The 39th Canadian Conference on Artificial Intelligence
DOI: 0

Canadian AI 2026
25—29 May 2026, Vancouver, BC

Back to the future: revival of evidence theory and modal logic for robust and interpretable AI

Elena Tsiporkova[†], Veselka Boeva[‡,*]
[†] EluciDATA Lab, Sirris, Ravensteinstraat 4, Brussels, Belgium
[‡] Department of Computer Science, Blekinge Institute of Technology, Sweden

Abstract

Years ago, when AI research was still predominantly a theoretical endeavor due to the scarcity of data, we developed a multi-valued modal logic interpretation of evidence measures. This work attracted some interest within the research community at the time but was gradually forgotten, even by us. Only recently have we revisited this modal logic interpretation of evidence theory, with the aim of developing neuro-symbolic modelling workflows capable of handling imperfect real-world data. The purpose of this position paper is to highlight the largely unexploited potential of these formalisms and to encourage the research community to further develop them toward more reliable, genuinely reasoning, and interpretable AI models. We begin by providing a concise overview of the key concepts underlying multi-valued mappings, evidence theory, and modal logic, along with our proposed multi-valued modal logic interpretation of evidence measures. We then present recent results demonstrating how these interpretations can be used to learn class expressions in weakly supervised learning scenarios. In addition, we show how these class representations can support reasoning under uncertainty in real-world applications. Finally, we discuss the untapped potential of the evidence-based approach for analyzing and quantifying the complexity of learning tasks, and we outline promising directions for future research.

Keywords: modal logic, evidence theory, neuro-symbolic AI, multi-valued mapping, evidential learning, class expression learning.

1. Rationale

Over the past decade, the rise of deep learning has largely displaced many traditional data science and machine learning paradigms from the research landscape. For a time, it seemed that these approaches were no longer necessary, as almost any task could be addressed with a sufficiently powerful deep learning architecture, until the limitations of the later became apparent. Consequently, it is interesting to observe how previously overlooked or somewhat neglected frameworks, such as ontologies, knowledge graphs, probabilistic and logical reasoning, and uncertainty modelling, are now being gradually revived, driven by the growing demand for robust, reliable, transparent, and explainable models. One such framework is Dempster–Shafer theory, commonly referred to as evidence theory. It provides measures of evidence, namely plausibility and belief, that capture both uncertainty and imprecision by defining upper and lower bounds on probabilities [1, 2]. More than a quarter of a century ago, we proposed, largely as a theoretical exercise, without specific real-world applications in mind, interpretations of Dempster–Shafer theory grounded in the syntactic and semantic structures of modal logic [3–5]. As is often the case, this work appears to have been ahead of its time as only now is its relevance and potential for modelling diverse application contexts in an elegant and transparent manner becoming fully apparent.

Recent research in neuro-symbolic AI has demonstrated that such approaches can be more effective and robust than purely deep learning (DL) methods, particularly in settings involving real-world, weakly labeled, and noisy data [6–10]. Building on this line of work, we have shown in [11] that integrating the modal logic interpretation of plausibility and

[*]veselka.boeva@bth.se
The authors contributed equally to the paper.

belief measures, as originally proposed in [3–5], with neural networks can lead to robust and interpretable AI models. As we further discuss in this paper, the resulting models exhibit increased robustness when dealing with imprecise scenarios and provide a logical framework for reasoning under uncertainty. Moreover, combining DL with modal logic–based reasoning enhances interpretability by enabling the transformation of available explicit knowledge into a rule-based representation.

The aim of this work is to raise awareness within the scientific community of the added value of such neuro-symbolic AI models, particularly those equipped with uncertainty reasoning capabilities grounded in modal logic interpretations of evidence theory. In what follows, we first briefly recall the theoretical foundations of multi-valued mappings, evidence theory, modal logic, and the corresponding interpretations of evidence measures. We then demonstrate the key properties and application potential of these interpretations in weakly supervised learning scenarios. Furthermore, we discuss how evidence theory can be leveraged to analyze and quantify the complexity of the learning task at hand. Finally, we outline several relevant directions for future research.

2. Evidence Theory Framed in Modal Logic

The core idea behind the inception of Dempster–Shafer theory [1, 2] was the introduction of the concepts of upper and lower probabilities (plausibility and belief measures), which are induced by a multi-valued mapping. We have explored this multi-valued approach in our original works [3–5] to successfully develop and establish the modal logic interpretation of Dempster–Shafer theory.

2.1. Some fundamentals of Demspter-Shafer theory

A *multi-valued mapping* $\mathcal{F}$ from X into Y associates to each element $x \in X$ a subset $\mathcal{F}(x)$ of Y [12, 13]. The *domain* of $\mathcal{F}$ is defined as $\text{dom}(\mathcal{F}) = \{x \mid x \in X \wedge \mathcal{F}(x) \neq \emptyset\}$. Consider a subset A of X and a subset B of Y. The *direct image* of A under $\mathcal{F}$ is the subset $\mathcal{F}(A)$ of Y, defined as $\mathcal{F}(A) = \bigcup_{x \in A} \mathcal{F}(x)$. Several inverse images of B under $\mathcal{F}$ can be defined. For instance, the *inverse* and *superinverse* images of B under $\mathcal{F}$ are the following subsets of X:

$$\begin{aligned} \mathcal{F}^-(B) &= \{x \mid x \in X \wedge \mathcal{F}(x) \cap B \neq \emptyset\} \\ \mathcal{F}^+(B) &= \{x \mid x \in \text{dom}(\mathcal{F}) \wedge \mathcal{F}(x) \subseteq B\}. \end{aligned} \tag{2.1}$$

Further, the *pure inverse image* of B under $\mathcal{F}$ is the subset $\mathcal{F}^{-1}(B)$ of X, defined as

$$\mathcal{F}^{-1}(B) = \{x \mid x \in X \wedge \mathcal{F}(x) = B\}.$$

A schematic illustration of the different inverse images can be consulted in Figure 1.

