A New Concordance Coefficients-Based Approach to Compare Improved FMECA Methods

Andrés Zúñiga [1], Paulo Branco [2], and João F. P. Fernandes [2]

[1]IDMEC
[2]Affiliation not available

October 30, 2023

Abstract

This work introduces the application of Cohen's kappa concordance coefficient as part of a comparative approach between different methods used to improve the FMECA analysis. The proposed approach considers the concordance assessment between different methodologies used in FMECA (Risk Isosurface function, VIKOR, ITWH, FWGM, Type-I and Type-II Fuzzy Inference System) when applied to the same problem and regarding an FMECA ranking selected as the reference one. The analyzed problem is a blood transfusion case study consisting of eleven failure modes widely used for benchmarking. Results show that Type-II fuzzy inference systems achieve the highest agreement regarding the reference FMECA ranking; one possible explanation for this result is that Type-II FIS considers uncertainty as an additional parameter. This approach proves effective to compare statistically different FMECA methods instead of the classical qualitative comparison between rankings.

A New Concordance Coefficients-Based Approach to Compare Improved FMECA Methods

Andrés A. Zúñiga, *Student Member, IEEE*, João F. P. Fernandes, *Member, IEEE* and Paulo J. Costa Branco, *Member, IEEE*

***Abstract*—The comparison between improved FMECA methods is commonly conducted by comparing qualitatively the resulting rankings and the potential balance between the three risk factors. This paper introduces the application of Cohen's kappa concordance coefficient as part of a comparative approach between different methods used to improve the FMECA analysis. The use of ranking agreement metrics allows comparing the rankings generated by independent raters; in this context, the application of Cohen's kappa in medical and social sciences is broad, but despite its relevance, its application in the FMECA context is limited. The proposed approach considers the concordance assessment between different methodologies used in FMECA (Risk Isosurface function, VIKOR, ITWH, FWGM, Type-I and Type-II Fuzzy Inference System) when applied to the same problem and regarding an FMECA ranking selected as the reference one. The analyzed problem is a blood transfusion case study consisting of eleven failure modes widely used for benchmarking. Results show that Type-II fuzzy inference systems achieve the highest agreement regarding the reference FMECA ranking; one possible explanation for this result is that Type-II FIS considers uncertainty as an additional parameter. This approach proves effective to compare statistically different FMECA methods instead of the classical qualitative comparison between rankings.**

***Index Terms*— FMECA; Risk assessment; Type-II fuzzy inference systems; Fuzzy weighted geometric mean; Concordance measurement; Cohen's kappa.**

I. Introduction

FAILURE Modes, Effects and Criticality Analysis is a qualitative risk assessment method designed to identify potential failure modes, their causes, and systems performance effects [1]. The objective of FMECA is to identify the possible ways a failure can occur, how often it occurs, how severe the failure affects the system performance, and what should be the preventive measures to avoid the failure.

The classical FMECA analysis is based on three factors, called risk factors, to characterize each failure mode [1]: the Severity (SEV) that characterize qualitatively the effect of the failure mode, the Frequency of Occurrence (OCC) that characterize how likely is it the failure mode to occur, and the Detectability (DET) that characterize how detectable is the failure mode before to occur. Each risk factor is classified in specific risk categories represented by a numerical scale, it can be a 1 to 10 scale as used in [1], or a 1 to 5 scale as in [2].

Each failure mode is assessed through a risk priority number (RPN); in general terms, the RPN results from the composition between SEV, OCC and DET as in (1), being the product the generally adopted approach.

$$RPN = (SEV) \circ (OCC) \circ (DET) \tag{1}$$

Because the RPN calculation in the classical FMECA approach results from the unique product between three integers, there is no associated computational complexity. Although FMECA is a very popular qualitative method for failure analysis, computation of the RPN has some disadvantages [3]–[5]. They are:

1) The RPN computation does not consider any difference degree between the three risk factors OCC, SEV, and DET (i.e., no weight averaging each risk factor).
2) Although a higher RPN is usually associated with a more critical failure modes, this is not always true [6], [7], and.
3) The scales for the three risk factors are generally considered arbitrarily and may not accurately represent the risk characteristics in specific problems.

To deal with these FMECA shortcomings, some approaches based on computational intelligence and decision-making methods were proposed in the past years. Bowles and Peláez [3] present one of the firsts applications of fuzzy inference systems FIS to improve the FMECA analysis; results shown that proposed FIS approach allows to overcome some FMECA issues like imprecise information related to the risk factors. In [6] authors conducted a literature review about FMECA methods published between 1998 and 2018; the review shows that publications about FMECA improvements increased in last

Manuscript received August 15, 2021. This work was supported by national funds through the Fundação para a Ciência e a Tecnologia, I.P., through IDMEC, under LAETA, Pproject UID/EMS/50022/2020, and by Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) of the Ecuadorian Government through fellowship CZ05-000291-2017. *(Corresponding author: Andrés A. Zúñiga.)*

Andrés A. Zúñiga, João F. P. Fernandes and Paulo J. Costa Branco are with the IDMEC, Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal (e-mail: andres.zuniga@tecnico.ulisboa.pt; joao.f.p.fernandes@tecnico.ulisboa.pt; pbranco@tecnico.ulisboa.pt).

ten years and methods like grey theory and fuzzy inference system are the most used to improve the FMECA analysis. In [7] it is shown the application of multi-criteria decision-making (MCDM) methods and uncertainty theory to model the vagueness related to FMECA processes. This book includes a broad review of academic works that apply MCDM methods to overcome FMECA issues.

In [8], authors shown the application of type-I fuzzy inference systems to improve FMECA prioritization in the context of Smart Grid architectures. The method was applied to eight selected failure modes, and the results obtained were very promising for the Smart Grid context. In [9] authors present a combination of fuzzy rules base and grey relation theory to improve the FMECA analysis conducted for an ocean going fish vessel; the proposed method includes the use of linguistic terms and allows to assign weights to each risk factor. Liu et al. in [4] proposed the application of interval 2-tuple hybrid weighted distance (ITHWD) in an FMECA analysis conducted in a blood transfusion problem. The FMECA analysis identified nineteen potential failure modes related to healthcare risks, and eleven failure modes with RPN higher than 80 were selected to apply the proposed FMECA approach; results proved to be a useful way to prioritizing the failure modes in the presence of uncertainty and incomplete information. In [10], uncertainty is considered using interval type-2 fuzzy sets to rank failure modes in a real oil spill incident. Based on five experts, their criteria were aggregated considering a rule-based approach, with the final fuzzy set subsequently defuzzified to find the RPN.

Reference [11] shows the application of type-2 fuzzy-based FMECA in the risk assessment of manufacturing facility in the automotive industry. The paper includes a fuzzy extension of the ordered weighted average (OWA) to assign an importance level to each of the fuzzy risk factors. Although the proposed method is limited to triangular membership functions; it is shown that the suggested approach offers additional flexibility to the experts in making judgments and provides better modeling of uncertainty in terms of intra and interpersonal uncertainty.

