

## Article

# A New Hybrid Possibilistic-Probabilistic Decision-Making Scheme for Classification

Basel Solaiman, Didier Guériot, Shaban Almouahed, Bassem Alsahwa and Éloi Bossé \*

Image & Information Processing Department (iTi), IMT-Atlantique, Technopôle Brest Iroise CS 83818, 29238 Brest, France; basel.solaiman@imt-atlantique.fr (B.S.); didier.gueriot@imt-atlantique.fr (D.G.); shaban.almouahed@bsb-marine.com (S.A.); bassems74@netcourrier.com (B.A.)

\* Correspondence: ebosse861@gmail.com

**Abstract:** Uncertainty is at the heart of decision-making processes in most real-world applications. Uncertainty can be broadly categorized into two types: aleatory and epistemic. Aleatory uncertainty describes the variability in the physical system where sensors provide information (hard) of a probabilistic type. Epistemic uncertainty appears when the information is incomplete or vague such as judgments or human expert appreciations in linguistic form. Linguistic information (soft) typically introduces a possibilistic type of uncertainty. This paper is concerned with the problem of classification where the available information, concerning the observed features, may be of a probabilistic nature for some features, and of a possibilistic nature for some others. In this configuration, most encountered studies transform one of the two information types into the other form, and then apply either classical Bayesian-based or possibilistic-based decision-making criteria. In this paper, a new hybrid decision-making scheme is proposed for classification when hard and soft information sources are present. A new Possibilistic Maximum Likelihood (PML) criterion is introduced to improve classification rates compared to a classical approach using only information from hard sources. The proposed PML allows to jointly exploit both probabilistic and possibilistic sources within the same probabilistic decision-making framework, without imposing to convert the possibilistic sources into probabilistic ones, and vice versa.

**Keywords:** possibility theory; possibilistic decision rule; possibilistic maximum likelihood; pattern classification; uncertainty; Bayesian decision; maximum *a posteriori*; image processing

**Citation:** Solaiman, B.; Guériot, D.; Almouahed, S.; Alsahwa, B.; Bossé, É. A New Hybrid Possibilistic-Probabilistic Decision-Making Scheme for Classification. *Entropy* **2021**, *23*, 67. <https://doi.org/10.3390/e23010067>

Received: 13 November 2020

Accepted: 30 December 2020

Published: 3 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

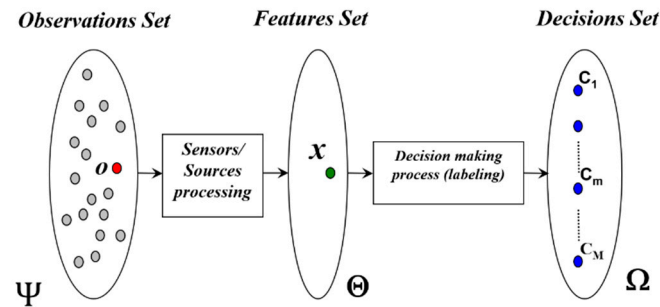
## 1. Introduction

Uncertainty can be categorized into two main kinds [1]: aleatory or randomness uncertainty, aka statistical uncertainty, due to the variability or the natural randomness in a process and epistemic uncertainty, aka systematic uncertainty, which is the scientific uncertainty in the model of the process. It is due to limited data and knowledge. Epistemic uncertainty calls for alternative methods of representation, propagation, and interpretation of uncertainty than just probability. Since the beginning of the 60 s, following fruitful cross-fertilization, a convergence is emerging between physics, engineering, mathematics, and the cognitive sciences to provide new techniques and models that shows a trend of inspiration from human brain mechanism towards a unified theory to represent knowledge, belief and uncertainty [2–9].

Uncertainty is a natural and unavoidable part in real-world applications. When observing a “real-world situation”, decision making is the process of selecting among several alternatives or decisions.

The problem here is to assign a label or a class to measurements or other types of observations (data) from sensors or other sources to which the observations are assumed to belong. This is a typical classification process.

As shown in Figure 1, the general classification process can be formulated as follows. An input set of observations  $\mathbf{o}$  ( $\mathbf{o} \in \Psi$ ) is “observed” using a sensor (or a set of sensors) delivering a feature vector  $\mathbf{x} \in \Theta$  ( $\Theta$  is called the features set). This feature vector  $\mathbf{x}$  is then injected into the decision-making system or labelling system in order to recognize the most likely decision (hypothesis, alternative, class) from a given exhaustive set  $\Omega = \{C_m, m = 1, \dots, M\}$  of  $M$  exclusive decisions [10].



**Figure 1.** Structure of a multisource classification system (Source: [11]).

The development of both classification algorithms and decision-making criteria are governed by several factors mainly depending on the nature of the feature vector, the nature of the imperfection attached to the observed features as well as the available knowledge characterizing each decision. Several global constraints also drive the conception of the global classification process: the “physical” nature and quality of the measures delivered by the sensors, the categories discrimination capacity of the computed features, the nature and the quality of the available knowledge used for the development of the decision-making system.

However, in much of the literature, the decision-making system is performed by the application of two successive functionalities: the *soft labeling* and the *hard decision (selection)* functionalities. The labeling functionality [12] uses the available a priori knowledge in order to perform a mapping  $\ell$  between the features set  $\Theta$  and the decisions set  $\Omega$  ( $\ell: \Theta \rightarrow \Omega$ ). For each feature vector  $x \in \Theta$ , a *soft decision label* vector  $\ell(x) = [\ell_{c_1}(x), \dots, \ell_{c_m}(x), \dots, \ell_M(x)] \in [0,1]^M$  is determined in the light of the available knowledge where  $\ell_{c_m}(x)$  measures the degree of belief or support, that we have in the occurrence of the decision  $C_m$ . For instance, if the available knowledge allows probabilistic computations, the soft decision label vector is given through  $\ell_{c_M}(x) = Pr\{C_M|x\}$  [13] where  $Pr\{C_M|x\}$  represents the *a posteriori* probability of the decision  $C_M$  given the observed feature vector  $x \in \Theta$  [14]. When the available knowledge is expressed in terms of ambiguous information, the possibility theory formalism (developed by L. Zadeh [7] and D. Dubois et al. [15–17]) can be used. The soft decision label vector  $\ell(x)$  is then expressed with an *a posteriori* possibility distribution  $\pi_x$  defined on the decisions set  $\Omega$ . In this case,  $\ell_{c_m}(x) = \pi_x(C_m)$  where  $\pi_x(C_m)$  represents the possibility degree for the decision  $C_m$  to occur, given the observed feature vector  $x \in \Theta$ .

The second functionality performed by the decision-making process is called the *hard decision* or the *selection functionality*. As the ultimate goal of most classification applications is to select one and only one class (associated with the observations “ $\mathbf{o}$ ” for which the feature vector  $x \in \Theta$  is extracted) out of the classes set  $\Omega$ , then a mapping has to be applied in order to transform the soft decision label vector  $\ell(x)$  into a hard decision label vector for which one and only one decision is selected. The goal is then to make a choice according to an optimality criterion.

In this paper, we propose a new criterion for decision-making process in classification systems called possibilistic maximum likelihood (PML). This criterion is framed within the possibility theory, but it uses corresponding notions from Bayesian decision-

making. The main motivation being the development of PML is for multisource information fusion where an object or a pattern may be observed through several channels and where the available information, concerning the observed features, may be of a probabilistic nature for some features, and of an epistemic nature for some others.

In the presence of both types of information sources, most encountered studies transform one of the two information types into the other form, and then apply either the classical Bayesian or possibilistic decision-making criteria. With the PML decision-making approach, the Bayesian decision-making framework is adopted. The epistemic knowledge is integrated into the decision-making process by defining possibilistic loss values instead of the usually used zero-one loss values. A set of possibilistic loss values is proposed and evaluated in the context of pixel-based image classification where a synthetic scene, composed of several thematic classes, is randomly generated using two types of probabilistic sensors: a Gaussian and a Rayleigh sensor, complemented by an expert type of information source. Results obtained with the proposed PML criterion show that the classification recognition rates approach the optimal case, being, when all the available information is expressed in terms of probabilistic knowledge.

When the sources of information can be modelled by probability theory, the Bayesian approach has sufficient decision-making tools to fuse that information and performs classification. However, in the case where the knowledge available for the decision-making process is ill-defined in the sense that it is totally or partially expressed in terms of ambiguous information representing limitations in feature values, or, encoding linguistic expert's knowledge about the relationship between the feature values and different potential decisions, new mathematical tools (i.e., PML) need to be developed. This type of available knowledge can be represented as a conditional possibilistic soft decision label vector  $\ell(x)$  defined on the decisions set  $\Omega$  such that,  $\ell_{C_m}(x) = \pi_x(C_m) = \pi(C_m|x)$  where  $\pi(C_m|x)$  represents the possibility degree for the decision  $C_m$  to occur, given the observed feature vector  $x \in \Theta$  and the underlying observations  $o$ .

Possibility theory constitutes the natural framework allowing to tackle this type of information imperfection (called the *epistemic uncertainty* type) when one and only one decision (*hard decision*) must be selected from the exhaustive decisions set  $\Omega$ , with incomplete, ill-defined or ambiguous available knowledge thus encoded as a possibility distribution over  $\Omega$ . This paper proposes a joint decision-making criterion which allows to integrate such extra possibilistic knowledge within a probabilistic decision-making framework taking into account both types of information: possibilistic and probabilistic. In spite of the fact that possibility theory deals with uncertainty, which means that a unique but unknown elementary decision is to occur, and the ultimate goal is to determine this decision, there are relatively few studies that tackle that decision-making issue [18–25]. We must however mention the considerable contributions of Dubois and Prade [26] on possibility theory as well as on clarification on the various semantics of fuzzy sets [27–30]. Denoeux et al. [31–33] contributed as well significantly on that topic but they consider epistemic uncertainty as a higher order uncertainty upon probabilistic models such as in imprecise probabilities of Walley [34,35] and fuzzy sets type-2 [36–38] which is not being the case in this current paper.