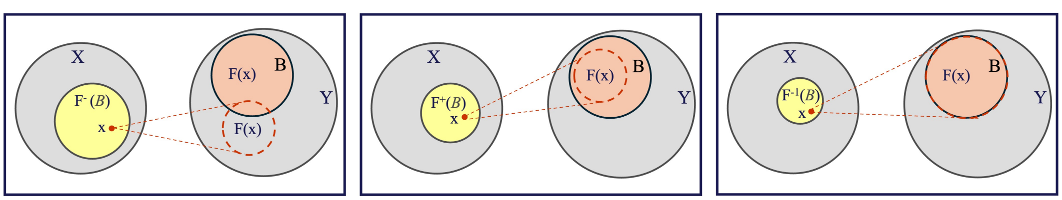

Figure 1. Schematic illustration of the inverse $\mathcal{F}^-(B)$ (left), the superinverse $\mathcal{F}^+(B)$ (middle) and the pure $\mathcal{F}^{-1}(B)$ (right) images of a set $B \subseteq Y$ under $\mathcal{F}$.

In his seminal work [1], Dempster showed that if P is a probability measure on $\mathcal{P}(X)$, then a multi-valued mapping $\mathcal{F}$ from a universe X into a universe Y (called a frame of discernment) induces *upper* and *lower probabilities* on $\mathcal{P}(Y)$, as follows:

$$P^*(B) = P\left(\mathcal{F}^-(B) \mid \text{dom}(\mathcal{F})\right) \quad \text{and} \quad P_*(B) = P\left(\mathcal{F}^+(B) \mid \text{dom}(\mathcal{F})\right). \tag{2.2}$$

It is clear that P^* and P_* are only well defined if $P(\mathrm{dom}(\mathcal{F})) > 0$. Note that P^* and P_* are dual, i.e., $P^*(B) = 1 - P_*(\mathrm{co}\, B)$.

Shafer reinterpreted upper and lower probabilities as degrees of *plausibility* Pl and *belief* Bel, abandoning Dempster's idea that they emerge as upper and lower bounds of Bayesian probabilities [2]. Furthermore, in case of a finite universe Y, Shafer introduced the concepts of a basic probability assignment and its focal elements. Formally, a $\mathcal{P}(Y) \to [0,1]$ mapping m is called a *basic probability assignment* (alternatively referred to as a *mass function*) on $\mathcal{P}(Y)$ if $m(\emptyset) = 0$ and $\sum_{B \in \mathcal{P}(Y)} m(B) = 1$. Note that the mass function m can also be defined in terms of the pure inverse image of the multi-valued mapping $\mathcal{F}$ as follows by [1]:

$$m(B) = P\left(\mathcal{F}^{-1}(B) \mid \mathrm{dom}(\mathcal{F})\right). \tag{2.3}$$

A subset B of Y for which $m(B) > 0$ is called a *focal element* of m. The union of all the focal elements of m is called its core. A mass function with all the focal elements concentrated on the singletons is referred to as a Bayesian mass function, i.e., a special case of a probability function. The belief Bel and plausibility Pl measures can be defined in terms of the mass function as follows:

$$\mathrm{Bel}(B) = \sum_{C \subseteq B} m(C) \quad \text{and} \quad \mathrm{Pl}(B) = \sum_{C \cap B \neq \emptyset} m(C).$$

2.2. **Modal logic interpretations of evidence measures**

Modal logic is an extension of classical propositional logic that is endowed with modal operators of possibility and necessity. Various systems of modal logic can be considered, depending on the additional axioms imposed. These systems can be clearly interpreted in terms of Kripke's semantics of *possible worlds* [14]. Thus necessary propositions are those that are true in all possible worlds, whereas possible propositions are those that are true in at least one possible world. The language of modal logic consists of a set of atomic propositions, logical connectives $\neg, \wedge, \vee, \to, \leftrightarrow$, and modal operators of *possibility* $\Diamond$ and *necessity* $\Box$. The propositions of the language can be the atomic propositions, and if p and q are propositions, then are so $\neg p$, $p \wedge q$, $p \vee q$, $p \to q$, $p \leftrightarrow q$, $\Box p$, $\Diamond p$.

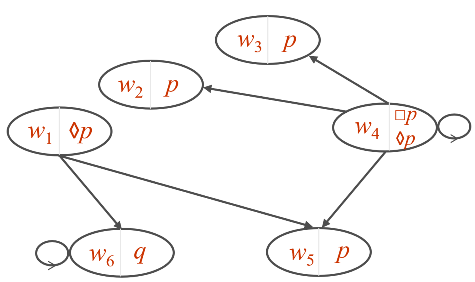

Figure 2. The truth set of a proposition p in a model $M = \langle W, R, V\rangle$, $\|p\|^M$, is a set of all worlds in which p is true. *Possibility* of a proposition p, $\Diamond p$, is true in a world w iff there is at least one world connected with w by the accessibility relation R in which p is true. *Necessity* of a proposition p, $\Box p$, is true in a world w iff p is true in all worlds connected with w by the accessibility relation R. In the above example, which depicts a model with six possible worlds, we have: $\|\Diamond p\|^M = \{w_1, w_4\}$ and $\|\Box p\|^M = \{w_4\}$.