In [12] authors developed an approach that combines interval type-2 fuzzy sets and evidential reasoning applied to FMECA analysis of a steam valve system, considering eight failure modes. The methodology revealed to be more precise than conventional methods like fuzzy VIKOR and fuzzy TOPSIS, also reducing the probability of producing the same RPN. The weighting scheme applied to the three risk factors has made the result more comprehensive and can better express the uncertainty than type-1 fuzzy methods.

Anes et al. [13] show an FMECA approach based on two mathematical functions: the first one deals with cases where the order of importance of risk variables is sufficient to prioritize failure modes; the second functions are an extension of the first one and allow taking into account the relative weight of each variable. The proposed approach is applied to the blood transfusion problem analyzed in [4] and compared with other approaches but fuzzy-based ones. The results show that the proposed risk isosurface method has a good potential to prioritize failure modes according to their risk.

Commonly, the efficacy of these new FMECA approaches is evaluated qualitatively by comparing the rankings obtained by each method and analyzing a potential balance between the three risk factors. When the number of failure modes is small, it is possible to conduct a simple qualitative analysis to identify how the different approaches rank the failure modes by considering a balance between the three risk factors SEV, OCC, and DET. Otherwise, for a high number of failure modes, a qualitative analysis becomes unpractical.

As a result of FMECA analysis, the failure modes are ordered according to their RPN, from higher to lower, and an ordinal number is assigned for each ranked failure mode; consequently, the results of FMECA can be viewed as a single ordinal ranking. When different approaches are used to conduct an FMECA analysis, it is possible to determine the concordance or agreement between these methods.

The measure of concordance can be defined as the level of agreement between two or more raters or judges and it is also known as a *rank agreement*, *rank concordance*, *reproducibility*, or *interrater reliability*. This is a well-known problem in biological and social sciences.

The application of concordance measurements in the FMECA context is limited and focused to evaluate concordance between FMECA’s team human experts and not between FMECA methods. In [14] is presented an extension of the FMECA analysis using a Bayesian Belief Network as classifier for the FMECA parameters in the context of manufacturing process; the authors state that Kendall’s concordance coefficient is used to measure the concordance between the FMECA team experts but does not provide the results of the concordance measure. In his Ph.D Thesis, Okwesili includes the application of Kendall’s coefficient to determine the agreement between human experts in medical risk analysis context [15]. In [16] authors show the application of FMECA’s web-based three-round Delphi technique in the context of risk assessment related to transition from paper based records to digital based record in radiotherapy department; the Wilcoxon matched-pairs signed-ranks test and the Kendall’s coefficient of concordance were used to establish the consensus between the FMECA’s risk factor. In the three above mentioned papers, the concordance was measured between human experts conducting the classical FMECA’s.

This paper introduces the application of Cohen’s coefficient of concordance in the FMECA analysis. The main goal of this work is to provide a methodology to conduct a statistical comparison between different approaches used to improve the FMECA analysis.

The paper is organized as follows. Section II details the rank agreement problem; the selected metric for the rank agreement used in this paper is explained in Sections II.B to II.D. Section III introduces the use of the concordance coefficient in the FMECA context. Section IV shows the study case and Section V describes the developed fuzzy-based FMECA methods. Section VI shows the results of agreement between the reference ranking and different FMECA methods. Section VII shows a discussion about the obtained results, and Section VIII shows the paper's conclusion and future developments.

II. Preliminaries

A. The measure of rank agreement

Consider a collection of n objects classified by a particular characteristic, and let m a finite number of judges or evaluators who rank the n objects according to their appreciation of the objects' characteristics. It is important to know the degree of agreement between the evaluators' decisions. This kind of problem is usually named as *the problem of m ranking*. Kendall and Smith define it as: "If m persons rank n objects according to some quality of the objects, there arises the problem: does the set of m rankings of n show any evidence of a community of judgment among the m individuals?" [17]. The community of judgment is usually called an *agreement*.

Agreement, also known as *concordance*, *reproducibility* [18], or *interrater reliability* [19], is a concept closely related to, but fundamentally different from correlation [18]–[21]. The agreement focuses on the *degree of concordance* between two or more individuals or results between two or more assessments of interest variables [18]. The existence of agreement implies correlation, but the reciprocal may not be true [22]. Correlation statistics are usually applied to represent the association between two or more variables that do not necessarily measure the same attribute. In contrast, agreement statistically describes the measure of concordance in the opinion between individuals regarding the same attribute or characteristic. The concordance can be measured between a pair of raters or between several raters.

Reference [19] contains an exhaustive analysis about some coefficients of agreement currently used in social and biological sciences: Cohen's kappa, Scott's Pi, Krippendorf's Alpha, Gewt's AC_1, Aicking's α, Cronbach Alpha, Kendall's Tau, among others.

This paper considers the concordance measures between pairs of raters, then the Cohen's Kappa coefficient was selected to conduct the concordance analysis.

B. Cohen's coefficient of concordance

Cohen's coefficient, usually known as Cohen's kappa and denoted by κ, is a statistic useful for inter-rater or intra-rater reliability measures [23], [24]. Cohen's coefficient compares the proportion of objects in which the raters agreed and the proportion of objects for which disagreement is expected [23]. Originally, the coefficient κ was proposed as a measure of the agreement between two raters but it can be extended for more than two raters [24].

Follow, we resume how Cohen's coefficient of concordance is computed. Let N objects, $n = 1, 2, \cdots, N$, be classified independently into k categories by two separated and independent raters, observers or judges, called A and B, as shown in Table I. Here, as an example, Object 1 was rated as Category 5 by Rater A and Category 3 by Rater B. The categories can represent an intrinsic characteristic of the classified objects or a single ordinal ranking from 1 to k.

TABLE I
EXAMPLE OF N OBJECTS RANKED BY TWO RATERS

Objects	Rater A	Rater B
Object 1	Category 5	Category 3
Object 2	Category 2	Category k
$\vdots$	$\vdots$	$\vdots$
Object n	Category k	Category 5
$\vdots$	$\vdots$	$\vdots$
Object N	Category 1	Category 1

Let p_{ij} be the proportion of objects that rater A classified in the category i, $i = 1, 2, \cdots, k$, and rater B classified in the category j, $j = 1, 2, \cdots, k$, respectively. Table II shows the proportion of classified objects.