The paper is organized this way. A brief recall of the Bayesian decision-making criteria, and of possibility theory is given in Sections 2 and 3. Three major possibilistic decision making criteria, i.e., maximum possibility, maximum necessity measure and confidence index maximization, are being detailed in Section 4. The PML criterion is presented in Section 5 followed by its evaluation in Section 6. Paper closes with conclusion in Section 7.

## 2. Hard Decision in the Bayesian Framework

In the Bayesian classification framework, the most widely used hard decision is based on minimizing an overall decision risk function [14]. Assuming  $o \in \Psi$  is the pattern for which the feature vector  $x \in \Theta$  is observed, let  $\lambda_{m,n}$  denotes a “predefined” conditional loss or penalty, incurred for deciding that the observed pattern  $o$  is associated with the

decision  $C_n$ , whereas the true decision (class or category) for  $\mathbf{o}$  is  $C_m (n, m \in \{1, \dots, M\})$ . Therefore, the probabilistic expected loss  $R(C_n|x)$ , also called the *Conditional risk*, associated with the decision  $C_n$  given the observed feature vector  $x \in \Theta$ , is given by:

$$R(C_n|x) = E\{\lambda_{m,n}\} = \sum_{m=1}^M \lambda_{m,n} Pr\{C_m|x\} \quad (1)$$

where  $E\{\cdot\}$  stands for the mathematical expectation. Bayes decision criterion consists in minimizing the overall risk  $R$ , also called *Bayes risk*, as defined in (2), by computing the conditional risk for all decisions and then, selecting the decision  $C_n$  for which  $R(C_n|x)$  is minimum:

$$R(C_n|x) = E_x\{R(C_n|x)\} = \int R(C_n|x) Pr\{x\} dx \quad (2)$$

Therefore, the *minimum-risk Bayes decision criterion* is based on the selection of the decision  $C_n$  which gives the smallest risk  $R(C_n|x)$ . This rule can thus be formulated as follows:

$$Decision[x(p)] = \arg \min_{n=1, \dots, M} \left( \sum_{m=1}^M \lambda_{m,n} Pr\{C_m|x\} \right) \quad (3)$$

If  $Pr\{C_m\}$  denotes the *a priori* probability of the decision  $C_m$  and  $Pr\{x|C_m\}$ , the likelihood function of the measured feature vector  $x$ , given the decision  $C_m$ , then using Bayes' rule, the minimum-risk Bayes decision criterion (3) can be rewritten as:

$$Decision[x(p)] = \arg \min_{n=1, \dots, M} \left( \sum_{m=1}^M Pr\{C_m\} Pr\{x|C_m\} \right) \quad (4)$$

In the two-category decision case, i.e.,  $\Omega = \{C_1, C_2\}$ , it can be easily shown that the minimum-risk Bayes decision criterion, simply called *Bayes criterion*, can be expressed as in (5):

$$LR = \frac{Pr\{x|C_1\}}{Pr\{x|C_2\}} \underset{Decision[x(p)]=C_2}{\overset{Decision[x(p)]=C_1}{\gtrless}} \underbrace{\frac{\lambda_{2,2} - \lambda_{2,1}}{\lambda_{1,1} - \lambda_{1,2}}}_{\eta} \cdot \frac{Pr\{C_2\}}{Pr\{C_1\}} \quad (5)$$

In other words, this decision criterion consists of comparing the likelihood ratio ( $LR$ )  $Pr\{x|C_1\}/Pr\{x|C_2\}$  to a threshold  $\eta$  independent of the observed feature vector  $x$ . The binary cost, or zero-one loss, assignment is commonly used in classification problems. This rule, expressed in (6), gives  $\lambda_{m,n}$  no cost for a correct decision (when the true pattern class/decision  $C_m$  is identical to the decided class/decision  $C_n$ ) and a unit cost for a wrong decision (when the true class/decision  $C_m$  is different from the decided class/decision  $C_n$ ).

$$\lambda_{m,n} = \begin{cases} 0 & \text{if } C_m = C_n \\ 1 & \text{if } C_m \neq C_n \end{cases} \quad (6)$$

It should be noticed that this binary cost assignment considers all errors as equally costly. It also leads to express the conditional risk as:

$$R(C_n|x) = \sum_{m=1}^M \lambda_{m,n} Pr\{C_m|x\} = 1 - Pr\{C_n|x\} \quad (7)$$

A decision minimizing the conditional risk  $R(C_n|x)$  becomes a decision maximizing the *a posteriori* probability  $Pr\{C_n|x\}$ . As shown in (8), this version of the Bayes criterion is

called the maximum a posteriori criterion (MAP) since it seeks to determine the decision maximizing the *a posteriori* probability value. It is also obvious that this decision process corresponds to the minimum-error decision rule which leads to the best recognition rate that a decision criterion can achieve:

$$Decision_{MAP}[x(p)] = \arg \max_{n=1, \dots, M} Pr\{C_n|x\} \quad (8)$$

When the decisions *a priori* probabilities  $Pr\{C_m\}$  and the likelihood functions  $Pr\{x|C_m\}$  are not available, or simply difficult to obtain, the *Minmax Probabilistic Criterion* (MPC) can be an interesting alternative to the minimum-risk Bayes decision criterion [39]. As expressed in (9), this hard decision criterion consists in selecting the decision that minimizes the maximum decision cost:

$$Decision_{MPC}[x(p)] = \arg \min_{n=1, \dots, M} \left[ \max_{m=1, \dots, M} \{\lambda_{m,n} Pr\{C_m|x\}\} \right] \quad (9)$$

### 3. Brief Review of Possibility Theory

Possibility theory is a relatively new theory devoted to handle uncertainty in the context where the available knowledge is only expressed in an ambiguous form. This theory was first introduced by Zadeh in 1978 as an extension of fuzzy sets and fuzzy logic theory, to express the intrinsic fuzziness of natural languages as well as uncertain information [7]. It is well established that probabilistic reasoning, based on the use of a probability measure, constitutes the optimal approach dealing with uncertainty. In the case where the available knowledge is ambiguous and encoded by a membership function, i.e., a fuzzy set, defined over the decisions set, the possibility theory transforms the membership function into a possibility distribution  $\pi$ . Then the realization of each event (subset of the decisions set) is bounded by a possibilistic interval defined through a possibility,  $\Pi$ , and a necessity,  $N$ , measures [16]. The use of these two dual measures in possibility theory makes the main difference compared with the probability theory. Besides, possibility theory is not additive in terms of beliefs combination, and makes sense on ordinal structures [17]. In the following subsections, the basic concepts of a possibility distribution and the dual possibilistic measures (possibility and necessity measures) will be presented. The possibilistic decision rules will be detailed in Section 4. Full details can be found in [11].

#### 3.1. Possibility Distribution

Let  $\Omega = \{C_1, C_2, \dots, C_M\}$  be a finite and exhaustive set of  $M$  mutually exclusive elementary decisions (e.g., decisions, thematic classes, hypothesis, etc.). Exclusiveness means that one and only one decision may occur at one time, whereas exhaustiveness states that the occurring decision certainly belongs to  $\Omega$ . Possibility theory is based on the notion of possibility distribution denoted by  $\pi$ , which maps elementary decisions from  $\Omega$  to the interval  $[0, 1]$ , thus encoding “our” state of knowledge or belief, on the possible occurrence of each class  $C_m \in \Omega$ . The value  $\pi(C_m)$  represents to what extent it is possible for  $C_m$  to be the unique occurring decision. In this context, two extreme cases of knowledge are given:

- Complete knowledge:  $\exists! C_m \in \Omega, \pi(C_m) = 1$  and  $\pi(C_n) = 0, \forall C_n \in \Omega, C_n \neq C_m$ .
- Complete ignorance:  $\forall C_m \in \Omega, \pi(C_m) = 1$  (all elements from  $\Omega$  are considered as totally possible).  $\pi(\cdot)$  is called a normal possibility distribution if it exists at least one element  $C_{m_0}$  from  $\Omega$  such that  $\pi(C_{m_0}) = 1$ .

#### 3.2. Possibility and Necessity Measures

Based on the possibility distribution concept, two dual set measures, possibility,  $\Pi$ , and a necessity,  $N$ , measures are derived. For every subset (or event)  $A \subseteq \Omega$ , these measures are defined by:

$$\Pi(A) = \max_{C_m \in A} [\pi(C_m)] \quad (10a)$$

$$N(A) = 1 - \Pi(A^c) = \min_{C_m \notin A^c} [1 - \pi(C_m)] \quad (10b)$$

where  $A^c$  denotes the complement of the event  $A$  (i.e.,  $A \cup A^c = \Omega$  with  $A \cap A^c = \emptyset$ ).

