahalanobis distance measures supported by choquet integral," *IEEE Transactions on Fuzzy Systems*, 2022.

[16] P. Liu and S.-M. Chen, "Group decision making based on heronian aggregation operators of intuitionistic fuzzy numbers," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2514–2530, 2017.

[17] T. Wilkin and G. Beliakov, "Weakly monotonic averaging functions," *International Journal of Intelligent Systems*, vol. 30, no. 2, pp. 144–169, 2015.

[18] P. Lindskog and L. Ljung, "Ensuring monotonic gain characteristics in estimated models by fuzzy model structures," *Automatica*, vol. 36, no. 2, pp. 311–317, 2000.

[19] E. Van Broekhoven and B. De Baets, "Only smooth rule bases can generate monotone mamdani–assilian models under center-of-gravity defuzzification," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 1157–1174, 2009.

[20] J.-M. Won and F. Karray, "Toward necessity of parametric conditions for monotonic fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 2, pp. 465–468, 2014.

[21] L. M. Pang, K. M. Tay, and C. P. Lim, "Monotone fuzzy rule relabeling for the zero-order tsk fuzzy inference system," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 6, pp. 1455–1463, 2016.

[22] Y. W. Kerk, K. M. Tay, and C. P. Lim, "Monotone fuzzy rule interpolation for practical modeling of the zero-order tsk fuzzy inference system," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 5, pp. 1248–1259, 2022.

[23] H. Bustince, J. Fernández, A. Kolesárová, and R. Mesiar, "Directional monotonicity of fusion functions," *European Journal of Operational Research*, vol. 244, no. 1, pp. 300–308, 2015.

[24] G. Beliakov, A. Pradera, T. Calvo *et al.*, *Aggregation functions: A guide for practitioners*. Springer, 2007, vol. 221.

[25] G. Beliakov, S. James, and J. Wu, *Discrete fuzzy measures*. Springer, 2020.

[26] P. Liu, S.-M. Chen, and G. Tang, "Multicriteria decision making with incomplete weights based on 2-d uncertain linguistic choquet integral operators," *IEEE Transactions on Cybernetics*, vol. 51, no. 4, pp. 1860–1874, 2021.

[27] M. Sugeno, "Theory of fuzzy integrals and its applications," *Doct. Thesis, Tokyo Institute of technology*, 1974.

[28] M. Grabisch, "K-order additive discrete fuzzy measures and their representation," *Fuzzy Sets and Systems*, vol. 92, no. 2, pp. 167–189, 1997, fuzzy Measures and Integrals.

[29] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, and H. Bustince, "Preaggregation functions: construction and an application," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 2, pp. 260–272, 2016.

[30] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin," *IEEE Transactions on Cybernetics*, vol. 45, no. 6, pp. 1209–1221, 2015.

[31] M. Hu, E. C. Tsang, Y. Guo, and W. Xu, "Fast and robust attribute reduction based on the separability in fuzzy decision systems," *IEEE transactions on cybernetics*, vol. 52, no. 6, pp. 5559–5572, 2021.

[32] H. Zeng and Y.-m. Cheung, "Feature selection and kernel learning for local learning-based clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1532–1547, 2011.

[33] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," *Advances in Neural Information Processing Systems*, vol. 18, 2005.

[34] F. Li, C. Shang, Y. Li, J. Yang, and Q. Shen, "Interpolation with just two nearest neighboring weighted fuzzy rules," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2255–2262, 2020.

[35] Q. Hu, X. Che, L. Zhang, D. Zhang, M. Guo, and D. Yu, "Rank entropy-based decision trees for monotonic classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2052–2064, 2012.

[36] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework." *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, 2011.

[37] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[38] L. X. Wang and J. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.

[39] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy sets and systems*, vol. 86, no. 3, pp. 251–270, 1997.

[40] T. Chen, C. Shang, P. Su, and Q. Shen, "Induction of accurate and interpretable fuzzy rules from preliminary crisp representation," *Knowledge-Based Systems*, vol. 146, pp. 152–166, 2018.

[41] C. Y. Teh, Y. W. Kerk, K. M. Tay, and C. P. Lim, "On modeling of data-driven monotone zero-order tsk fuzzy inference systems using a system identification framework," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3860–3874, 2018.

[42] K. W. Wong, D. Tikk, T. Gedeon, and L. Koczy, "Fuzzy rule interpolation for multidimensional input spaces with applications: a case study," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 809–819, 2005.

[43] S.-M. Chen, Y.-C. Chang, and J.-S. Pan, "Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 gaussian fuzzy sets and genetic algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 412–425, 2013.

[44] Y. Tan, J. Li, M. Wonders, F. Chao, H. P. H. Shum, and L. Yang, "Towards sparse rule base generation for fuzzy rule interpolation," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2016, pp. 110–117.

[45] X. Hu, Y. Shen, W. Pedrycz, Y. Li, and G. Wu, "Granular fuzzy rule-based modeling with incomplete data representation," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 6420–6433, 2021.

[46] T. Chen, C. Shang, J. Yang, F. Li, and Q. Shen, "A new approach for transformation-based fuzzy rule interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 12, pp. 3330–3344, 2020.

[47] F. Li, C. Shang, Y. Li, and Q. Shen, "Interpretable mammographic mass classification with fuzzy interpolative reasoning," *Knowledge-Based Systems*, vol. 191, p. 105279, 2020.

[48] I. Kononenko, "Estimating attributes: Analysis and extensions of relief," in *European Conference on Machine Learning*. Springer, 1994, pp. 171–182.

[49] Q. Shen and R. Leitch, "Fuzzy qualitative simulation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 4, pp. 1038–1061, 1993.

[50] E. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, 1975.

[51] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116–132, 1985.

[52] J.-S. R. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[53] J. Yang, C. Shang, Y. Li, F. Li, and Q. Shen, "Anfis construction with sparse data via group rule interpolation," *IEEE Transactions on Cybernetics*, vol. 51, no. 5, pp. 2773–2786, 2019.

[54] P. Zhang, C. Shang, and Q. Shen, "Fuzzy rule interpolation with k-neighbours for tsk models," *IEEE Transactions on Fuzzy Systems*, vol. 40, no. 10, pp. 4031–4043, 2022.

[55] J. Sartori, G. Lucca, T. Asmus, H. Santos, E. Borges, B. Bedregal, H. Bustince, and G. P. Dimuro, "d-cc integrals: Generalizing cc-integrals by restricted dissimilarity functions with applications to fuzzy-rule based systems," in *Brazilian Conference on Intelligent Systems*. Springer, 2023, pp. 243–258.

[56] H. X. Song, P. Yu, and H. Liu, "From pre-(quasi-) grouping functions to directional monotonic fuzzy implications," *Fuzzy Sets and Systems*, vol. 466, p. 108445, 2023.

[57] J. B. Bowles and C. E. Peláez, "Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis," *Reliability Engineering & System Safety*, vol. 50, no. 2, pp. 203–213, 1995.