TABLE II
THE PROPORTION OF CLASSIFIED OBJECTS FOR EACH CATEGORY

		Rater B						
	Categories	1	2	…	j	…	k	Total
Rater A	1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1k}	p_{1+}
	2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2k}	p_{2+}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{ik}	p_{i+}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	k	p_{k1}	p_{k2}	$\cdots$	p_{kj}	$\cdots$	p_{kk}	p_{k+}
	Total	p_{+1}	p_{+2}	$\cdots$	p_{+j}	$\cdots$	p_{+k}	1

The proportions p_{i+} and p_{+j}, where the symbol + represents summation over the index, are the frequencies or marginal probabilities for an object to be assigned into category i for rater A and category j for rater B:

$$p_{i+} = \sum_{j=1}^{k} p_{ij} \tag{2}$$

$$p_{+j} = \sum_{i=1}^{k} p_{ij} \tag{3}$$

Where $\sum_{i=1}^{k} p_{i+} = 1$ and $\sum_{j=1}^{k} p_{+j} = 1$. Let p_0 be the observed proportion of agreement between raters [23] and expressed by (4):

$$p_0 = \sum_{i=1}^{k} p_{ii} \tag{4}$$

The observed proportion of agreement does not take into account the agreement obtained only by chance (this means not really "agreeing" at all) [25]. Therefore, the expected proportion of agreement obtained by chance, denoted by p_e, is based on the probability that rater A assigns the objects in the category i overall and rater B assigns the objects in the same category overall, that is for all $i = j$:

$$p_e = \sum_{i=1}^{k} \left(p_{i+} \cdot p_{+i} \right) \tag{5}$$

Then, Cohen's κ coefficient can be defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{6}$$

The lower and upper limits for κ are -1 and 1, respectively, but usually falls between 0 and 1 [25]. When the observed agreement is greater than the agreement expected by chance, κ takes positive values. When the observed agreement is less than the agreement expected by chance, κ takes negative values [23].

$\kappa = 1$ occurs when (and only when) there is a perfect agreement between raters. For perfect agreement, there is a necessary condition where $p_{i+} = p_{+j}$ [23]. $\kappa = 0$ indicates that the observed agreement is no better than that expected by chance as if the raters had simply *guessed* every rating [25]. $\kappa < 0$ would mean that the agreement is worse than the one expected by chance. Because the upper limit of κ is 1, then it is likely that values less than 0 mean poor agreement [23].

The κ coefficient does not indicate at all whether the disagreement is due to random differences or systematic differences between raters [25], The value of κ can be interpreted using labels assigned for different ranges, as proposed in [20] and shown in Table III.

TABLE III
LABELS TO INTERPRET DIFFERENT VALUES OF κ IN TERMS OF THE STRENGTH OF AGREEMENT

κ Range	Strength of agreement
$\kappa < 0.00$	Poor agreement
$0.00 < \kappa \leq 0.20$	Slight agreement
$0.20 < \kappa \leq 0.40$	Fair agreement
$0.40 < \kappa \leq 0.60$	Moderate agreement
$0.60 < \kappa \leq 0.80$	Substantial agreement
$0.80 < \kappa \leq 1.00$	Almost perfect agreement

In some circumstances, the original κ coefficient produces unexpected results; this problem has been referred in literature as the *kappa paradoxes* [19]. These paradoxes are related the use of marginal probabilities to quantify the agreement expected by chance p_e . In book [19] the author lists two main paradoxes:

1) If p_e is large, the correction process included in (6) can convert a relatively high value of p_0 into a relatively low value of κ.
2) If unbalanced proportions p_{i+} (or p_{+j}) produce higher values of κ than more balanced proportions.

As indicated in book [19], the application of wights on the original κ coefficient overcomes the paradoxes. Section II.C introduce the weighted version of Cohen's kappa.

C. *Cohen's weighted kappa*

The development of Cohen's weighted kappa coefficient, denoted by κ_w, was motivated by the "appearance of some disagreements in assignments, that is, some off-diagonal cells in the k x k matrix (Table II) that can have greater gravity than others" [26], and to avoid unexpected results or *kappa paradoxes* [19].

Let w_{ij} the weight for agreement assigned to the $i^{th} - j^{th}$ cell of Table II. The weighted kappa coefficient can be defined by (7) [27]:

$$\kappa_w = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} p_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} p_{i+}\, p_{+j}}{1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} p_{i+}\, p_{+j}} \tag{7}$$

Considering (7), the unweighted kappa is a special case of weighted kappa where all disagreements are given the same weight equal to 1 [26], [27]. The weights can be assigned using any judgment procedure, or, in many instances, they may result from a consensus of a committee of substantive experts [26]; in [19] the author proposed six weighting schemes for κ_w. Nevertheless, the linear and quadratic weighting schemes are the most applied in the Cohen's kappa calculation [25]–[32].

The linear weighting scheme considers the difference between categories *i* and *j*. Linear weighting scheme is defined by (8) [33]:

$$w_{ij}^{(2)} = 1 - \frac{|i - j|}{n - 1} \tag{8}$$

The quadratic weighting scheme considers the squared difference between categories *i* and *j*. Quadratic weighting scheme is defined by *equation reference goes here* as in [33]:

$$w_{ij}^{(2)} = 1 - \left(\frac{i - j}{n - 1} \right)^2 \tag{9}$$

D. *Cohen's weighted kappa test of significance*

Let H_0 be the null hypothesis stated as *raters' agreement is no better than agreement expected by chance* and let H_1 be the alternative hypothesis stated as *raters' agreement is better than agreement expected by chance*. The probability distribution of κ_w can be approximated by the Normal distribution [28]. The estimated variance when there is no association between raters' assignments, that is, when the agreement is not better than the agreement expected by chance (null hypothesis), can be calculated using (10) [33]:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} \left(p_{i+}\, p_{+j} \left[w_{ij} \left(\overline{w}_{i+} + \overline{w}_{+j} \right)^2 \right] \right) - p_e^2}{n \left(1 - p_e \right)^2} \tag{10}$$

Where $\overline{w}_{i+} = \sum_{j=1}^{k} w_{ij} p_{+j}$ represents the weighted average of the weights in the i^{th} row and $\overline{w}_{+j} = \sum_{i=1}^{k} w_{ij} p_{i+}$ represents the weighted average of the weights in the j^{th} column [33].

Assuming that $\kappa_w/\hat{\sigma}$ follows a normal distribution, it is possible to test the hypothesis of agreement expected by chance by reference to the standard normal distribution [27]. The test statistics is thus defined by (11):

$$z = \frac{\kappa_w}{\hat{\sigma}} \tag{11}$$

For a one-sided alternative, the null hypothesis H_0 is rejected if $|z| \geq z_\alpha$, where z_α is the value that leaves α in the upper tail of the standard normal distribution. In this work, the level of significance was selected as $\alpha = 0.05$ and the critical value $z_\alpha = 1.645$ [34]. Then, if the test statistics $|z| \geq 1.645$ the null hypothesis will be rejected.

III. The Measurement of Rank Agreement in the FMECA Context

In terms of the rank agreement problem, the FMECA's prioritized failure modes represent the *n* ranked objects, and the improved FMECA methods represent the *m* different judges or raters. Therefore, the rank agreement metrics can be used to evaluate the concordance between improved FMECA methods.

Cohen's concordance coefficient is a suitable metric to compare nominal rankings between two raters, therefore it was selected to compare different methods used to improve the FMECA analysis.

The Cohen's kappa can be used to measure the concordance between different improved FMECA methods applied on the same system, considering the following assumptions:

1) The failure modes represent the *n* classified objects.
2) The FMECA methods represent the m independent raters.
3) The ordinal failure modes' ranking represents the categories.

In this context, there are two possible ways to use Cohen's kappa to assess the concordance in the FMECA context to:

1) Assess the concordance between two different methods when applied to the same problem and without considering a reference one, and;
2) Assess the concordance between two different methods when applied to the same problem and regarding an FMECA ranking selected as the reference one.