The possibility measure  $\Pi(A)$  estimates the level of consistency about event  $A$  occurrence, given the available knowledge encoded by the possibility distribution  $\pi$ . Thus,  $\Pi(A) = 0$  means that  $A$  is an impossible event while  $\Pi(A) = 1$  means that the event  $A$  is totally possible. The necessity measure  $N(A)$  evaluates the level of certainty about event  $A$  occurrence, involved by possibility distribution  $\pi$ .  $N(A) = 0$  means that the certainty about the occurrence of  $A$  is null. On the contrary,  $N(A) = 1$  means that the occurrence of  $A$  is totally certain. In a classification problem, where each decision  $C_m$  refers to a given class or category, the case where all events  $A$  are composed of a single decision ( $A_m = \{C_m\}, m = 1, \dots, M$ ), is of particular interest. In this case, the possibility  $\Pi(\cdot)$ , and the necessity  $N(\cdot)$ , measures are reduced to:

$$\Pi(A_m) = \Pi(\{C_m\}) = \pi(C_m) \quad (11a)$$

$$N(A_m) = N(\{C_m\}) = 1 - \Pi(\{C_m\}^c) = 1 - \max_{n \neq m} \pi(C_n) \quad (11b)$$

#### 4. Decision-Making in the Possibility Theory Framework

In this section, we will investigate existing possibilistic decision-making rules. Two families of rules can be distinguished: rules based on the direct use of the information encapsulated in the possibility distribution, and rules based on the use of uncertainty measures associated with this possibility distribution. Let  $\Omega = \{C_1, C_2, \dots, C_M\}$  be a finite and exhaustive set of  $M$  mutually exclusive elementary decisions. Given an observed pattern  $\mathbf{o} \in \Psi$  for which the feature vector  $\mathbf{x} \in \Theta$  is observed, let  $\pi_x(C_m)$  denotes the *a posteriori* possibility distribution  $\pi(C_m|\mathbf{x})$  defined on  $\Omega$ . The possibility,  $\Pi_x(\{C_m\})$ , and necessity,  $N_x(\{C_m\})$ , measures are obtained as expressed in Equation (11), using the possibility distribution  $\pi_x(C_m)$ .

##### 4.1. Decision Rule Based on the Maximum of Possibility

The decision rule based on the maximum of possibility is certainly the most widely used in possibilistic classification—decision-making applications. Indeed, as shown in (12), this rule is based on the selection of the elementary decision  $C_{m_0} \in \Omega$  having the highest possibility degree of occurrence  $\Pi_x(\{C_{m_0}\})$ :

$$\text{Decision } [x(p)] = C_{m_0} \text{ if and only if } m_0 = \arg \max_{n=1, \dots, M} \Pi_x(\{C_m\}) \quad (12)$$

A “first” mathematical justification of this “intuitive” possibilistic decision-making rule can be derived from the *Minmax Probabilistic Criterion* (MPC), Equation (9), using a binary cost assignment rule. Indeed, ‘converting’ the *a posteriori* possibility distributions  $\pi_x(\cdot)$  into *a posteriori* probability distributions  $Pr\{\cdot|\mathbf{x}\}$  is assumed to respect the three following constraints [30]: (a) the consistency principle, (b) the preference ordering preservation, and (c) the least commitment principle. The preference ordering preservation, on which we focus the attention here, means that if decision  $C_{m_1}$  is preferred to decision  $C_{m_2}$ , i.e.,  $\pi_x(C_{m_1}) > \pi_x(C_{m_2})$ , then the *a posteriori* probability distribution  $Pr\{\cdot|\mathbf{x}\}$  obtained from  $\pi_x(\cdot)$  should satisfy  $Pr\{C_{m_1}|\mathbf{x}\} > Pr\{C_{m_2}|\mathbf{x}\}$ . Equation (13) sums up this preference ordering preservation constraint:

$$\pi_x(C_{m_1}) > \pi_x(C_{m_2}) \Leftrightarrow Pr\{C_{m_1}|\mathbf{x}\} > Pr\{C_{m_2}|\mathbf{x}\} \quad (13)$$

Therefore, selecting the decision maximizing the *a posteriori* probability or selecting the decision maximizing the *a posteriori* possibility decision is identical: using the MPC

associated with the binary cost assignment rule or using the maximum possibility decision rule led to an identical result as expressed in (14).

This decision-making criterion is called the *Naive Bayes style* possibilistic criterion

$$Decision [x(p)] = C_{m_0} \text{ iff } m_0 = \arg \max_{m=1, \dots, M} Pr\{C_m|x\} = \max_{m=1, \dots, M} \pi_x(C_m) \quad (14)$$

Refs. [40–42] and most ongoing efforts are oriented into the computation of the *a posteriori* possibility values using numerical data [43]. An extensive study of properties and equivalence between possibilistic and probability approaches is presented in [20]. Notice that this decision rule, strongly inspired from probabilistic decision reasoning, does not provide a hard decision mechanism when several elementary decisions have the same maximum possibility measure.

#### 4.2. Decision Rule Based on Maximizing the Necessity Measure

It is worthwhile to notice that the *a posteriori* measures of possibility  $\Pi_x$  and necessity  $N_x$  coming from a normal *a posteriori* possibility distribution  $\pi_x(\cdot)$ , constitute a bracketing for the *a posteriori* probability distribution  $Pr\{\cdot|x\}$  [17]:

$$N_x(\{C_m\}) = 1 - \max_{n \neq m} \pi_x(C_n) \leq Pr\{C_m|x\} \leq \Pi_x(\{C_m\}) = \pi_x(C_m) \quad (15)$$

Therefore, the maximum possibility decision criterion can be considered as an optimistic decision criterion as it maximizes the upper bound of the *a posteriori* probability distribution. On the contrary, a pessimistic decision criterion based on maximizing the *a posteriori* necessity measure can be considered as a maximization of the lower bound of the *a posteriori* probability distribution. Equation (16) expresses this pessimistic decision criterion:

$$Decision [x(p)] = C_{m_0} \text{ iff } m_0 = \arg \max_{n=1, \dots, M} [N_x(\{C_m\})] \quad (16)$$

The question that we must raise concerns the “links” between the optimistic and the pessimistic decision criteria. Let us consider the *a posteriori* possibility distribution  $\pi_x(\cdot)$  for which  $C_{m_1}$  (resp.  $C_{m_2}$ ) is the “winning decision” obtained using the maximum possibility (resp. necessity measure) decision criteria as given in (17):

$$\pi_x(C_{m_1}) = \max_m \pi_x(C_m) \text{ and } N_x(\{C_{m_2}\}) = \max_m N_x(\{C_m\}) \quad (17)$$

The following important question can be formulated as follows: “Is the winning decision  $C_{m_1}$  (according to the maximum possibility criterion) is the same as the winning decision  $C_{m_2}$  according to maximum necessity measure criterion?”

First, notice that if several elementary decisions share the same maximum possibility value  $v = \pi_x(C_{m_1})$ , then, the necessity measure becomes a useless decision criterion since:

$$N_x(\{C_m\}) = 1 - \max_{k \neq m} \pi_x(C_k) = 1 - v \text{ for all the elementary decisions.}$$

Now, suppose that only one decision  $C_{m_1}$  assumes the maximum possibility value  $v = \pi_x(C_{m_1})$ , it is important to raise the question whether the decision  $C_{m_1}$  will (or will not) be the decision assuming the maximum necessity measure value. Let us note  $v'$ , the possibility value for the “second best” decision according to the possibility value criterion. As  $C_{m_1}$  is the unique decision having the maximum possibility value  $v$ , we have  $v' < v$ . Therefore, as shown in (18), the necessity measure value  $N_x(\{C_m\})$  only gets maximum for the decision  $C_{m_1}$  since  $1 - v' > 1 - v$ .

$$N_x(\{C_m\}) = 1 - \max_{k \neq m} \pi_x(C_k) = \begin{cases} 1 - v' & \text{if } m = m_1 \\ 1 - v & \text{if } m \neq m_1 \end{cases} \quad (18)$$

As a conclusion, when the maximum necessity measure criterion is useful for application (i.e., only one elementary decision assumes the maximum possibility value), then, both decision criteria (maximum possibility and maximum necessity) produce the same winning decision. In order to illustrate the difference between the maximum possibility and the maximum necessity measure criteria, Figure 2 presents an illustrative example.

In Figure 2 example, four different *a posteriori* possibility distributions  $\pi_1, \pi_2, \pi_3, \pi_4$ , all defined on a five elementary decisions set  $\Omega = \{C_1, C_2, C_3, C_4, C_5\}$  are considered. The necessity measures  $N_k(\{C_m\})$  have been computed from the corresponding possibility distribution  $\pi_k$ . The underlined values indicate which decisions result from the maximum possibility decision criterion as well as the maximum necessity measure decision criterion, for the four possibility distributions  $\pi_k$ . Note that the necessity measure assumes at most two values whatever the considered possibility distribution. When the *a posteriori* possibility distribution has one and only one decision having the highest possibility degree, then both decision rules produce the same winning decision. This is the case of the normal possibility distribution  $\pi_1$  as well as the subnormal possibility distribution  $\pi_3$ , indicated as cases (a) and (c) in Figure 2.

When several elementary decisions share the same highest possibility degree, then the maximum possibility decision criterion can randomly select one of these potential winning decisions. In this case, the maximum necessity measure decision criterion will affect a single necessity measure degree to all elementary decisions from  $\Omega$ , and thus, it will be impossible to select any of the potential winning decisions. This behavior can be observed with a normal possibility distribution  $\pi_2$  as well as with a subnormal possibility distribution (like  $\pi_4$ ), cases (b) and (d) in Figure 2. This example clearly shows the weakness of the decisional capacity of the maximum necessity measure decision criterion when compared to the maximum possibility decision criterion.

	$\pi_1(C_m)$	$N_1(\{C_m\})$	$\pi_2(C_m)$	$N_2(\{C_m\})$	$\pi_3(C_m)$	$N_3(\{C_m\})$	$\pi_4(C_m)$	$N_4(\{C_m\})$
$C_1$	0.1	0	0.1	0	0.1	0.2	0.1	0.2
$C_2$	0.5	0	0.5	0	0.5	0.2	0.5	0.2
$C_3$	0.3	0	0.3	0	0.3	0.2	0.3	0.2
$C_4$	0.8	0	<u>1</u>	0	<u>0.8</u>	<u>0.5</u>	<u>0.8</u>	0.2
$C_5$	<u>1</u>	<u>0.2</u>	<u>1</u>	0	0	0.2	<u>0.8</u>	0.2

(a)
(b)
(c)
(d)

**Figure 2.** Comparative example of the maximum possibility and maximum necessity measures decision criteria using four *a posteriori* possibility distributions.