Our interpretation of Dempster-Shafer theory, as published originally in [4, 5], is based on extending the concept of a *standard model* of modal logic. The latter is a triplet $M = \langle W, R, V\rangle$, where W denotes a set of possible worlds, R is a binary relation on W called *accessibility relation*, and V is the *value assignment function* by which truth T or falsity F of each atomic proposition in each world is assigned (see an illustrative example in Figure 2).

The model is extended by adding a probability measure on the power set $\mathcal{P}(W)$ of W, i.e., $M = \langle W, R, V, P \rangle$. Furthermore, it is considered that the propositions have the form $e_B =$"*a given incompletely characterized element* ϵ *is classified in some set* B", where $\epsilon \in Y$ and $B \in \mathcal{P}(Y)$. The set of atomic propositions is composed of all $e_{\{y\}}$, for $y \in Y$. It is additionally assumed that exactly one $e_{\{y\}}$ is true in each world. This implies that e_Y and also $e_B \leftrightarrow \neg e_{\mathrm{co}\,B}$ are always true in M and also induces a mapping f from W into Y associating with each world $w \in W$ the unique $y \in Y$ for which $e_{\{y\}}$ is true in w. Subsequently, regarding the accessibility relation R as a multivalued mapping from W into W, the composition $\mathcal{F}$ of R and f, i.e., $\mathcal{F} = f \circ R$, can be seen as a multivalued mapping from W into Y. Thus, given the probability measure P on $\mathcal{P}(W)$, $\mathcal{F}$ will induce upper and lower probabilities on $\mathcal{P}(Y)$ (see Section 2.1). In this context, it was shown in [4, 5] that a plausibility measure and a belief measure can be expressed in terms of conditional probabilities of truth sets of possibilities and necessities:

$$\mathrm{Pl}_M(B) = P\left(\|\Diamond e_B\|^M \mid \|\Diamond e_Y\|^M\right) \quad \text{and} \quad \mathrm{Bel}_M(B) = P\left(\|\Box e_B\|^M \mid \|\Diamond e_Y\|^M\right). \tag{2.4}$$

The basic probability assignment m_M corresponding to Pl_M and Bel_M is given by

$$m_M(B) = \begin{cases} 0 & \text{, if } B = \emptyset \\ P\left(\|\Box e_B \wedge q_B\|^M \mid \|\Diamond e_Y\|^M\right), & \text{otherwise} \end{cases}, \tag{2.5}$$

where

$$q_B = \bigwedge_{y \in B} \Diamond e_{\{y\}}.$$

3. Reasoning under Uncertainty in Real-World Applications

Many real-world application contexts, which can benefit from data-driven modeling, are concerned with a set of entities (objects), e.g., images, industrial machinery, vehicles, patients, described by a set of instances (properties). For example, image classification/annotation task requires that the concerned images are well described in terms of the segments or objects they contain. In case of failure prediction for industrial assets, each concerned machine, device, vehicle, is characterised in terms of its technical specifications, field usage, and/or operating conditions. Analogously in healthcare applications, each patient is supplied by a profile composed of demographic, physiological, clinical and other relevant information. Note that each entity (image, device, machine, ...) may have multiple properties, and a property may be associated with multiple entities.

3.1. Multi-instance Learning

Most of the above scenarios can be formulated as multi-instance learning (MIL) problems. These types of problems arise in complex machine learning applications where the learning system has partial or incomplete knowledge of each training example. For example, MIL has been widely used in image analysis for tasks such as scene classification and medical diagnosis, where detailed annotations are scarce. In traditional supervised learning problems, the learning system is given training examples in the form of $(y_i, \mathcal{L}_i)$, where $y_i \in Y$ is a training sample, $\mathcal{L}_i \in \mathcal{L}$ is its label, and $(Y, \mathcal{L})$ is the training set with $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_c\}$ representing a set of c classes. However, as applications become more complex, the learner may have incomplete information about the available training set. Instead of knowing that each training example can be represented by the same set of feature vectors, the learner only knows that each example can be represented by some subset of all potential feature vectors (instances). In an MIL scenario, each training example $y \in Y$ is associated with a subset of instances $X_y \subseteq X$, where X is the set of all instances used to describe the entities in the training dataset $(Y, \mathcal{L})$.

We have developed two distinct real-world MIL scenarios in [11, 15], demonstrating the application potential of the modal logic interpretations of evidence theory. These scenarios are briefly summarized in Section 3.4, where we also describe how the interpretations can be combined with neural network models to address the specific challenges associated with each use case.

3.2. **Possibility, Necessity and Ambiguity Conditions**

Let us consider the above MIL generic scenario with X denoting the set of all instances used to describe the entities (training samples) Y in a given training dataset $(Y, \mathcal{L})$. In order to be able to apply the modal logic interpretations of the evidence measures, as discussed in Section 2.2, it is necessary to transform/redefine each instance (property or characteristics) into a binary one that can be either satisfied or not by an entity. In this context, a multi-valued mapping $\mathcal{F}$ from the set of instances X to the set of entities Y can be defined. This mapping associates with each instance $x \in X$ a set of entities $\mathcal{F}(x) \subseteq Y$ that possess it. In addition, consider that the entities in Y are distributed (labeled) across c different categories (classes), i.e., $Y = \bigcup_{i=1}^{c} Y_i$, where $Y_i \subset Y$ and $Y_i \cap Y_j = \emptyset$, for $i \neq j$. Subsequently, the mapping $\mathcal{F}$ can be exploited to characterize each class Y_i, for $i = 1, \ldots, c$, in terms of its possibility and necessity conditions, by constructing inverse and superinverse images of the class as defined in (2.1). Formally, the possibility and the necessity conditions referring to class Y_i can be described by the following two expressions [11]:

$$\Diamond Y_i = \bigvee_{x \in \mathcal{F}^-(Y_i)} x \quad \text{and} \quad \Box Y_i = \bigvee_{x \in \mathcal{F}^+(Y_i)} x. \tag{3.1}$$

Intuitively, the possibility condition for class Y_i is defined as the disjunction of all single properties, for which at least one entity in their direct image under $\mathcal{F}$ satisfies the property. Analogously, the necessity condition for class Y_i is defined as the disjunction of all single properties, for which all entities in their direct image under $\mathcal{F}$ satisfy the property.