Our approach considers applying the second way to measure the agreement between a pair of methods, considering a previously selected method as the reference one. That is, for the measurement of rank agreement between two FMECA results, as described below.

IV. FMECA Case Study

The selected FMECA case study corresponds to a risk assessment in the blood transfusion process analyzed using the classical FMECA in [35] and subsequently analyzed using fuzzy-based and MCDM-based FMECA approaches in [4], [7], [13], [36]. According to [35], a total number of 19 failure modes were originally identified, being the 11 failure modes with RPN higher than 80 selected for further analysis.

This FMECA case study was selected as a reference to apply the proposed concordance assessment approach considering the following reasons:

1) This study case was already used for benchmarking in some studies;
2) Because the study case has only 11 failure modes, comparing the different methods used to improve the FMECA prioritization in terms of the influence of the risk factors becomes more intuitive.

Table IV shows the FMECA analysis for the case study including the ranking obtained using the classical RPN [35].

TABLE IV
FMECA Table for the Case Study

Failure mode	Failure mode	SEV	OCC	DET	RPN	RANK
FM1	Insufficient and/or incorrect clinical information on request form	7	6	3	126	5
FM2	Blood plasma abuse	6	6	5	180	4
FM3	Insufficient preoperative assessment of the blood product requirement	7	5	7	245	1
FM4	Blood group verification incomplete	7	5	3	105	8
FM5	Delivery of blood sample and/or request form delayed	5	3	6	90	9
FM6	Incorrect blood components issued	10	1	8	80	10
FM7	Quality checks not performed on blood products	8	2	5	80	10
FM8	Preparation time before infusion >30 min	8	6	5	240	2
FM9	Transfusion cannot be completed within the appropriate time	7	4	4	112	6
FM10	Blood transfusion reaction occurs during the transfusion process	8	4	7	224	3
FM11	Bags of blood products are improperly disposed of bags	7	4	4	112	6

The failure modes FM9 and FM11 have RPN equal to 112 and both was ranked as priority 6, and failure modes FM6 and FM7 have RPN 80 and both was ranked as priority 10.

This case study was analyzed using different methods to improve the failure modes' prioritization. Table V shows the ranking for the classical FMECA analysis included in [35], the Fuzzy VIKOR based FMECA shown in [37], the interval 2-tuple hybrid weighted distance measure performed in [4], and the Risk Isosurfaces RPI(SC_4) and RPI(SC_5) proposed in [13].

Because our proposed approach requires an FMECA ranking as reference, the FMECA method denoted as RPI(SC4) proposed in [13] was selected as the reference ranking to be used to measure the concordance between it and other FMECA methods. The reasons to justify this selection are those ones declared by its authors in [13]:

1) The selected FMECA, RPI(SC4), does not require additional previous knowledge about the problem, and;
2) The failure modes prioritization agrees with the expectation made for the risk scenario.

TABLE V
DIFFERENT RANKINGS FOR FMECA IMPROVEMENT METHODS

Failure mode	RPN Rank	Fuzzy VIKOR	ITHWD	RPI(SC_4)	RPI(SC_5)
FM1	5	4	4	4	5
FM2	4	7	6	5	7
FM3	1	2	1	2	4
FM4	8	8	10	7	9
FM5	9	11	11	11	11
FM6	10	1	3	6	3
FM7	10	6	9	9	6
FM8	2	5	5	1	1
FM9	6	10	7	8	8
FM10	3	3	2	3	2
FM11	6	9	8	10	10

V. FUZZY-BASED FMECA METHODS

In addition to the FMECAs listed in Table V, this paper includes the application of FMECA approaches based on the Type-I and Type-II Fuzzy Inference Systems, and the Fuzzy Weighted Geometric Mean. These Fuzzy-FMECA approaches were also applied to the case study.

The following membership functions were considered for the Type-I and Type-II Fuzzy Inference System and for the Fuzzy Weighted Geometric Mean [8], [38], [39]:

1) Triangular membership function, denoted by *trimf*.
2) Trapezoidal membership function, denoted by *trapmf*.
3) Gaussian membership function, denoted by *gaussmf*.
4) Generalized bell membership function, denoted by *gbellmf*.

We defined eight fuzzy configurations for the Type-I Fuzzy Inference System, denoted follow as Type-I FIS, T1-FIS 01 to T-FIS 08 and as shown in Table VI.

TABLE VI
CONFIGURATIONS FOR THE FMECA BASED ON TYPE-I FUZZY INFERENCE SYSTEM

Config	Symmetry	MFSEV	MFOCC	MFDET	MFRPN
T1-FIS 01	symm	trimf	trimf	trimf	trimf
T1-FIS 02	symm	trapmf	trapmf	trapmf	trapmf
T1-FIS 03	symm	gaussmf	gaussmf	gaussmf	gaussmf
T1-FIS 04	symm	gbellmf	gbellmf	gbellmf	gbellmf
T1-FIS 05	asymm	trimf	trimf	trimf	trimf
T1-FIS 06	asymm	trapmf	trapmf	trapmf	trapmf
T1-FIS 07	asymm	gaussmf	gaussmf	gaussmf	gaussmf
T1-FIS 08	asymm	gbellmf	gbellmf	gbellmf	gbellmf

The term *symm* means that the used membership functions were all symmetrical, *asymm* means that the used membership functions were all asymmetrical, column MFSEV represents the type of membership function used to represent the severity (SEV), column MFOCC means the type of membership function used to represent the occurrence (OCC), column MFDET represents the type of membership function used to represent the detection (DET), and column MFRPN represents the type of membership function used to represent the RPN number.

We defined eight fuzzy configurations for the Fuzzy Weighted Geometric Mean, denoted as FWGM, FWGM 01 to FWGM 08, as Table VII shows.

TABLE VII
CONFIGURATIONS FOR THE FMECA BASED ON FUZZY WEIGHTED GEOMETRIC MEAN

Config	Symmetry	MFSEV	MFOCC	MFDET	MFRPN
FWGM 01	symm	trimf	trimf	trimf	trimf
FWGM 02	symm	trapmf	trapmf	trapmf	trapmf
FWGM 03	symm	gaussmf	gaussmf	gaussmf	gaussmf
FWGM 04	symm	gbellmf	gbellmf	gbellmf	gbellmf
FWGM 05	asymm	trimf	trimf	trimf	trimf
FWGM 06	asymm	trapmf	trapmf	trapmf	trapmf
FWGM 07	asymm	gaussmf	gaussmf	gaussmf	gaussmf
FWGM 08	asymm	gbellmf	gbellmf	gbellmf	gbellmf

Regarding the application of the Type-II Fuzzy Inference System (Type-II FIS), we considered an exhaustive combination of four different types of membership functions for the severity (MFSEV), occurrence (MFOCC), detection (MFDET), and RPN (MFRPN). Different Footage of Uncertainty FOU [39] and symmetry/asymmetry for each membership function, both establishing the FIS configuration, were used. The total combination of these set of parameters results in 41472 Type-II FIS configurations, denoted as T2-FIS 01 to T2-FIS 41472.