#### 4.3. Decision Rule Based on Maximizing the Confidence Index

Other possibilistic decision rules based on the use of uncertainty measures are also encountered in literature. The most frequently used criterion (proposed by Kikuchi et al. [44]) is based on the maximization of the confidence index  $Ind$  defined as a combination of the possibility and the necessity measures for each event  $A \subseteq \Omega$ , given a possibility distribution  $\pi(\cdot)$ :

$$Ind: 2^\Omega \rightarrow [-1, +1]$$

$$A \rightarrow Ind(A) = \Pi(A) + N(A) - 1, \forall A \subseteq \Omega \quad (19)$$



where  $2^\Omega$  denotes the power set of  $\Omega$ , i.e., the set of all subsets from  $\Omega$ .

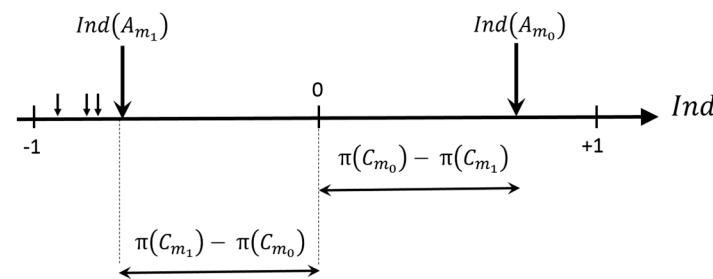
For an event  $A$ , this index ranges from  $-1$  to  $+1$ :

- $Ind(A) = -1$ , iff  $\Pi(A) = N(A) = 0$  (the occurrence of  $A$  is totally impossible and uncertain);
- $Ind(A) = +1$ , iff  $\Pi(A) = N(A) = 1$  (the occurrence of  $A$  is totally possible and certain).

Restricting the application of this measure to events  $A_m$  having only one decision  $A_m = \{C_m\}$  shows that  $Ind(A_m)$  measures the difference between the possibility measure of the event  $A_m$  (which is identical to the possibility degree of the decision  $C_m$ ) and the highest possibility degree of all decisions contained in  $(A_m)^c$  (the complement of  $A_m$  in  $\Omega$ ):

$$Ind(A_m) = \Pi(A_m) + N(A_m) - 1 = \pi(C_m) - \max_{m \neq n} \pi(C_n) \quad (20)$$

Therefore, if  $A_{m_0} = \{C_{m_0}\}$  is the only event having the highest possibility measure value  $\pi(C_{m_0})$ , then,  $A_{m_0}$  will be the unique event having a positive confidence index value, whereas all other events will have negative values, as illustrated in Figure 3 where we assume  $\pi(C_{m_0}) > \pi(C_m), \forall m \neq m_0$ , and,  $C_{m_1}$  refers to the decision having the second highest possibility degree.



**Figure 3.** Confidence indices associated with different decisions ( $A_{m_0}$ : event having the highest possibility degree,  $A_{m_1}$ : event with the second highest possibility degree). (Source: [11]).

In a classification decision-making problem, the decision criterion associated with this index can be formulated as follows:

$$Decision_{Ind}[x(p)] = A_{m_0} \text{ iff } Ind(A_{m_0}) = \max_{m=1, \dots, M} [Ind(A_m)] \quad (21)$$

The main difference between the maximum possibility and the maximum confidence index decision criteria lies in the fact that the maximum possibility decision criterion is only based on the maximum possibility degree whereas the maximum confidence index decision criterion is based on the difference between the two highest possibility degrees associated with the elementary decisions. As already mentioned, it is important to notice that the event  $A_{m_0} = \{C_{m_0}\}$  having the highest possibilistic value, will be the unique event producing a positive confidence index measuring the difference with the second highest possibility degree. All other events  $A_m = \{C_m\}, \forall m \neq m_0$ , will produce negative confidence indices.

When several decisions share the same highest possibility degree, their confidence index (the highest one) will be null. This shows the real capacity of this uncertainty measure for the decision-making process. However, this criterion brings the same resulting decisions as the two former ones.

## 5. Possibilistic Maximum Likelihood (PML) Decision Criterion

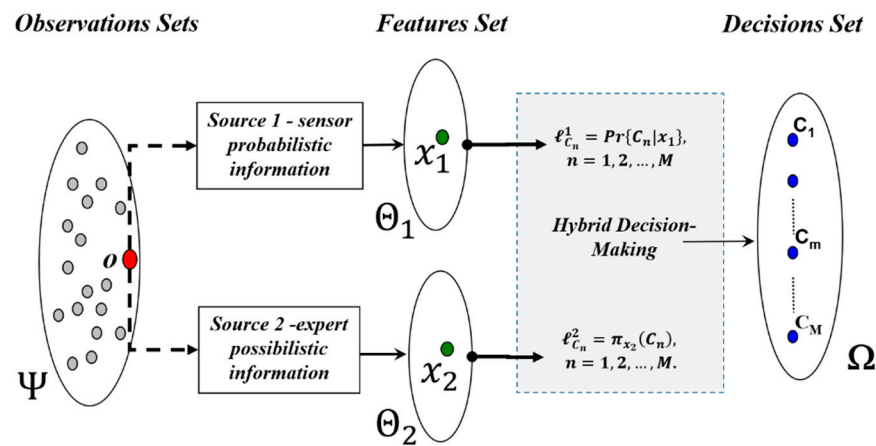
In the formulation of the Bayesian classification approach, all information sources are assumed to have probabilistic uncertainty where the available knowledge describing this uncertainty is expressed, estimated or evaluated in terms of probability distributions. In the

possibilistic classification framework, the information sources are assumed to suffer from possibilistic (or epistemic) uncertainty where the available knowledge describing this uncertainty is expressed in terms of possibility distributions. In this section, the Bayesian pattern recognition framework is generalized in order to integrate both probabilistic and epistemic sources of knowledge. A joint probabilistic—possibilistic decision criterion called *Possibilistic Maximum Likelihood* (PML) is proposed to handle both types of uncertainties.

### 5.1. Sources with Probabilistic and Possibilistic Types of Uncertainties

In some situations, an object from the observation space is observed through several feature sets. This is the case, for instance, in multi-sensor environment for classification applications. In such situations, the information available for the description of the feature vectors may be of different natures: probabilistic, epistemic, etc. Yager [24,45,46] addresses the same sort of problems: multi-source uncertain information fusion in the case when the information can be both from hard sensors of a probabilistic type and from soft knowledge-expert linguistic source of a possibilistic type. He uses t-norms (“and” operations) to combine possibility and probability measures. As will be explained below, Yager’s product of possibilities and probabilities coincides with our ‘decision variables’ optimized through the proposed PML approach.

Let us consider the example illustrated in Figure 4, where each pattern  $o$  (from the patterns set  $\Psi$ ) is “observed” through two channels. Source 1 (resp. source 2) measures a sub-feature vector  $x_1 \in \Theta_1$  (resp.  $x_2 \in \Theta_2$ ). Therefore, the resulting feature vector  $x(o)$  is obtained as the concatenation of the two sub-feature vectors:  $x(o) = [x_1 \ x_2]$ . In this configuration, the available information in the sub-feature vector  $x_1$  (resp.  $x_2$ ) undergoes probabilistic (resp. epistemic) uncertainties and is encoded as an *a posteriori* probability soft decision label vector  $\ell_{c_n}^1 = \Pr\{C_n|x_1\}, n = 1, 2, \dots, M$  (resp. *a posteriori* possibility soft decision label vector  $\ell_{c_n}^2 = \pi_{x_2}(C_n), n = 1, 2, \dots, M$ ).



**Figure 4.** Multi-source information context for pattern classification.

As an example, in a remote sensing system, *Source 1* may be considered as a multi-spectral imaging system, where all potential *a posteriori* probability distributions,  $\Pr\{C_n|x_1\}, n = 1, 2, \dots, M$ , are assumed to be known and well established. The second sensor, *Source 2*, could be a radar imaging system where the available information concerning the different thematic classes is expressed by an expert using ambiguous linguistic variables like: “the thematic class  $C_n$  is observed as “Bright”, “Slightly Dark”, etc. in the sub-feature set  $\Theta_2$ ”. Each linguistic variable can be used to generate an *a posteriori* possibility distribution associated with each thematic class  $\pi_{x_2}(C_n), n = 1, 2, \dots, M$ .

### 5.2. Possibilistic Maximum Likelihood (PML) Decision Criterion: A New Hybrid Criterion

In the Bayesian decision framework, detailed in previous sections, the binary cost assignment approach suffers from two constraints. On one hand, all errors are considered as equally costly: the penalty (or cost) of misclassifying an observed pattern  $\mathbf{o}$  as being associated with a decision  $C_n$  whereas the true decision for “ $\mathbf{o}$ ” is  $C_m$  is the same (unit loss). This situation does not reflect real applications constraints. For instance, deciding that an examined patient is healthy whereas he suffers a cancer is much more serious than the other way around. On the other hand, the loss function values  $\lambda_{m,n}, \forall m, n \in \{1, 2, \dots, M\}$  are static (or, predefined) and do not depend on the feature vectors of the observed patterns. The possibilistic maximum likelihood (PML) criterion, proposed in this paper, is based on the use of the epistemic source of information (the *a posteriori* possibility distribution, defined on the sub-feature space  $\Theta_2$ ) in order to define possibilistic loss values and to inject, afterwards, these values into the Bayesian decision criterion.