The above multi-valued interpretations of possibility and necessity conditions (3.1) facilitate reasoning about how well the existing evidence (instances) support the partitioning of Y in c classes and can be used to define logic rules for the purpose of predicting the most probable class of unseen entities. For instance, the fact that only the necessity condition of one of the disjoint classes can be fulfilled at any time, because of $\bigcap_{i=1}^{c} \mathcal{F}^+(Y_i) = \emptyset$, can be used for this purpose. Moreover, the common parts (the pairwise intersections) of the inverse images of the c classes in Y contain the instances in X, which are satisfied by entities belonging to at least two classes:

$$\mathcal{F}_c^- = \bigcup_{i=1}^{c} \bigcup_{j=i+1}^{c} \left(\mathcal{F}^-(Y_i) \cap \mathcal{F}^-(Y_j) \right),$$

capturing the so-called *ambiguity conditions* $\mathcal{A}_c$ associated with the partition $\{Y_i\}_{i=1}^{c}$:

$$\mathcal{A}_c = \bigvee_{x \in \mathcal{F}_c^-} x. \tag{3.2}$$

Note that in the binary case, i.e., $c = 2$, it can be shown that $\Diamond Y_i = \Box Y_i \vee \mathcal{A}_c$, for $i = \{1, 2\}$. In other words, any of the two classes is possible whenever the class is either necessary or the ambiguity conditions are fulfilled. In the latter case, both classes are possible, i.e., the situation is ambiguous.

3.3. **Evidence Scores and Ambiguity Reject**

According to the modal logic interpretations of evidence measures as expressed in (2.4) and the possibility and necessity conditions in (3.1), the plausibility (Pl) and belief (Bel) values of any new entity y to belong to class i ($i = 1, \ldots, c$) can be calculated as follows:

$$\mathrm{Pl}_i(y) = |\, \Diamond Y_i(X_y)\, | \,/\, |\, \Diamond Y_i\, | \quad \text{and} \quad \mathrm{Bel}_i(y) = |\, \Box Y_i(X_y)\, | \,/\, |\, \Diamond Y_i\, |, \tag{3.3}$$

where X_y is the subset of instances associated with entity y. The above plausibility and belief values of any entity $y \in Y$ form a set of intervals, $[\mathrm{Bel}_i(y),\, \mathrm{Pl}_i(y)]$. The width of these intervals is correlated with the uncertainty associated with y belonging to class i, for $i = 1, \ldots, c$. This uncertainty can be quantified by introducing a scoring function $\mathcal{E}_i(y)$, as proposed in [11] and referred here as an *evidence score*:

$$\mathcal{E}_i(y) = \frac{\mathrm{Pl}_i(y) + \mathrm{Bel}_i(y)}{\sum_{k=1}^{c} \big(\mathrm{Pl}_k(y) + \mathrm{Bel}_k(y)\big)}. \tag{3.4}$$

Note that the evidence scores are in the interval $[0, 1]$ and $\mathcal{E}_i(y)$ can be interpreted as the likelihood that an entity y belongs to class i. Thus, a score-based classifier can be defined by using equation (3.4) together with a pre-defined threshold to decide whether a new entity belongs to a class of interest. Such a decision rule can be further enriched with an ambiguity reject option. For instance, Chow [16] studied the optimal error-reject trade-off, proposing to postpone decision-making when the conditional error of making a decision is high. This scenario usually occurs in regions of the feature space where there is significant overlap between classes. In our case, the ambiguity between the classes is captured by the ambiguity conditions (see (3.2)) and can be quantified by introducing the concept of an *ambiguity index* as follows:

$$\mathcal{A}I_c(y) = |\, \mathcal{A}_c(X_y)\, | \,/\, |\, \Diamond Y(X_y)\, |\, .$$

The latter quantifies the portion of ambiguity conditions satisfied by the new entity y in comparison to all the conditions (instances) satisfied by y.

To handle ambiguous scenarios, Denoeux introduced in its evidential version of KNN [17], the most popular evidential machine learning model, the ambiguity reject option by imposing minimum thresholds on plausibility and belief. In later work [18], Denoeux considered two different situations involving entity rejection based on the approach in [19]. The first situation occurs when several classes appear to be almost equally likely, while the second one arises when the entity under consideration differs significantly from previous data. Let us discuss how these two situations can be handled within our approach. In our context, having several classes being almost equally likely will result into comparable evidence scores for those classes, most probably below the decision threshold due to the fact that $\sum_{i=1}^{c} \mathcal{E}_i(y) = 1$ for any y. In addition, it is expected that the ambiguity index will be rather high in situation when multiple classes are possible. Thus, low evidence score and high ambiguity index will induce a rejection. In the second situation discussed by Denoeux [18], i.e., when the new entity has not been seen in the training data, it may occur that low plausibility and belief values across all classes will boost the evidence scores, leading to false recognition. However, the ambiguity index is supposed to capture such situations by maintaining relatively high values. Thus, emergence of anomalous trends of high evidence scores and high ambiguity induces might point out to the occurrence of a concept drift and need for a model update.