VI. RESULTS

Results are organized in two sections: Section VI.A shown the results for the linear weighted kappa and Section VI.B contains results for the quadratic weighted kappa. Each section contains the results of agreement between the reference ranking RPI(SC_4) and the following methods: 1) Fuzzy VIKOR, 2) ITHWD; 3) RPI(SC5); 4) Type-I Fuzzy Inference System; 5) Fuzzy weighted geometric mean FWGM; and 6) Type-II Fuzzy Inference System.

At the end of the section is a comparison between the reference ranking and the ranking obtained by 5 methods with highest concordance coefficient.

A. Results considering the linear weighted kappa

Table VIII shows the linear weighted concordance coefficient κ_{w-lin}, the value of the test statistics *z*, the strength of agreement and the result of the hypothesis test for the FMECA methods RPI(SC5), Fuzzy VIKOR and ITHWD, when compared with the reference ranking RPI(SC4).

The computed κ_{w-lin} takes values from 0.55 to 0.65, and the scenario RPI(SC_5) achieves the better concordance with κ_{w-lin} equal to 0.65, which can be considered a substantial agreement according to the strength of agreement proposed in Table III.

TABLE VIII
LINEAR WEIGHTED KAPPA κ_{w-lin} BETWEEN REFERENCE RANKING RPI(SC4) AND METHODS RPI(SC5), VIKOR, AND ITHWD

	RPI(SC_5)	Fuzzy VIKOR	ITHWD
κ_{w-lin}	0.65	0.55	0.6
Strength of agreement	Substantial	Moderate	Moderate
z	3.346	2.832	3.226
H_0 test	Reject	Reject	Reject

Regarding the hypothesis test, the critical value for the test statistics is $z_{0.05} = 1.645$. The value of z for the three methods is greater than test statistics, then the null hypothesis H_0 *raters' agreement is no better than agreement expected by chance* is rejected.

Table IX shows now the concordance results between RPI(SC4) and the FMECA based on the eight Type1-FIS proposed configurations (Table VI). The best concordance coefficient value is 0.55 and corresponds to configurations T1-FIS-05 and T1-FIS-08 and can be considered moderate agreement.

There are two differences between T1-FIS-05 and T1-FIS-08: the ranking of the failure modes and the shape of the membership functions. If the FMECA has few failure modes the most appropriated configuration can be selected through a qualitative comparison between the rankings. For larger FMECAs an analysis of the shape of the membership function could be necessary.

Two of the Type1-FIS configurations, T1-FIS-03 and T1-FIS-04 achieve the worst value for κ_{w-lin} as 0.3, which can be considered fair agreement. Also, in both cases, the null hypothesis H_0 was accepted because $z < z_{0.05}$.

TABLE IX
LINEAR WEIGHTED KAPPA κ_{w-lin} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-I FUZZY INFERENCE SYSTEM

	T1-FIS 01	T1-FIS 02	T1-FIS 03	T1-FIS 04	T1-FIS 05	T1-FIS 06	T1-FIS 07	T1-FIS 08
κ_{w-lin}	0.45	0.40	0.30	0.30	0.55	0.45	0.45	0.55
Strength of agreement	Moderate	Moderate	Fair	Fair	Moderate	Moderate	Moderate	Moderate
z	2.317	2.059	1.545	1.545	2.832	2.317	2.317	2.832
H_0 test	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject

Table X shows the results for the FMECA that was based on eight FWGM proposed configurations (Table VII). In this case, the concordance coefficient κ between RPI(SC_4) and each FWGM approach achieved a value between 0.40 and 0.50. In all cases the agreement is moderate and the hypothesis tests were satisfactory.

TABLE X
LINEAR WEIGHTED KAPPA κ_{w-lin} BETWEEN REFERENCE RANKING RPI(SC4) AND FUZZY WEIGHTED GEOMETRIC MEAN

	FWGM 01	FWGM 02	FWGM 03	FWGM 04	FWGM 05	FWGM 06	FWGM 07	FWGM 08
κ_{w-lin}	0.50	0.45	0.45	0.50	0.40	0.40	0.50	0.50
Strength of agreement	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
z	2.574	2.317	2.317	2.574	2.059	2.059	2.574	2.574
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

We present now in Table XI the results for the FMECA that now was based on the 41472 Type-II FIS proposed configurations. These eight configurations shown in Table XI achieved the highest value of all types of FMECA for the concordance coefficient κ as being 0.85. This can be considered as an almost perfect concordance, also having the hypothesis test H_0 being satisfactory in all scenarios. However, it must be notice that the fact that κ is equal for all the eight Type-II FIS, it does not imply that the rankings are the same.

Since the concordance coefficient is equal for the eight Type-II FIS configurations, the most appropriate configuration can be selected by conduct a qualitative analysis if the FMECA has few failure modes, or, in the case of larger FMECAs can be performed an analysis that considers the membership functions and their footprint of uncertainty FOU.

TABLE XI

LINEAR WEIGHTED KAPPA κ_{w-lin} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-II FUZZY INFERENCE SYSTEM

	T2-FIS 25361	T2-FIS 29969	T2-FIS 30033	T2-FIS 34580	T2-FIS 35089	T2-FIS 35153	T2-FIS 38673	T2-FIS 38675
κ_{w-lin}	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Strength of agreement	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect
z	4.376	4.376	4.376	4.376	4.376	4.376	4.376	4.376
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

B. Results considering the quadratic weighted kappa

Table XII shows the quadratic weighted concordance coefficient κ_{w-quad}, the value of the test statistics z, the strength of agreement and the result of the hypothesis test for the FMECA methods RPI(SC5), Fuzzy VIKOR and ITHWD, when compared with the reference ranking RPI(SC4).

TABLE XII

QUADRATIC WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND RPI(SC5), VIKOR, AND ITHWD

	RPI(SC_5)	Fuzzy VIKOR	ITHWD
κ_{w-quad}	0.855	0.727	0.809
Strength of agreement	Perfect	Substantial	Substantial
z	2.834	2.412	2.683
H_0 test	Reject	Reject	Reject

As shown, quadratic weighted kappa takes values between 0.727 and 0.855, revealing the scenario RPI(SC5) as those one achieving better concordance of 0.855, which can be considered as an almost perfect agreement.

Using now coefficient κ_{w-quad}, Table XIII shows the results for the FMECA based on the eight Type1-FIS proposed configurations.

Results obtained show the best agreement coefficient value equal to 0.8 corresponding to the configuration T1-FIS-08 (all membership functions type gbell and asymmetrical). Two of the Type1-FIS configurations (T1-FIS 03, T1-FIS 04) achieved the worst value for the agreement coefficient, 0.536, which can be considered as moderate agreement. However, the null hypothesis was rejected in both cases.

Table XIV shows the results for the FMECA based on the eight FWGM proposed configurations. wherein this case, the agreement coefficient achieves a value between 0.6 and 0.7. The agreement between these scenarios and the reference ranking can be considered a substantial agreement. In all cases, the hypothesis tests were rejected.