Assume that, for each object  $\mathbf{o} \in \Psi$ , the observed feature vector is given by  $\mathbf{x}(\mathbf{o}) = [x_1 \ x_2] \in \Theta_1 \times \Theta_2$ , and denote  $Pr\{\cdot | x_1\}$  (resp.  $\pi_{x_2}(\cdot)$ ) as encoding the *a posteriori* probability (resp. possibility) soft decision label vectors defined over the sub-feature set  $\Theta_1$  (resp.  $\Theta_2$ ). The proposed PML criterion relies on the use of loss values  $\lambda_{m,n}$  ranging from -1 (i.e., no loss) to +1 (i.e., maximum loss), and  $\lambda_{m,n}$  refers to the risk of choosing  $C_n$  whereas the real decision for the considered pattern is  $C_m$ . Depending on the epistemic information available through Source 2, the proposed loss values are given by:

$$\lambda_{m,n} = \begin{cases} \max_{k \neq m} \pi_{x_2}(C_k) & = \Pi_{x_2}(\{C_m\}^c) \quad \forall m \neq n \\ \max_{k \neq m} \pi_{x_2}(C_k) - \pi_{x_2}(C_m) & = -Ind_{x_2}(\{C_m\}) \quad \text{if } m = n \end{cases} \quad (22)$$

In the case of a wrong decision, the decision penalty values, i.e.,  $\lambda_{m,n}$  where  $m \neq n$ , are considered as positive loss values ranging in the interval:  $0 \leq \lambda_{m,n} = \max_{k \neq m} \pi_{x_2}(C_k) \leq 1$ . Thus, the wrong decision unit cost in the framework of binary-cost assignment, is “softened”, in this possibilistic approach, and assumes its maximum value, i.e., unit cost, only when the wrong decision  $C_n$  has a total possibility degree of occurrence.

When a correct decision  $C_m$  is selected, the zero-loss value (used by the binary cost assignment approach) is substituted by  $\lambda_{m,n} = \max_{k \neq m} \pi_{x_2}(C_k) - \pi_{x_2}(C_m) = -Ind_{x_2}(\{C_m\})$ . If the occurrence possibility degree  $\pi_{x_2}(C_m)$  of the true decision  $C_m$ , is the highest degree  $\pi_{x_2}(C_m) > \max_k \pi_{x_2}(C_k)$ , then the resulting loss value  $\lambda_{m,n}$  becomes negative. The smallest penalty value is reached, i.e.,  $\lambda_{m,n} = -1$ , when  $\pi_{x_2}(C_m) = 1$  (i.e., true decision  $C_m$  has a total possibility degree of occurrence), with a null possibility degree of occurrence for all the remaining decisions (leading to  $\max_{m \neq k} \pi_{x_2}(C_k) = 0$ ). Two special cases are present:

- (1) If the true decision  $C_m$  shares the same maximum possibility value with, at least one different wrong decision  $C_m$ , then, the correct decision  $C_m$  loss value becomes null  $\lambda_{m,n} = \max_{k \neq m} \pi_{x_2}(C_k) - \pi_{x_2}(C_m) = 0$ ;
- (2) If the true decision  $C_m$  does not produce the maximum occurrence possibility degree, i.e.,  $\pi_{x_2}(C_m) < \max_{k \neq m} \pi_{x_2}(C_k)$ , then the loss value  $\lambda_{m,n}$  is positive and will increase the conditional risk, associated with the true decision  $C_m$ .

Using the proposed possibilistic loss values, the conditional risk  $R(C_k | x_1)$  of choosing decision  $C_k$  can thus be computed as follows:

$$R(C_k | x) = -Ind_{x_2}(\{C_k\}) \cdot Pr\{C_k | x_1\} + \sum_{\substack{i=1 \\ i \neq k}}^M \Pi(\{C_m\}^c) \cdot Pr\{C_i | x_1\} \quad (23)$$

As already mentioned, Bayes decision criterion computes the conditional risk for all decisions, then, selects the decision  $C_n$  for which  $R(C_n | x)$  is minimum. Based on Equation (23), and to select the minimum conditional risk decision, the comparison of conditional risks related to two decisions  $C_k$  and  $C_p$ , can be straightforward performed leading to:

$$R(C_k|x_1) \leq R(C_p|x_1) \Leftrightarrow \pi_{x_2}(C_k) \cdot Pr\{C_k|x_1\} \geq \pi_{x_2}(C_p) \cdot Pr\{C_p|x_1\} \quad (24)$$

Therefore, the application of the PML criterion, for the selection of the minimum conditional risk decision (out of  $M$  potential elementary decisions) can be simply formulated by the following decision rule:

$$Decision[x(p)] = \arg \max_{n=1, \dots, M} \pi_{x_2}(C_n) \cdot Pr\{C_n|x_1\} \quad (25)$$

This “intuitive” decision criterion allows the joint use both probabilistic and epistemic sources of information in the very same Bayesian minimum risk framework. As an example, the application of the proposed possibilistic loss values in the two-class decision case, where  $\Omega = \{C_1, C_2\}$ , leads to the following loss matrix  $[\lambda]$ :

$$[\lambda] = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} \\ \lambda_{2,1} & \lambda_{2,2} \end{bmatrix} = \begin{bmatrix} \pi_{x_2}(C_2) - \pi_{x_2}(C_1) & \pi_{x_2}(C_2) \\ \pi_{x_2}(C_1) & \pi_{x_2}(C_1) - \pi_{x_2}(C_2) \end{bmatrix} \quad (26)$$

The use of this loss matrix  $[\lambda]$  into the minimum-risk Bayes decision approach (as defined in (5)), leads to express the PML decision as follows:

$$\frac{Pr\{x_1|C_1\}}{Pr\{x_1|C_2\}} \underset{Decision[x(p)]=C_2}{\overset{Decision[x(p)]=C_1}{\gtrless}} \frac{\pi_{x_2}(C_2)}{\pi_{x_2}(C_1)} \cdot \frac{Pr\{C_2\}}{Pr\{C_1\}} \quad (27)$$

Notice that when the proposed possibilistic loss values are considered, then the PML induces a “weighting adjustment” of the *a priori* probabilities where the weighting factors are simply the *a posteriori* possibility degrees issued from the possibilistic Source 2. In the case of equal *a priori* probabilities,  $Pr\{C_2\} = Pr\{C_1\}$ , this decision criterion turns to an intuitive form using jointly probabilistic and epistemic sources of information, in the Bayesian minimum risk framework as shown by:

$$\pi_{x_2}(C_1) \cdot Pr\{x_1|C_1\} \underset{Decision[x(p)]=C_2}{\overset{Decision[x(p)]=C_1}{\gtrless}} \pi_{x_2}(C_2) \cdot Pr\{x_1|C_2\} \quad (28)$$

It is worthwhile to notice that when the two following conditions prevail:

- when the available probabilistic information (issued from source 1) is non-informative; and,
- when the only meaningful and available information is reduced to the epistemic expert information on the sub-feature vector issued from source 2; then,

the proposed PML criterion is simply reduced to the maximum possibility decision criterion:

$$\pi_{x_2}(C_1) \underset{Decision[x(p)]=C_2}{\overset{Decision[x(p)]=C_1}{\gtrless}} \pi_{x_2}(C_2) \quad (29)$$

This raises a fundamental interpretation of the maximum possibility decision criterion as being a very special case of the possibilistic Bayesian decision making process under the total ignorance assumption of the probabilistic source of information.

### 5.3. PML Decision Criterion Behavior

Let  $S_1$  denotes a probabilistic source of information measuring a sub-feature vector  $\mathbf{x}_1 \in \Theta_1$  and attributing to each elementary decision  $C_m, m = 1, 2, \dots, M$ , an *a posteriori* probability soft decision label  $Pr\{C_m|\mathbf{x}_1\}$ . Under the assumption of equal *a priori* probabilities and using the binary-cost assignment, the application of the maximum *a posteriori* criterion (MAP), Equation (8), turns to be the “optimal” criterion ensuring the minimum-error decision rate.

Assume that an additional possibilistic source of information,  $S_2$ , (measuring a sub-feature vector  $\mathbf{x}_2 \in \Theta_2$ ) is available, see Figure 4. Based on the use of the sub-feature vector  $\mathbf{x}_2 \in \Theta_2$ ,  $S_2$  attributes to each elementary decision,  $C_m$ , an *a posteriori* possibility soft decision label  $\pi_{x_2}(C_m), m = 1, 2, \dots, M$ . To obtain a hard decision, the application of the *maximum of possibility* decision criterion, Equation (12), is considered.

In the previous section, we have proposed the possibilistic maximum likelihood, PML, decision criterion, Equation (25), as a hybrid decision criterion allowing the coupled use of both sources of information,  $S_1$  and  $S_2$ , by considering the possibilistic information issued from  $S_2$ , i.e.,  $\pi_{x_2}(C_m), m = 1, 2, \dots, M$ , for the definition of the loss values in the framework of the minimum-risk Bayes decision criterion (instead of the use of the binary-cost assignment approach). In this section, we will briefly discuss, from a descriptive point of view through an illustrative example, the “decisional behavior” of the PML criterion when compared to the decisions obtained with the “individual” application of the MAP and the *maximum of possibility* decision criteria.