3.4. **Neuro-Symbolic Symbiosis: Use Cases**

We have explored several strategies for integrating the above symbolic reasoning framework with deep learning. Such integration enables the combination of incomplete symbolic knowledge with the power of neural learning, thereby addressing ambiguity while improving both accuracy and interpretability of the results. More specifically, we have combined our multi-valued modal logic interpretations of evidence theory [3–5] with deep learning and investigated their potential for representing and modelling ambiguous knowledge in two distinct real-world use cases, both formulated as MIL problems. A brief summary of the proposed modelling approach, along with the obtained results in terms of possibility and necessity conditions (plausibility and belief measures) for each use case, is presented below.

3.4.1. Use Case 1. Detecting of abandoned luggage in surveillance videos

This use case scenario, studied in [11], is concerned with a surveillance scene classification task involving the detection of abandoned luggage. The training set consists of images (frames) labeled as either positive (abandoned luggage present) or negative (otherwise). In other words, a binary classifier discriminating between these two situations needs to be learnt. Scene classification problems typically require a two-stage approach involving an attribute extractor method and a classifier. Our neuro-symbolic workflow chains a pre-trained deep learning model for object detection with a rule-based classifier utilising the modal logic and evidence-based approach described in the foregoing subsections.

More concretely, a pre-trained deep learning model is used to detect objects of interest (in this case, people and bags) and estimate their depth within the image. Subsequently, each image is characterised with a set of meaningful attributes expressing the presence/absence of objects of interests (persons and bags) in the image, as well as, the relationships between the bag(s) of interest and the people detected in the image, e.g., the overlap between their bounding boxes and the distance between the bag and the person closest to it. Note that, it may occur that not each image can be associated with all the possible attributes, e.g., if neither a person nor a bag is detected in a image than attributes related to estimating the distance between them cannot be assigned to this image. Thus we are clearly dealing with an MIL context. Thus, a multi-valued mapping can be constructed from each attribute, also called instance in our context, to the subset of images in the training set in which the attribute appears. Subsequently, applying (3.1), the inverse and superinverse images of the positive and negative classes are used to express the possibility and necessity conditions of each of the two classes in terms of binary attributes.

For illustration, the necessity conditions of the positive class (abandoned bag) and the negative class (non-abandoned bag) for the PETS2006 dataset [20], which we used in our experiments in [11], can be expressed as follows:

$$\begin{aligned}\Box\{\mathit{abandoned}\} &= \mathit{min\_distance\_above\_}0.5\\ \Box\{\mathit{non\text{-}abandoned}\} &= \mathit{contains\_person\_but\_no\_bag} \vee \mathit{has\_total\_overlap}\\ &\quad \vee \mathit{has\_partial\_overlap} \vee \mathit{min\_distance\_below\_}0.1.\end{aligned}$$

Note that if none of the above decision rules is satisfied than "none of known" class is assigned to avoid misclassification. The ambiguity conditions, determined as $\mathcal{A}_c = \Diamond\{\mathit{abandoned}\} \cap \Diamond\{\mathit{non\text{-}abandoned}\}$, are composed of the following binary attributes:

$$\begin{aligned}\mathcal{A}_c = &\ \mathit{contains\_bag} \vee \mathit{contains\_person} \vee \mathit{has\_no\_overlap}\\ &\vee \mathit{min\_distance\_above\_}0.1 \vee \mathit{min\_distance\_above\_}0.25.\end{aligned}$$

The possibility conditions of the two classes can then be obtained by $\Diamond\{\mathit{abandoned}\} = \Box\{\mathit{abandoned}\} \vee \mathcal{A}_c$ and $\Diamond\{\mathit{non\text{-}abandoned}\} = \Box\{\mathit{non\text{-}abandoned}\} \vee \mathcal{A}_c$.

3.4.2. Use Case 2. Estimating the predisposition to failure of an industrial vehicle

This use case, studied in [15], involves predicting failures in a large fleet of heavy-duty trucks. The application context is completely different to that of the scene classification use case. It deals with multi-source data [21], composed of operational multivariate time series data collected by on-board sensors, including maintenance, repair and servicing information, and technical component specifications collected by the production system. The latter describes each vehicle by several specifications, each of which has a few possible values/categories. In addition, labels are available for each vehicle to indicate whether it experienced a failure during the observation period. Thus, we are dealing with a binary classification task again: healthy (the negative class, i.e., vehicles not experiencing failures) and failing vehicles (the positive class). Note that again we are dealing with MIL problem since not all available specifications can be assigned to each vehicle.

The multi-source nature of this use case makes a neuro-symbolic modelling integration strategy particularly beneficial, as explained below. Specifically, we consider two modelling components: (1) a neural component for the multivariate time series data, and (2) a symbolic component for the technical specification data. The symbolic component employs our modal logic, evidence-based approach introduced in the previous sections to determine a vehicle's predisposition to failure based solely on its technical characteristics. A multi-valued mapping is constructed between the set of technical specifications and the set of vehicles in the training dataset. Its inverse and super-inverse images are then used to derive the possibility and necessity conditions for each class, respectively. Subsequently, the plausibility and belief measures of the two classes are computed as outlined in (3.3), representing the lower and upper bounds of the predisposition to failure for each class.

Two different neuro-symbolic integration strategies are explored in [15]. The first strategy uses the outputs of the symbolic component as additional features for the neural component, thereby infusing it with background knowledge contained in the data. The second strategy uses the neural component to generate features, which are then combined with the outputs of the symbolic component to create a rule-based model. More concretely, an LSTM-autoencoder is trained to reconstruct sequences of time series data for healthy vehicles. The reconstruction error obtained is then combined with the evidence measures from the symbolic component's output to define the following logic rule:

$$\text{IF } \mathit{reconstruction\_error}(y) > T_1 \text{ } \mathit{and} \text{ } (\mathrm{Pl}_+(y) > T_2 \text{ } \mathit{or} \text{ } \mathrm{Bel}_+(y) > 0)$$
$$\text{THEN } \mathit{failure\_detected},$$

where y is the entity that must be classified, Pl_+ and Bel_+ are plausibility and belief scores produced by the positive class, and T_1 and T_2 are empirically determined thresholds.