TABLE XIII

LINEAR WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-I FUZZY INFERENCE SYSTEM

	T1-FIS 01	T1-FIS 02	T1-FIS 03	T1-FIS 04	T1-FIS 05	T1-FIS 06	T1-FIS 07	T1-FIS 08
κ_{w-quad}	0.70	0.60	0.536	0.536	0.764	0.682	0.736	0.80
Strength of agreement	Substantial	Substantial	Moderate	Moderate	Substantial	Substantial	Substantial	Substantial
z	2.322	1.990	1.799	1.779	2.533	2.261	2.442	2.653
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

TABLE XIV

LINEAR WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND FUZZY WEIGHTED GEOMETRIC MEAN

	FWGM 01	FWGM 02	FWGM 03	FWGM 04	FWGM 05	FWGM 06	FWGM 07	FWGM 08
κ_{w-quad}	0.70	0.636	0.627	0.70	0.60	0.627	0.70	0.70
Strength of agreement	Substantial	Substantial	Substantial	Substantial	Substantial	Substantial	Substantial	Substantial
z	2.322	2.111	2.080	2.322	1.990	2.080	2.322	2.322
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

At last, the quadratic weighted kappa is used in the FMECA based the 41472 Type-2 FIS proposed configurations. Table XV shows the results achieved for the eight best scenarios. Concordance coefficient achieves its highest value, 0.973, which can be considered as an almost perfect concordance, with the null hypothesis H_0 rejected in all scenarios.

VII. DISCUSSION

Results confirm that the quadratic weighted kappa achieves concordance values greater than linear weighted kappa, as it was documented in [32]. In the FMECA context, the relationship between categories is not always linear and is difficult to establish; this relationship should determine the weighting scheme that will be used for the calculation of κ_w, however, in this paper were used the linear and quadratic weighting schemes.

TABLE XV
QUADRATIC WEIGHTED KAPPA κ_{w-quad} BETWEEN REFERENCE RANKING RPI(SC4) AND TYPE-II FUZZY INFERENCE SYSTEM

	T2-FIS 25361	T2-FIS 29969	T2-FIS 30033	T2-FIS 34580	T2-FIS 35089	T2-FIS 35153	T2-FIS 38673	T2-FIS 38675
κ_{w-quad}	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973
Strength of agreement	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect	Perfect
z	3.226	3.226	3.226	3.226	3.226	3.226	3.226	3.226
H_0 test	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

The results of the simulations show that FMECA methods that achieve a higher linear weighted kappa also achieve a higher quadratic weighted kappa. In practical terms, it can be stated that the main difference between the obtained results of kappa using the two weighting schemes (linear and quadratic) can be determined by the labels associated to the strength of agreement detailed in Table III. For example, the κ_{w-lin} obtained for the method FWGM 01 can be considered a "moderate" and its respective κ_{w-quad} obtained using the same approach can be considered as "substantial".

Nevertheless, for the scenarios T1-FIS 03 and T1-FIS 04, the null hypothesis is accepted when the linear weighting scheme is used and rejected when the quadratic weighting scheme is used. This could mean that the weighting scheme influence the statistical significance of the kappa value. A more in-depth study is needed to quantify the influence of the weighting scheme on the Cohen's kappa.Table XVI shows the ranking for the reference FMECA RPI(Sc4), the RPI(Sc5), ITHWD, T1-FIS 05, FWGM 08, and T2-FIS 38675, and their corresponding linear and quadratic weighted concordance coefficients; the rankings were ordered from highest to lowest kappa.

TABLE XVI
DIFFERENT RANKINGS FOR FMECA IMPROVEMENT METHODS

Failure mode	RPI(SC_4)	T2-FIS 38675	RPI(SC_5)	ITHWD	T1-FIS 05	FWGM 08
FM1	4	4	5	4	8	6
FM2	5	5	7	6	5	10
FM3	2	1	4	1	3	1
FM4	7	6	9	10	9	8
FM5	11	11	11	11	11	11
FM6	6	7	3	3	2	4
FM7	9	8	6	9	10	9
FM8	1	2	1	5	1	3
FM9	8	9	8	7	6	7
FM10	3	3	2	2	4	2
FM11	10	10	10	8	7	5
κ_{w-lin}	Reference	0.850	0.650	0.60	0.550	0.50
κ_{w-quad}	Reference	0.973	0.855	0.809	0.764	0.70

Because the FMECA case study has only a few failure modes, it is possible to identify the concordances and discordances between the five FMECA methods. The ranking for failure modes FM1, FM2, FM5, FM10 and FM11 are the same for the reference RPI(SC_4) and T2-FIS 38675; both models agree 5 times and disagree 6 times. Comparing the base case with RPI(SC5), the rankings agree 4 times and disagree 7 times. For ITHWD and T1-FIS 05, the rankings agree 3 times and disagree 8 times.

For FWGM 08, the rankings agree 2 times and disagree 9 times. Notice, that the number of agreements and disagreements can indicate the level of concordance between two raters in a simple way, however, it does not provide an effective metric to measure it; the Cohen's coefficient deals with this issue and also gives a concordance level based on the coincidences between ratings and the agreement that occurs by chance.

Fig. 1 shows a radar chart for the three best FMECA methods (RPI(Sc5), ITHWD, and T2-FIS 38675) and the reference one. The chart greatly simplifies the comparison between the different rankings assigned to each failure mode. The blue line in Fig. 1 represents the reference FMECA ranking and the red line represents the ranking for the method with highest κ (T2-FIS 38675).

When compared with the reference raking RPI(Sc4), the approach T2-FIS 38675 has perfect agreement in 5 failure modes (FM1, FM2, FM5, FM10 and FM11), the approach RPI(Sc5) has perfect agreement in 4 failure modes (FM5, FM8, FM9 and FM11), and the approach ITHWD has perfect agreement in 3 failure modes (FM1, FM5 and FM7).

VIII. CONCLUSION AND FUTURE WORK

This paper introduces an approach based on the Cohen's kappa concordance coefficient to compare different methods used in the FMECA context. A simple and further analyzed case study was selected to conduct the comparisons. FMECA approaches based on Type-I Fuzzy Inference System, Fuzzy Weighted Geometric Mean, and Type-II Fuzzy Inference System were developed and conducted to rank the failure modes. From our results and its previous discussion, one pulls out four critical conclusions:

1) The comparison between different FMECA methods is commonly based on the qualitative comparison of ranking and balance between the three risk factors; nevertheless, this approach can be impractical for more extensive problems.
2) The proposed approach aims to contribute to the quantitative comparison between methods used to improve the prioritization of failure modes regarding a

reference ranking.

3) The results shown that the Cohen's κ coefficient gives a quantitative level for the agreement between two different rankings in the FMECA analysis context.
4) The ranking based on Type-II Fuzzy Inference System's achieves the best agreement regarding the reference FMECA method. This occurs due to the uncertainty being considered now as an additional parameter in the fuzzy inference process.
5) The selection of the weighting scheme is another essential aspect to take into account in the proposed approach; since the relationship between categories in FMECA's risk factors is not linear, results show that quadratic weighting scheme allows obtaining a better strength of agreement.
6) The reference FMECA's ranking identification is a critical aspect for the success of the proposed approach.