First, it is worthwhile to notice that the “decision variable” to be maximized by the PML criterion is simply the direct product  $v(C_m) = \pi_{x_2}(C_m) \cdot Pr\{C_m|\mathbf{x}_1\}, m = 1, 2, \dots, M$ , which is a T-norm fusion operator (considering both probabilistic and possibilistic information as two “similar” measures of the degree of truthfulness related to the occurrence of different elementary decisions, see also p.101 of Yager [24]). This also means that both sources of information,  $S_1$  and  $S_2$ , are considered as having the same informative level. It is also important to notice that the PML criterion, as a decision fusion operator merging decisional information from both sources,  $S_1$  and  $S_2$ , constitutes a coherent decision fusion criterion in the sense that:

- when both sources  $S_1$  and  $S_2$  are in full agreement (i.e., leading to the same decision  $C_{m_0}$ ), then, the decision obtained by the application of the PML criterion will be the same as  $C_{m_0}$ ;
- when one of the two sources  $S_1$  and  $S_2$  suffers from total ignorance (i.e., producing equal *a posteriori* probabilities, for  $S_1$  and equal *a posteriori* possibilities, for  $S_2$ ), then the PML criterion will “duplicate” the same elementary decision as the one proposed by the remaining reliable source of information;
- when the two sources  $S_1$  and  $S_2$  lack decisional agreement, then, the decision obtained by the application of the PML criterion will be the most “plausible” elementary decision that may be different from individual decisions resulting from the MAP (resp. maximum possibility) criterion using the sub-feature vector  $\mathbf{x}_1 \in \Theta_1$  (resp. sub-feature vector  $\mathbf{x}_2 \in \Theta_2$ ).

This decision fusion coherence is illustrated through the examples given in Table 1. The decisions set is formed by five elementary decisions, i.e.,  $\Omega = \{C_1, C_2, C_3, C_4, C_5\}$ , and we assume that, given the observed feature  $\mathbf{x}_1 \in \Theta_1$ , the probabilistic source  $S_1$  produces the following *a posteriori* probability distribution:  $Pr\{\cdot|\mathbf{x}_1\} = [0.1 \ 0.4 \ 0.1 \ 0.3 \ 0.1]$ . Each example fits in one sub-array which presents  $S_1$  and  $S_2$  specific configuration ( $Pr\{\cdot|\mathbf{x}_1\}$ ,  $\pi_{x_2}(\cdot)$  and  $v = Pr\{\cdot|\mathbf{x}_1\} \cdot \pi_{x_2}(\cdot)$ ) with the resulting decision for each decision parameter. The cases presented in Table 1 are explained as the following:

*Case 1:* when both sources  $S_1$  and  $S_2$  agree, with a winning decision  $C_2$ , the PML criterion maintains this agreement and obtains the same decision,  $C_2$ .

Cases 2 and 3: it shows that when one of the two sources presents a total ignorance, then the PML criterion “duplicates” the same elementary decision as the one offered by the remaining reliable source of information.

Cases 4, 5 and 6: when sources  $S_1$  and  $S_2$  lack agreement (i.e., dissonant sources), then, the resulting decision obtained through the application of the PML criterion is the most reasonable decision. That may not necessarily be one of the winning decisions offered by the two sources (this is specifically shown in case 6).

**Table 1.** PML decision making behavior for several cases.

Case 1				Case 2			Case 3		
	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$
$C_1$	0.1	0.2	0.02	0.1	1.0	0.1	0.2	0.8	0.16
$C_2$	0.4	0.7	0.28	0.4	1.0	0.4	0.2	0.2	0.04
$C_3$	0.1	0.3	0.03	0.1	1.0	0.1	0.2	0.1	0.02
$C_4$	0.3	0.7	0.21	0.3	1.0	0.3	0.2	0.1	0.02
$C_5$	0.1	0.1	0.10	0.1	1.0	0.1	0.2	0.2	0.04
Case 4				Case 5			Case 6		
	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$	$Pr\{\cdot   x_1\}$	$\pi_{x_2}(\cdot)$	$v(\cdot)$
$C_1$	0.1	1.0	0.1	0.1	0.8	0.08	0.1	0.8	0.08
$C_2$	0.4	0.2	0.08	0.4	0.2	0.08	0.4	0.3	0.12
$C_3$	0.1	0.0	0.0	0.1	0.1	0.01	0.1	0.4	0.04
$C_4$	0.3	0.0	0.0	0.3	0.3	0.03	0.35	0.5	0.16
$C_5$	0.1	0.0	0.0	0.1	0.1	0.02	0.1	0.5	0.05

## 6. Experimental and Validation Results

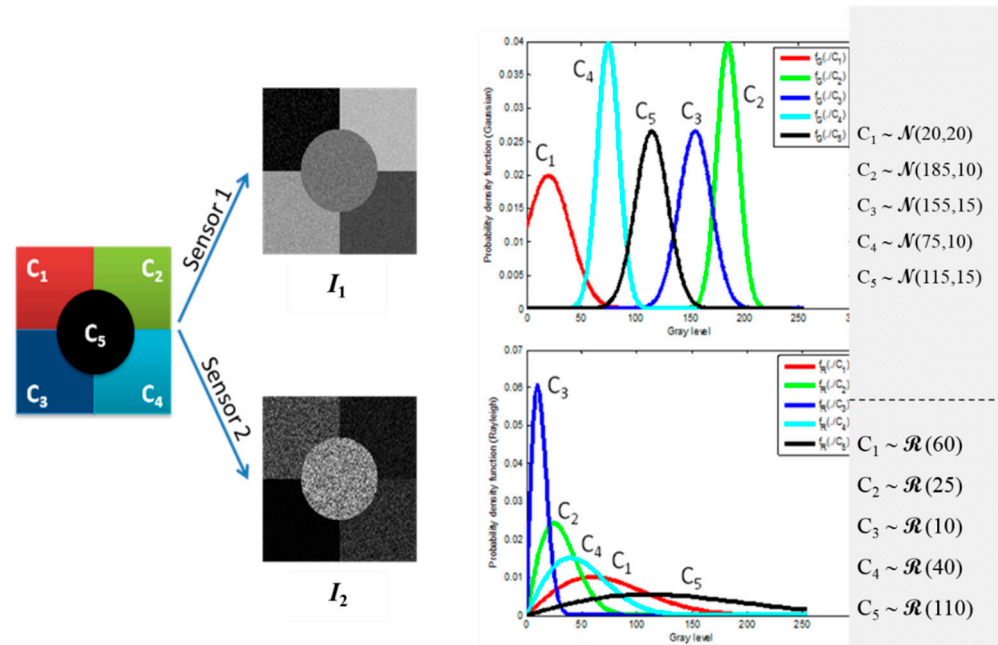
In this section, the proposed PML decision-making criterion is evaluated in a pixel-based image classification context. A synthetic scene composed of five thematic classes  $\Omega = \{C_1, C_2, C_3, C_4, C_5\}$  is assumed to be observed through two independent imaging sensors. Sensor  $S_1$  (resp. sensor  $S_2$ ) provides an image  $I_1$  (resp.  $I_2$ ) of the simulated scene. The two considered sensors are assumed to be statistically independent. Without loss of generalization, pixels from both images  $I_1$  and  $I_2$  are assumed to have the same spatial resolution, thus, they represent the same observed spatial cell or *object o*. The value of the pixel  $I_1(i, j)$  (resp.  $I_2(i, j)$ ) provides the observed feature  $x_1$  (resp.  $x_2$ ) delivered by the first (resp. second) sensor. According to sensors characteristics, the measured feature  $x_1$  (resp.  $x_2$ ) follows a Gaussian  $\mathcal{N}(m_C, \sigma_C^2)$  (resp. Rayleigh  $\mathcal{R}(\sigma_C^2)$ ) probability distribution with related parameters  $m_C, \sigma_C^2$  depending on the thematic class “C” of the observed object.

Figure 5 depicts the experimental simulated images  $I_1$  (resp.  $I_2$ ) assumed to be delivered at the output of the two sensors. Figure 6 shows the possibility distributions encoding expert’s information, for the five thematic classes. Parameter values considered for each thematic class are given in the same figure. This configuration of classes’ parameters is considered as a reasonable configuration that may be encountered when real data is observed. Nevertheless, other configurations have been generated and the obtained results are in full accordance with those obtained by the considered configuration.

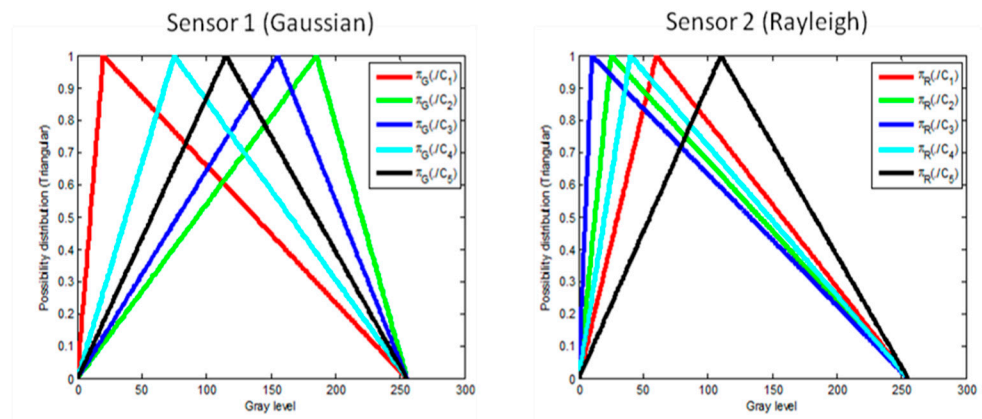
In addition to the previously mentioned probabilistic information, we assume that each thematic class is described, by an expert, using the “simplest” linguistic variable “Close to  $v_{S, C_k}$ ” where  $v_{S, C_k}$  denotes the thematic class  $C_k$  feature mean value, observed through sensor  $S_s$ . Therefore, the only information given by the expert is  $v_{1, C_k} = m_{C_k}$  for sensor  $S_1$  (underlying Gaussian distributions) and  $v_{2, C_k} = \sigma_{C_k} \sqrt{\pi/2}$  for sensor  $S_2$  (underlying Rayleigh distributions). For each sensor  $S_s$  and thematic class  $C_k$ , a standard triangular possibility distribution is considered to encode this epistemic knowledge with the summit positioned at the mean value  $v_{S, C_k}$  and the support covering the whole range

of the features set. It is clearly seen that the possibility distributions (considered as encoding the expert's knowledge), represent a weak knowledge which is *less* informative than the initial, or even estimated, probabilistic density functions.