4. Unraveling the Complexity of the Learning Task

The research results discussed in the foregoing subsections provide only a partial view of how evidence theory and modal logic can be applied to analyze and reason about uncertain scenarios. Their full potential remains largely unexplored. For instance, within our MIL context, the concepts of evidence theory can be leveraged to evaluate the quality and strength of a training set, quantify the uncertainty it contains, and assess how well it supports the learning task at hand.

4.1. Mass Assignment and Measures of Uncertainty

The uncertainty presented in a given training dataset $(Y, \mathcal{L})$ can be quantified by analyzing the mass function generated by its multi-valued model. More concretely, the multi-valued mapping $\mathcal{F}$ defined above induces a basic probability assignment (a mass function), which assigns to each $B \subseteq Y$ the value $m(B)$ as specified in (2.3). Alternatively, considering the modal logic interpretation in (2.5), $m(B)$ can be expressed as follows:

$$m(B) = P\big(\|\Box B \wedge (\bigwedge_{y \in B} \Diamond y)\| \mid \|\Diamond Y\|\big).$$

Thus the mass of B is expressed as the probability of the conjunctive combination of the necessity of B and the possibility of each of its individual members. The value of $m(B)$ should be interpreted as a measure of the exact belief assigned to B, and not to any of its subsets and it can be employed to quantify the uncertainty presented in the training dataset.

For instance, to capture epistemic uncertainty within the evidential framework, Dubois and Prade introduced the concept of *non-specificity* (NS) [22]. The NS of a mass function m is defined as follows:

$$\mathrm{NS}(m) = \sum_{B \subseteq Y} m(B) log_2(\mid B \mid). \tag{4.1}$$

Note that the mass function m, being defined on the power set of Y, says nothing particular about the elements (singletons) of Y, i.e., the individual entities. A transformation leading to a Bayesian mass functions (i.e., mass function whose focal elements are singletons) is preferred when dealing with real-world application contexts. The so-called pignistic transformation is very often used for this purpose converting m into the pignistic probability distribution BetP_m as proposed by Smets and Kennes in [23]. More concretely, BetP_m is defined for all the subsets B of Y by:

$$\mathrm{BetP}_m(B) = \sum_{C \subseteq Y} m(C) \frac{\mid B \cap C \mid}{\mid C \mid}. \tag{4.2}$$

Note that in case B is a singleton $\{y\}$ then

$$\mathrm{BetP}_m(\{y\}) = \sum_{y \in C} \frac{m(C)}{\mid C \mid},$$

or in other words the pignistic probability of any $y \in Y$ can be expressed with the equally distributed masses of the focal elements of m to which y belongs. BetP_m is a probability distribution on $\mathcal{P}(Y)$, but it was called pignistic by Smets and Kennes ([23]) for the sake of emphasizing that it is derived from a belief (mass) function. It is supposed, as argued by Smets and Kennes ([23]), to serve situations when one must make a forced decision given a belief function. In which case it is necessary to find a rule that allows for the construction of a probability distribution from the belief function. Subsequently, another measure of uncertainty, induced by the pignistic transformation of a belief function into a probability distribution and called *ambiguity measure*, was proposed in [24]:

$$\mathrm{AM}(m) = \sum_{y \in Y} \mathrm{BetP}_m(y) log_2 \left(\mathrm{BetP}_m(y)\right). \tag{4.3}$$

4.2. **Feature Dispersion and Class Support**

Let us illustrate the usefulness of the above concepts by considering a concrete example of a probability distribution on $\mathcal{P}(X)$. A straightforward approach is to exploit the frequency of occurrence of the different features $x \in X$ in the training set $(Y, \mathcal{L})$. In this way, a probability value can be assigned to each $A \subseteq X$ as follows:

$$P(A) = \frac{1}{\mathcal{D}} \sum_{x \in A} \mid \mathcal{F}(x) \mid, \quad \text{where} \quad \mathcal{D} = \sum_{x \in \mathrm{dom}(\mathcal{F})} \mid \mathcal{F}(x) \mid .$$

Feature dispersion and support. Note that $\mathcal{D}$ can be interpreted as expressing the degree of *dispersion* of the feature set X into the training set Y. Subsequently, using (2.3), it can be demonstrated that for any subset $B \subseteq Y$

$$m(B) = \mid \mathcal{F}^{-1}(B) \mid \frac{\mid B \mid}{\mathcal{D}}. \tag{4.4}$$

We call $\mid \mathcal{F}^{-1}(B) \mid$ the feature *support* of B, which is the number of features/instances $x \in X$ for which $\mathcal{F}(x) = B$, i.e., the ones supporting B. Thus the mass assigned to B is proportional to its feature support weighted by the ratio of its size and the feature dispersion $\mathcal{D}$ of X into Y. For noisy training sets, $\mathcal{D}$ is expected to be relatively large, in which case $m(B)$ will be small (negligible) number.