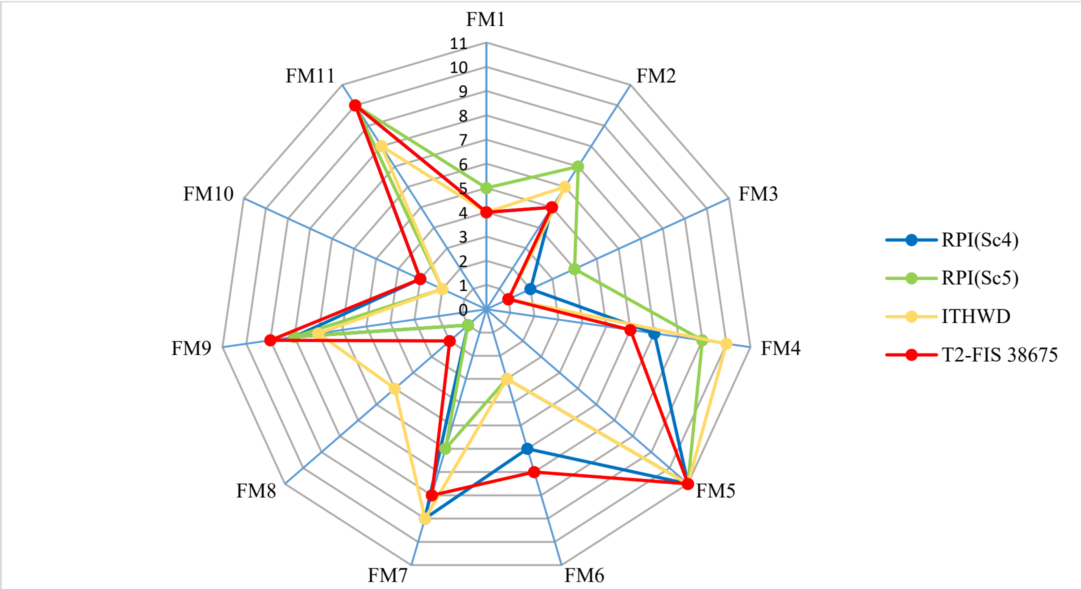

Fig. 1. Radar chart showing the reference FMECA ranking (blue line) and the three best approaches.

The main shortcoming in our proposed approach is the selection of the reference FMECA ranking. In practical applications, it can be demanding to identify a suitable FMECA reference ranking. An acceptable procedure to conduct this kind of comparison could be to apply different FMECA approaches to a well-known problem whose failure modes' ranking can be considered as optimal and then compute the concordance coefficient to identify the best FMECA method respect to this reference. Once the best FMECA method is identified, it can be applied to another case study with similar characteristics. The solution to this shortcoming is being addressed and will be included in future works.

Additional aspects are currently in development and will be include in forthcoming works:

1) The application of the proposed approach in the context of smart substations;
2) The definition of *tailor-made* scales for the FMECA risk factors in the context of smart substations;
3) The proposal of a new weighting scheme based on the above-mentioned risk factors' scales;
4) The use of paradox-resistant concordance coefficients.

Acknowledgment

The authors express special thanks to Professor Blas Galván González for reading an earlier version of this paper and offer many insightful comments and suggestions.

References

[1] International Electrotechnical Comission, *IEC 60812:2006 - Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA)*. 2006.

[2] X. Li, M. He, and H. Wang, "Application of failure mode and effect analysis in managing catheter-related blood stream infection in intensive care unit," *Medicine (United States)*, vol. 96, no. 51, 2017, doi: 10.1097/MD.0000000000009339.

[3] J. B. Bowles and C. E. Peláez, "Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis," *Reliability Engineering and System Safety*, vol. 50, no. 2, pp. 203–213, 1995, doi: 10.1016/0951-8320(95)00068-D.

[4] H. C. Liu, J. X. You, and X. Y. You, "Evaluating the risk of healthcare failure modes using interval 2-tuple hybrid weighted distance measure," *Computers and Industrial Engineering*, vol. 78, pp. 249–258, 2014, doi: 10.1016/j.cie.2014.07.018.

[5] A. A. Zúñiga, A. Baleia, J. Fernandes, and P. J. da Costa Branco, "Classical failure modes and effects analysis in the context of smart grid cyber-physical systems," *Energies*, vol. 13, no. 5, pp. 1–26, 2020, doi: 10.3390/en13051215.

[6] J. Huang, J. X. You, H. C. Liu, and M. S. Song, "Failure mode and effect analysis improvement: A systematic literature review and future research agenda," *Reliability Engineering and System Safety*, vol. 199, no. February, p. 106885, 2020, doi: 10.1016/j.ress.2020.106885.

[7] H. C. Liu, *FMEA using uncertainty theories and MCDM methods*. Singapore: Springer Science+Business Media Singapore, 2016.

[8] A. A. Zúñiga, J. F. P. Fernandes, and P. J. C. Branco, "A Fuzzy-Based Failure Modes and Effects Analysis (FMEA) in Smart Grids," in *International Conference on Information Technology & Systems ICITS 2019. Advances in Intelligent Systems and Computing*, vol. 918, Á. Rocha, C. Ferrás, and M. Paredes, Eds. Cham: Springer,