To evaluate the efficiency of the proposed possibilistic maximum likelihood decision making criterion, the adopted procedure consists, first, on the random generation of 1000 statistical realizations of the two synthetic Gaussian and Rayleigh images (with the five considered thematic classes) representing the analyzed scene. Second, the following average pixel-based recognition rates are evaluated:



**Figure 5.** Two-sensors simulated images representing a scene of five thematic classes. Pixels from  $I_1$  (resp.  $I_2$ ) are generated using Gaussian (resp. Rayleigh) density functions.



**Figure 6.** Triangular-shaped possibility distributions encoding expert's knowledge, for the five thematic classes and the two sensors.

- $\tau(Pr_{C_m}^G)$ : Minimum-risk Bayes average pixel-based decision recognition rate, Equation (8), using zero-one loss assignment, for each thematic class  $C_m, m = 1, 2, \dots, 5$ , based on the use of sensor  $S_1$  Gaussian feature  $x_1$  only.
- $\tau(\pi_{C_m}^G)$ : Maximum possibility average pixel-based decision recognition rate, Equation (12), exploiting the epistemic expert knowledge for the description of each considered thematic class  $C_m, m = 1, 2, \dots, 5$ , in the features set  $\Theta_1$  only.

- $\tau(Pr_{C_m}^G \cdot \pi_{C_m}^R)$ : Possibilistic maximum likelihood average pixel-based decision recognition rate, Equation (24), jointly exploiting the epistemic expert knowledge for the description of each considered thematic class  $C_m, m = 1, 2, \dots, 5$ , in the features set  $\Theta_2$  (sensor  $S_2$ ), and the Gaussian probabilistic knowledge for the description of the same thematic class in the features set  $\Theta_1$  (sensor  $S_1$ ).
- $\tau(Pr_{C_m}^R)$ : Minimum-risk Bayes average pixel-based decision recognition rate, Equation (8), using zero-one loss assignment, for each thematic class  $C_m, m = 1, 2, \dots, 5$ , based on the use of sensor  $S_2$  Rayleigh feature  $x_2$  only.
- $\tau(\pi_{C_m}^G)$ : Maximum possibility average pixel-based decision recognition rate, Equation (12), exploiting the epistemic expert knowledge for the description of each considered thematic class  $C_m, m = 1, 2, \dots, 5$ , in the features set  $\Theta_1$  (sensor  $S_1$ ).
- $\tau(Pr_{C_m}^R \cdot \pi_{C_m}^G)$ : Possibilistic maximum likelihood average pixel-based decision recognition rate, Equation (24), jointly exploiting the epistemic expert knowledge for the description of each thematic class  $C_m, m = 1, 2, \dots, 5$ , in sensor  $S_1$  features set  $\Theta_1$ , and the Rayleigh probabilistic knowledge for the description of  $C_m, m = 1, 2, \dots, 5$ , in sensor  $S_2$  features set  $\Theta_2$ .
- $\tau(Pr_{C_m}^G \cdot Pr_{C_m}^R)$ : Minimum-risk Bayes average pixel-based decision recognition rate, Equation (8), using zero-one loss assignment, for each thematic class  $C_m, m = 1, 2, \dots, 5$ , and based on the joint use of both sensors  $S_1$  (associated with Gaussian probabilistic knowledge in the features set  $\Theta_1$ ) and  $S_2$  (associated with Rayleigh probabilistic knowledge in the features set  $\Theta_2$ ). Sensors  $S_1$  and  $S_2$  are considered as being statistically independent. This criterion as well as all the criteria above have been calculated for the example in Table 2.

**Table 2.** PML decision average pixel-based recognition rates for the five thematic classes using various configurations of knowledge sources.

Knowledge Sources	Source S1: Probabilistic (G: Gaussian) Source S2: Epistemic (Expert)			Source S2: Probabilistic (R: Rayleigh) Source S1: Epistemic (Expert)			Both sources are probabilistic
	$S_1$	$S_2$	$S_1 \oplus_{PML} S_2$	$S_2$	$S_1$	$S_1 \oplus_{PML} S_2$	$S_1 \oplus S_2$
	$\tau(Pr_{C_m}^G)$	$\tau(\pi_{C_m}^G)$	$\tau(Pr_{C_m}^G \cdot \pi_{C_m}^R)$	$\tau(Pr_{C_m}^R)$	$\tau(\pi_{C_m}^G)$	$\tau(Pr_{C_m}^R \cdot \pi_{C_m}^G)$	$\tau(Pr_{C_m}^G \cdot Pr_{C_m}^R)$
Criterion	$\tau(Pr_{C_m}^G)$	$\tau(\pi_{C_m}^G)$	$\tau(Pr_{C_m}^G \cdot \pi_{C_m}^R)$	$\tau(Pr_{C_m}^R)$	$\tau(\pi_{C_m}^G)$	$\tau(Pr_{C_m}^R \cdot \pi_{C_m}^G)$	$\tau(Pr_{C_m}^G \cdot Pr_{C_m}^R)$
$C_1$	0.96	0.33	0.96	0.34	0.96	0.64	0.96
$C_2$	0.90	0.32	0.91	0.49	0.98	0.68	0.92
$C_3$	0.78	0.91	0.81	0.89	0.67	0.90	0.91
$C_4$	0.94	0.25	0.95	0.32	0.96	0.61	0.97
$C_5$	0.83	0.72	0.86	0.58	0.75	0.71	0.94
<b>Average Recognition Rate</b>	0.88	0.51	0.90	0.52	0.86	0.71	0.94

The obtained average recognition rates are summarized in Table 2 (last row). As expected, at the global scene level, the average recognition rates when a probabilistic information source is used (for modelling the observed features) are higher than those obtained by the use of epistemic knowledge (i.e.,  $\tau(Pr^G) \geq \tau(\pi^G)$ , and  $\tau(Pr^R) \geq \tau(\pi^R)$ ). Nevertheless, at the thematic classes' level, this property does not hold for some classes. This is mainly due to the fact that for "sharp classes" probability density functions, i.e., small variance, (for instance, thematic classes  $C_2$  and  $C_4$ ), the possibility distributions shape used to encode the expert knowledge (i.e., a wide-based triangular possibility shape) may bias each class influence, leading to a better recognition rate to the detriment of other neighboring classes (for instance, class  $C_3$ ). In this case, this leads to obtain  $\tau(\pi_{C_2}^G) > \tau(Pr_{C_2}^G)$  and  $\tau(\pi_{C_4}^G) > \tau(Pr_{C_4}^G)$ .

Poorer recognition performances of the maximum possibility decision criterion clearly come from the "weak epistemic knowledge" produced by the expert (indicating



just the mean values) compared to the “strong probabilistic” knowledge involved by full probability density functions (resulting from either *a priori* information or the densities estimation using some learning data). The most interesting and promising result can be witnessed in terms of recognition rate improvement when the epistemic knowledge is jointly used with the probabilistic one as proposed by the PML decision criterion. Indeed, Table 2 (columns 4 and 7, bold numbers) shows that for all classes  $C_m, m = 1, 2, \dots, 5$ , we have  $\tau(Pr_{C_m}^G \cdot \pi_{C_m}^R) \geq \tau(Pr_{C_m}^G)$  and  $\tau(Pr_{C_m}^R \cdot \pi_{C_m}^G) \geq \tau(Pr_{C_m}^R)$ .

It is worthwhile to notice (columns 4 and 7, bold numbers) that the level of performance improvement depends on the “informative” capacity of the “additional” knowledge source. For instance, embedding the Gaussian source of knowledge (in terms of epistemic knowledge form) into the decisional process based on the probabilistic Rayleigh source of knowledge, improves much more the performance level than the reverse (i.e., embedding epistemic Rayleigh source of knowledge into the decisional process based on the probabilistic Gaussian source of knowledge):  $\tau(Pr_{C_m}^G \cdot \pi_{C_m}^R) \approx \tau(Pr_{C_m}^G)$  whereas  $\tau(Pr_{C_m}^R \cdot \pi_{C_m}^G) > \tau(Pr_{C_m}^R)$ .

Finally, it is important to notice that the PML decision performances are lower-upper bounds delimited as follows:

$$\begin{aligned} \tau(Pr_{C_m}^G) &\leq \tau(Pr_{C_m}^G \cdot \pi_{C_m}^R) \leq \tau(Pr_{C_m}^G \cdot Pr_{C_m}^R) \\ \tau(Pr_{C_m}^R) &\leq \tau(Pr_{C_m}^R \cdot \pi_{C_m}^G) \leq \tau(Pr_{C_m}^R \cdot Pr_{C_m}^G) \end{aligned} \quad (30)$$

Given the fact that the two sources  $S_1$  and  $S_2$  are assumed to be statistically independent, then the joint probability distribution of the augmented feature vector  $\mathbf{x} = [\mathbf{x}_1 \ \mathbf{x}_2]$  is the direct product of marginal ones. This simply means that the upper bounds given in Equation (30) constitute the optimal recognition rate (obtained by considering both probabilistic sources of knowledge). Therefore, the PML criterion improves the performances of the use of a “single” probabilistic source of knowledge, and approaches for some thematic classes the optimal recognition rate upper bound (last column of Table 2).