Class support and non-specificity. Subsequently, it would be interesting to examine the mass $m(Y_i)$ assigned to each class Y_i $(i = 1, \ldots, c)$ verifying whether any of the classes belongs to the set of focal elements of m, i.e., $m(Y_i) > 0$. Note that the mass values can be considered as expressing the belief assigned by the available evidence in support of each class. The total mass assigned to all the classes, i.e.,

$$\mathcal{S}_c = \sum_{i=1}^{c} m(Y_i) = \frac{1}{\mathcal{D}} \sum_{i=1}^{c} \mid \mathcal{F}^{-1}(Y_i) \mid \cdot \mid Y_i \mid$$

captures the strength of the overall belief in the partition of Y into the respective classes. Values of $\mathcal{S}_c$ close to 1 will indicate a solid evidence supporting the construction of a model discriminating between the distinct c classes in Y.

The non-specificity of m, as outlined in (4.3), can then be computed as follows:

$$\mathrm{NS}(m) = \frac{1}{\mathcal{D}} \sum_{B \subseteq Y} \mid B \mid log_2 \left(\mid B \mid^{|\mathcal{F}^{-1}(B)|} \right)$$

and used to quantify the uncertainty present in the training set Y.

Class probability and ambiguity. Further, the pignistic transformation of m into the pignistic probability distribution BetP_m (see (4.2)) can be expressed for any $B \subseteq Y$ and any $y \in Y$, respectively:

$$\mathrm{BetP}_m(B) = \frac{1}{\mathcal{D}} \sum_{C \subseteq Y} \mid \mathcal{F}^{-1}(C) \mid \cdot \mid B \cap C \mid \qquad \text{and} \qquad \mathrm{BetP}_m(y) = \frac{1}{\mathcal{D}} \sum_{y \in C} \mid \mathcal{F}^{-1}(C) \mid .$$

Thus, the (pignistic) probability of any $y \in Y$ is the sum of the feature supports of the focal elements of m to which y belongs and it can be considered expressing the degree of (evidence) support to each training sample in the training set. The probability of each class Y_i (for $i = 1, \ldots, c$) can be calculated, using the above formulas, for the purpose of examining how well each of the classes is supported by the extracted feature set. The overall ambiguity contained in the training set can be subsequently estimated, as proposed in (3.2).

In summary, the following measures can be used to characterize and assess each MIL scenario in advance: feature dispersion; training set non-specificity and ambiguity; individual and overall class support; and individual class probability.

5. **The Way Forward: Combination of Evidence Sources**

In real-world applications, where typically annotated data is scarce, it might be very beneficiary to be able to combine several evidence sources in order to leverage the available labels. There are many different approaches to do so described in the literature, e.g., consult [25, 26]. The combination of evidence sources in Dempster–Shafer theory is based on the assumption that these sources are independent and provide different assessments of the same frame of discernment, i.e., the same set Y in our case. *Dempster's rule of combination* [2] was the first one defined in the framework of evidence theory. According to it, the conjunctive combination of two mass functions m_1 and m_2, both defined on $\mathcal{P}(Y)$, is denoted by $m_1 \oplus m_2$ and defined for any $B \subseteq Y$ as follows:

$$m_1 \oplus m_2(B) = \frac{1}{1 - \mathbf{k}} \sum_{C \cap D = B} m_1(C) m_2(D), \tag{5.1}$$

where $C, D \subseteq Y$ and $\mathbf{k}$ is the degree of conflict obtained by:

$$\mathbf{k} = \sum_{C \cap D = \emptyset} m_1(C) m_2(D).$$

Intuitively, according to (5.1), the new mass function $m_1 \oplus m_2$ combines the information from the two mass functions m_1 and m_2 by aggregating the contributions of all intersecting focal elements of m_1 and m_2, respectively, that form B. The denominator, $1-\mathbf{k}$, is a normalization factor that has the effect of completely ignoring conflict and attributing any probability mass associated with conflict to the null set [27]. Consequently, several alternatives have been proposed in the literature attempting to mitigate this shortcoming, e.g., the disjunctive combination proposed by Dubois and Prade in [28]:

$$m_1 \cup m_2(B) = \sum_{C \cap D = B} m_1(C)m_2(D) + \sum_{C \cup D = B \,\wedge\, C \cap D = \emptyset} m_1(C)m_2(D). \tag{5.2}$$

Note that both rules can be extended for more that two mass functions.

Although evidence combination has been the subject of intensive research over the past two decades, and some approaches, such as the two outlined above, have already found their way into various application contexts, the combination rules defined so far remain limited in both scope and applicability. In particular, the assumption that different evidence sources always refer to the same frame of discernment does not adequately reflect real-world scenarios. Consider the MIL tasks discussed in this paper. One can envisage integration scenarios involving the addition of new features, new training samples, or both. Such extensions lead to an enriched training set and clearly violate the aforementioned assumption. To address these scenarios, new combination rules are required, rules that can update class expressions without necessitating the retraining of models on extended datasets. For example, in the use case presented in Section 3.4.1, which focuses on surveillance scene classification for detecting abandoned luggage, one may collect labeled images from a new location. In such a case, it would be highly beneficial to integrate the corresponding modal logic, evidence-based models into a single model that consolidates the knowledge acquired from both locations. However, neither of the combination rules introduced in (5.1) and (5.2) is suitable for this purpose. In the second use case, concerning the estimation of failure risk for industrial vehicles (Section 3.4.2), even more complex integration scenarios may arise.

There is a clear need for the development of novel combination rules that address the specific requirements of contemporary real-world applications, which often involve multiple evidence sources of diverse nature and reliability within complex integration settings. In this work, we aim to demonstrate that evidence theory and modal logic provide powerful formal frameworks for modelling and reasoning under uncertainty. They enable the seamless incorporation of domain knowledge, support the definition of transparent decision logic, and facilitate the development of hybrid models that integrate other modelling paradigms, such as deep learning. We believe that the full potential of these conceptual frameworks remains largely underexplored by the research community and that they can play a crucial role in the ongoing quest for reliable, explainable, and truly reasoning AI models.