Cham, 2019, pp. 507–516.
[9] A. Pillay and J. Wang, "Modified failure mode and effects analysis using approximate reasoning," *Reliability Engineering and System Safety*, vol. 79, no. 1, pp. 69–85, 2003, doi: 10.1016/S0951-8320(02)00179-5.
[10] E. Akyuz and E. Celik, "A quantitative risk analysis by using interval type-2 fuzzy FMEA approach: the case of oil spill," *Maritime Policy and Management*, vol. 45, no. 8, pp. 979–994, 2018, doi: 10.1080/03088839.2018.1520401.
[11] E. Bozdag, U. Asan, A. Soyer, and S. Serdarasan, "Risk prioritization in failure mode and effects analysis using interval type-2 fuzzy sets," *Expert Systems with Applications*, vol. 42, no. 8, pp. 4000–4015, 2015, doi: 10.1016/j.eswa.2015.01.015.
[12] J. Qin, Y. Xi, and W. Pedrycz, "Failure mode and effects analysis (FMEA) for risk assessment based on interval type-2 fuzzy evidential reasoning method," *Applied Soft Computing Journal*, vol. 89, p. 106134, 2020, doi: 10.1016/j.asoc.2020.106134.
[13] V. Anes, E. Henriques, M. Freitas, and L. Reis, "A new risk prioritization model for failure mode and effects analysis," *Quality and Reliability Engineering International*, vol. 34, no. 4, pp. 516–528, 2018, doi: 10.1002/qre.2269.
[14] T. Karaulova, M. Kostina, and E. Shevtshenko, "Reliability assessment of manufacturing processes," *International Journal of Industrial Engineering and Management*, vol. 3, no. 3, pp. 143–151, 2012.
[15] P. C. Okwesili Jr., "Risk Assessmnt Using Paired Comparison Expert Elicitation for Ranking of Compounding Outsourcing Facilities," The George Washington University, 2016.
[16] H. Frewen, E. Brown, M. Jenkins, and A. O'Donovan, "Failure mode and effects analysis in a paperless radiotherapy department," *Journal of Medical Imaging and Radiation Oncology*, vol. 62, no. 5, pp. 707–715, 2018, doi: 10.1111/1754-9485.12762.
[17] M. G. Kendall and B. B. Smith, "The Problem of m Rankings," *The Annals of Mathematical Statistics*, vol. 10, no. 3, pp. 275–287, 1939, doi: 10.1214/aoms/1177732186.
[18] J. Liu, W. Tang, G. Chen, Y. Lu, C. Feng, and X. M. Tu, "Correlation and agreement: overview and clarification of competing concepts and measures," *Shanghai Archives of Psychiatry*, vol. 28, no. 2, pp. 115–120, 2016, doi: 10.11919/j.issn.1002-0829.216045.
[19] K. L. Gwet, *Handbook of Inter-Rater Reliability*, 4th ed. Gaithersburg: Advanced Analytics, LLC, 2014.
[20] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, p. 159, 1977, doi: 10.2307/2529310.
[21] T. R. Vetter and P. Schober, "Agreement analysis: What he said, she said versus you said," *Anesthesia and Analgesia*, vol. 126, no. 6, pp. 2123–2128, 2018, doi: 10.1213/ANE.0000000000002924.
[22] J. Teles, "Concordance coefficients to measure the agreement among several sets of ranks," *Journal of Applied Statistics*, vol. 39, no. 8, pp. 1749–1764, 2012, doi: 10.1080/02664763.2012.681460.
[23] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.
[24] M. L. McHugh, "Lessons in biostatistics interrater reliability: the kappa statistic," *Biochemica Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[25] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: Use, interpretation, and sample size requirements," *Physical Therapy*, vol. 85, no. 3, pp. 257–268, 2005, doi: 10.1093/ptj/85.3.257.
[26] J. Cohen, "Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968, doi: https://doi.org/10.1037/h0026256.
[27] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, "Beyond kappa: A review of interrater agreement measures," *Canadian Journal of Statistics*, vol. 27, no. 1, pp. 3–23, 1999, doi: 10.2307/3315487.
[28] A. Ben-David, "Comparison of classification accuracy using Cohen's Weighted Kappa," *Expert Systems with Applications*, vol. 34, no. 2, pp. 825–832, 2008, doi: 10.1016/j.eswa.2006.10.022.
[29] I. Guggenmoos-Holzmann, "The meaning of kappa: Probabilistic concepts of reliability and validity revisited," *Journal of Clinical Epidemiology*, vol. 49, no. 7, pp. 775–782, 1996, doi: 10.1016/0895-4356(96)00011-X.
[30] P. W. Mielke and K. J. Berry, "A note on Cohen's weighted kappa coefficient of agreement with linear weights," *Statistical Methodology*, vol. 6, no. 5, pp. 439–446, 2009, doi: 10.1016/j.stamet.2009.03.002.
[31] M. J. Warrens, "Cohen's kappa is a weighted average," *Statistical Methodology*, vol. 8, no. 6, pp. 473–484, 2011, doi: 10.1016/j.stamet.2011.06.002.
[32] M. J. Warrens, "Cohen's quadratically weighted kappa is higher than linearly weighted kappa for tridiagonal agreement tables," *Statistical Methodology*, vol. 9, no. 3, pp. 440–444, 2012, doi: 10.1016/j.stamet.2011.08.006.
[33] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, no. 5, pp. 323–327, 1969, doi: 10.1037/h0028106.
[34] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Sixth Edit. London: Academic Press, 2021.
[35] Y. Lu, F. Teng, J. Zhou, A. Wen, and Y. Bi, "Failure mode and effect analysis in blood transfusion: A proactive tool to reduce risks," *Transfusion*, vol. 53, no. 12, pp. 3080–3087, 2013, doi: 10.1111/trf.12174.
[36] H.-C. Liu, *Improved FMEA Methods for Proactive Healthcare Risk Analysis*. Singapore: Springer, 2019.
[37] H. C. Liu, L. Liu, N. Liu, and L. X. Mao, "Risk evaluation in failure mode and effects analysis with extended VIKOR method under fuzzy environment," *Expert Systems with Applications*, vol. 39, no. 17, pp. 12926–12934, 2012, doi: 10.1016/j.eswa.2012.05.031.
[38] C. Kao and S. T. Liu, "Fractional programming approach to fuzzy weighted average," *Fuzzy Sets and Systems*, vol. 120, no. 3, pp. 435–444, 2001, doi: https://doi.org/10.1016/S0165-0114(99)00137-2.
[39] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, 2nd ed. Cham: Springer International Publishing, 2017.

Andrés A. Zúñiga (M'17) received the B.S. degree in electrical engineering from Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador, in 2015. He is currently pursuing the Ph.D. degree in electrical engineering at University of Lisbon, Lisbon, Portugal. From 2018 to 2019, he was Invited Lecturer at the Department of Electrical and Computer Engineering, Scientific Area of Energy, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.

Since 2017, his research has been developed with the collaboration of Instituto de Engenharia Mecânica (IDMEC) from the Laboratório Associado de Energia Transportes e Aeronáutiva (LAETA) and supported by a fellowship from Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) of the Ecuadorian Government since 2017. His current research interests include reliability and risk analysis in smart grids environment.

João F. P. Fernandes received the Ph.D. degree in electrical and computers engineering from Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal, in 2017.

From 2006 to 2013 he was Managing Partner of the company Emove, Innovative Technologies, Lda, Lisbon, Portugal, and Electrical Engineer with company Efacec, Engenharia e Sistemas. Since 2018, he is Assistant Professor with the Department of Electrical and Computer Engineering, Scientific Area of Energy, of Instituto Superior Técnico, Universidade de Lisboa. His research has been developed with the collaboration of Instituto de Engenharia Mecânica (IDMEC) from the Laboratório Associado de Energia

Transportes e Aeronáutica (LAETA). He has published several works these areas, in international refereed journals and international refereed conferences. His research interests include the development of new machines' designs, the application of new materials in electrical machines, as soft composites and superconducting materials, and the development of hybrid field computing tools.

Dr. Fernandes is a Reviewer of several international journals, as IEEE, Elsevier, and others.

Paulo J. Costa Branco (M'91) received the B.S. and M.S. degrees in electrical and computer engineering from the Federal University of Rio de Janeiro (UFRJ) and COPPE/UFRJ, Rio de Janeiro, Brazil, in 1990 and 1992, respectively.

He is currently an Associate Professor (with habilitation) with the Department of Electrical and Computer Engineering, Scientific Area of Energy, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal. His research interests include design of innovative spherical induction machines for biomedical and robotic applications, project of innovative electromechanical systems supported on ionic polymer metal-composite materials, and the use of high-temperature superconducting materials in increasing the electric energy efficiency of electrical machines and MagLev.