## 7. Conclusions

In this paper, a new criterion for decision-making process in classification systems is proposed. After a brief recall of the Bayesian decision-making criteria, three major possibilistic decision making criteria, i.e., maximum possibility, maximum necessity measure and confidence index maximization, have been detailed. It was clearly shown that the three considered decision criteria lead to, at best, the maximum possibility decision criterion. However, the maximum possibility criterion has no physical justification. A new criterion called, possibilistic maximum likelihood (PML) framed within the possibility theory, but using notions from Bayesian decision-making, has been presented and its behavior evaluated. The main motivation being the development of such criterion is for multi-source information fusion where a pattern may be observed through several channels and where the available knowledge, concerning the observed features, may be of a probabilistic nature for some features, and of an epistemic nature for some others.

In this configuration, most encountered studies transform one of the two knowledge types into the other form, and then apply either the classical Bayesian or possibilistic decision-making criteria. In this paper, we have proposed a new approach called the Possibilistic maximum likelihood (PML) decision-making approach, where the Bayesian decision-making framework is adopted and where the epistemic knowledge is integrated into the decision-making process by defining possibilistic loss values instead of the usually used zero-one loss values.

A set of possibilistic loss values is proposed and evaluated in the context of pixel-based image classification where a synthetic scene, composed of several thematic classes, was randomly generated using two types of sensors: a Gaussian and a Rayleigh sensor. The evaluation of the proposed PML criterion has clearly shown the interest of the application of PML; where the obtained recognition rates approach the optimal rates (i.e.,

where all the available knowledge is expressed in terms of probabilistic knowledge). Moreover, the proposed PML decision criterion offers a physical interpretation of the maximum possibility decision criterion as a special case of the possibilistic Bayesian decision-making process when all the available probabilistic information indicates equal decisions probabilities.

**Author Contributions:** Conceptualization, B.S., D.G., S.A. and B.A.; Methodology, B.S., D.G., B.A., and É.B.; Software, S.A. and B.A.; Supervision, B.S. and É.B.; Writing—original draft, B.S.; Writing—review & editing, and É.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dubois, D. Uncertainty theories: A unified view. In *SIPTA School 08—UEE 08*; SIPTA: Montpellier, France, 2008.
2. Klir, G.J.; Wierman, M.J. *Uncertainty-Based Information: Elements of Generalized Information Theory*; Physica-Verlag HD: Heidelberg, Germany, 1999.
3. Denoeux, T. 40 years of Dempster-Shafer theory. *Int. J. Approx. Reason.* **2016**, *79*, 1–6.
4. Yager, R.R.; Liu, L. (Eds.) *Classic Works of the Dempster-Shafer Theory of Belief Functions*; Springer: Berlin, Germany, 2008; Volume 219.
5. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976; Volume 42.
6. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353.
7. Zadeh, L. Fuzzy Sets as the Basis for a Theory of Possibility. *Fuzzy Sets Syst.* **1978**, *1*, 3–28.
8. Zadeh, L.A. Generalized theory of uncertainty (GTU)—principal concepts and ideas. *Comput. Stat. Data Anal.* **2006**, *51*, 15–46.
9. Dubois, D.; Prade, H. The legacy of 50 years of fuzzy sets: A discussion. *Fuzzy Sets Syst.* **2015**, *281*, 21–31.
10. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
11. Solaiman, B.; Bossé, É. *Possibility Theory for the Design of Information Fusion Systems*; Springer: Berlin, Germany, 2019.
12. Frélicot, C. On unifying probabilistic/fuzzy and possibilistic rejection-based classifiers. In *Advances in Pattern Recognition*; Springer: Berlin, Germany, 1998; pp. 736–745.
13. Solaiman, B.; Bossé, E.; Pigeon, L.; Guériot, D.; Florea, M.C. A conceptual definition of a holonic processing framework to support the design of information fusion systems. *Inf. Fusion* **2015**, *21*, 85–99.
14. Tou, J.T.; Gonzalez, R.C. *Pattern Recognition Principles*; Addison-Wesley: Boston, MA, USA, 1974.
15. Dubois, D.; Prade, H. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*; Plenum Press: New York, NY, USA, 1988.
16. Dubois, D.J. *Fuzzy Sets and Systems: Theory and Applications*; Academic Press: Cambridge, MA, USA, 1980; Volume 144.
17. Dubois, D.; Prade, H. When upper probabilities are possibility measures. *Fuzzy Sets Syst.* **1992**, *49*, 65–74.
18. Aliev, R.; Pedrycz, W.; Fazlollahi, B.; Huseynov, O.H.; Alizadeh, A.; Guirimov, B. Fuzzy logic-based generalized decision theory with imperfect information. *Inf. Sci.* **2012**, *189*, 18–42.
19. Buntao, N.; Kreinovich, V. How to Combine Probabilistic and Possibilistic (Expert) Knowledge: Uniqueness of Reconstruction in Yager's (Product) Approach. *Int. J. Innov. Manag. Inf. Prod. (IJIMIP)* **2011**, *2*, 1–8.
20. Coletti, G.; Petturiti, D.; Vantaggi, B. Possibilistic and probabilistic likelihood functions and their extensions: Common features and specific characteristics. *Fuzzy Sets Syst.* **2014**, *250*, 25–51.
21. Fargier, H.; Amor, N.B.; Guezguez, W. On the complexity of decision making in possibilistic decision trees. *arXiv* **2012**, arXiv:1202.3718.
22. Guo, P. Possibilistic Decision-Making Approaches. In *The 2007 International Conference on Intelligent Systems and Knowledge Engineering*; Atlantis Press: Beijing, China, 2007; pp. 684–688.
23. Weng, P. Qualitative decision making under possibilistic uncertainty: Toward more discriminating criteria. *arXiv* **2012**, arXiv:1207.1425.
24. Yager, R.R. A measure based approach to the fusion of possibilistic and probabilistic uncertainty. *Fuzzy Optim. Decis. Mak.* **2011**, *10*, 91–113.
25. Yager, R.R. On the fusion of possibilistic and probabilistic information in biometric decision-making. In *Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, Paris, France, 11–15 April 2011; IEEE: Piscataway Township, NJ, USA, 2011; pp. 109–114.

26. Bouyssou, D.; Dubois, D.; Prade, H.; Pirlot, M. *Decision Making Process: Concepts and Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
27. Dubois, D.; Foulloy, L.; Mauris, G.; Prade, H. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.* **2004**, *10*, 273–297.
28. Dubois, D.; Prade, H. Possibility theory: Qualitative and quantitative aspects. In *Quantified Representation of Uncertainty and Imprecision*; Springer: Berlin, Germany, 1998; pp. 169–226.
29. Dubois, D.; Prade, H. The three semantics of fuzzy sets. *Fuzzy Sets Syst.* **1997**, *90*, 141–150.
30. Dubois, D.; Prade, H.; Sandri, S. On possibility/probability transformations. In *Fuzzy logic*; Springer: Berlin, Germany, 1993; pp. 103–112.
31. Denœux, T. Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets Syst.* **2000**, *116*, 167–199.
32. Denœux, T. Maximum likelihood estimation from fuzzy data using the EM algorithm. *Fuzzy Sets Syst.* **2011**, *183*, 72–91.
33. Denœux, T. Maximum likelihood estimation from uncertain data in the belief function framework. *Knowl. Data Eng.* **2013**, *25*, 119–130.
34. Walley, P. *Statistical Reasoning With Imprecise Probabilities*; Chapman and Hall: London, UK, 1991.
35. Walley, P. Towards a unified theory of imprecise probability. *Int. J. Approx. Reason.* **2000**, *24*, 125–148.
36. Linda, O.; Manic, M.; Alves-Foss, J.; Vollmer, T. Towards resilient critical infrastructures: Application of Type-2 Fuzzy Logic in embedded network security cyber sensor. In Proceedings of the 2011 4th International Symposium on Resilient Control Systems (ISRCs), Boise, ID, USA, 9–11 August 2011; IEEE: Piscataway Township, NJ, USA, 2011; pp. 26–32.
37. Mendel, J.M.; John, R.I.B. Type-2 fuzzy sets made simple. *Fuzzy Syst.* **2002**, *10*, 117–127.
38. Ozen, T.; Garibaldi, J.M. Effect of type-2 fuzzy membership function shape on modelling variation in human decision making. Presented at the IEEE International Conference on Fuzzy Systems, Budapest, Hungary, 25–29 Junly 2004.
39. Luce, R.D.; Raiffa, H. *Games and Decisions*; Wiley: New York, NY, USA, 1957.
40. Haouari, B.; Amor, N.B.; Elouedi, Z.; Mellouli, K. Naïve possibilistic network classifiers. *Fuzzy Sets Syst.* **2009**, *160*, 3224–3238.
41. Benferhat, S.; Tabia, K. An efficient algorithm for naive possibilistic classifiers with uncertain inputs. In *Scalable Uncertainty Management*; Springer: Berlin, Germany, 2008; pp. 63–77.
42. Bounhas, M.; Hamed, M.G.; Prade, H.; Serrurier, M.; Mellouli, K. Naive possibilistic classifiers for imprecise or uncertain numerical data. *Fuzzy Sets Syst.* **2014**, *239*, 137–156.
43. Bounhas, M.; Mellouli, K.; Prade, H.; Serrurier, M. Possibilistic classifiers for numerical data. *Soft Comput.* **2013**, *17*, 733–751.
44. Kikuchi, S.; Perincherry, V. Handling Uncertainty in Large Scale Systems with Certainty and Integrity. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.7069> (accessed on 30 December 2020).
45. Yager, R.R. Set measure directed multi-source information fusion. *Fuzzy Syst.* **2011**, *19*, 1031–1039.
46. Yager, R.R. Hard and soft information fusion using measures. In Proceedings of the 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, Hangzhou, China, 15–16 November 2010.