Acknowledgments

This research received funding from the Flemish Government (AI Research Program).

Veselka Boeva's research was funded partly by the Knowledge Foundation, Sweden, through the Human-Centered Intelligent Realities (HINTS) Profile Project (contract 20220068).

References

[1] A. Dempster. "Upper and lower probabilities induced by a multivalued mapping". In: *The Annals of Mathematical Statistics* 38 (1967), pp. 325–339.

[2] G. Shafer. "A Mathematical Theory of Evidence". In: Princeton University Press, Princeton, 1976.

[3] V. Boeva, E. Tsiporkova, and B. D. Baets. "Modelling uncertainty with kripke's semantics". In: *Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 1998. LNCS.* Ed. by F. Giunchiglia. Vol. 1480. Springer, Berlin, Heidelberg, 1998.

[4] E. Tsiporkova, V. Boeva, and B. De Baets. “Dempster–Shafer theory framed in modal logic”. In: *International journal of approximate reasoning* 21.2 (1999), pp. 157–175.

[5] E. Tsiporkova, B. D. Baets, and V. Boeva. “Evidence theory in multivalued models of modal logic”. In: *Journal of Applied Non-Classical Logics* 10.1 (2000), 55–81.

[6] T. R. Besold et al. “Neural-Symbolic Learning and Reasoning: A Survey and Interpretation”. In: *ArXiv* abs/1711.03902 (2017).

[7] A. d. Garcez et al. “Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning”. In: *arXiv preprint arXiv:1905.06088* (2019).

[8] A. d. Garcez and L. C. Lamb. “Neurosymbolic ai: The 3 rd wave”. In: *Artificial Intelligence Review* 56.11 (2023), pp. 12387–12406.

[9] G. Marcus. “The next decade in AI: four steps towards robust artificial intelligence”. In: *arXiv preprint arXiv:2002.06177* (2020).

[10] P Hitzler, M. Sarker, T. Besold, A. Garcez, S Bader, H Bowman, P Domingos, P Hitzler, K. Kühnberger, L. Lamb, et al. “Neural-symbolic learning and reasoning: A survey and interpretation”. In: *Frontiers in artificial intelligence and applications* 342 (2022), pp. 1–51.

[11] G. Murtas, V. Boeva, and E. Tsiporkova. “An evidence-based neuro-symbolic framework for ambiguous image scene classification”. In: *19th International Conference on Neurosymbolic Learning and Reasoning*. 2025. URL: https://openreview.net/forum?id=6UnuZcQ2zY.

[12] J.-P. Aubin and H. Frankowska. “Set-Valued Analysis”. In: Birkhäuser, Boston–Basel–Berlin, 1990.

[13] B. De Baets and E. Kerre. “A revision of Bandler-Kohout compositions of relations”. In: *Mathematica Pannonica* 4 (1993), pp. 59–78.

[14] B. Chellas. “Modal Logic, an Introduction”. In: Cambridge University Press, Cambridge, 1980.

[15] G. Murtas, V. Boeva, and E. Tsiporkova. “Neuro-LENS: a neuro-symbolic framework integrating incomplete background knowledge and deep learning”. In: *Neurosymbolic Artificial Intelligence* (2026). accept. URL: https://neurosymbolic-ai-journal.com/system/files/nai-paper-961.pdf.

[16] C Chow. “On optimum recognition error and reject tradeoff”. In: *IEEE Transactions on information theory* 16.1 (2003), pp. 41–46.

[17] T. Denoeux. “A k-nearest neighbor classification rule based on Dempster-Shafer theory”. In: *IEEE transactions on systems, man, and cybernetics* 25.5 (1995), pp. 804–813.

[18] T. Denoeux. “A neural network classifier based on Dempster-Shafer theory”. In: *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans* 30.2 (2000), pp. 131–150.

[19] B. Dubuisson and M. Masson. “A statistical decision rule with incomplete knowledge about classes”. In: *Pattern recognition* 26.1 (1993), pp. 155–165.

[20] P. Spagnolo, A. Caroppo, M. Leo, T. Martiriggiano, and T. D’Orazio. “An Abandoned/Removed Objects Detection Algorithm and Its Evaluation on PETS Datasets”. In: *2006 IEEE International Conference on Video and Signal Based Surveillance*. 2006, pp. 17–17.

[21] Z. Kharazian, T. Lindgren, S. Magnússon, O. Steinert, and O. Andersson Reyna. “Scania component x dataset: A real-world multivariate time series dataset for predictive maintenance”. In: *Scientific Data* 12.1 (2025), p. 493.

[22] D. Dubois and H. Prade. “Properties of measures of information in evidence and possibility theories”. In: *Fuzzy sets and systems* 24.2 (1987), pp. 161–182.

[23] P. Smets and R. Kennes. “The transferable belief model”. In: *Artificial Intelligence* 66.2 (1994), 191–234.

[24] A.-L. Jousselme, C. Liu, D. Grenier, and E. Bosse. “Measuring ambiguity in the evidence theory”. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 36.5 (2006), pp. 890–903.

[25] E. Lefevre, O. Colot, and P. Vannoorenberghe. “Belief function combination and conflict management”. In: *Information Fusion* 3.2 (2002), pp. 149–162.

[26] K. Sentz and S. Ferson. “Combination of evidence in Dempster-Shafer theory”. In: (2002).

[27] R. R. Yager. “On the Dempster-Shafer framework and new combination rules”. In: *Information sciences* 41.2 (1987), pp. 93–137.

[28] D. Dubois and H. Prade. “Representation and combination of uncertainty with belief functions and possibility measures”. In: *Computational Intelligence* 4.3 (1988), pp. 244–264.