Fusing Expert Knowledge and Internet of Things Data for Digital Twin Models: Addressing Uncertainty in Expert Statements

Michelle Jungmann
AIFB Institute
Karlsruhe Institute of Technology
Germany
michelle.jungmann@kit.edu

Sanja Lazarova-Molnar
AIFB Institute
Karlsruhe Institute of Technology
Germany
Mærsk Mc-Kinney Møller Institute
University of Southern Denmark
Denmark
lazarova-molnar@kit.edu

ABSTRACT

Extracting Digital Twin models by fusing expert knowledge with Internet of Things data remains a challenging and open research area. Existing literature offers very limited approaches for seamless and systematic extraction of Digital Twin models from these combined sources. In this paper, we address the research gap by proposing a novel approach that considers and integrates the uncertainty inherent in human expert knowledge into the extraction processes of Digital Twin models. Given that experts possess unique experiences, contextual understandings and judgements, their knowledge can be highly divergent, complex, ambiguous, and even incorrect or incomplete. Consequently, not all expert knowledge statements should be equally weighted in the resulting simulation models. Our contributions include a comprehensive literature review on the uncertainty in expert knowledge and the proposal of an approach to integrate this uncertainty in the extraction of Digital Twin models from fused expert knowledge and IoT data. We demonstrate our approach through a case study in reliability assessment.[1]

CCS CONCEPTS

• **Computer methodologies → Modeling and simulation → Modeling methodologies**

KEYWORDS

ACM proceedings, Digital Twins, Fusion of Data and Expert Knowledge, Uncertainty in Expert Knowledge, Industry 4.0

ACM Reference format:

Michelle Jungmann and Sanja Lazarova-Molnar. 2025. In *SAC '25: The 40th ACM/SIGAPP Symposium on Applied Computing Proceedings, Catania, Italy, March 31 - April 4, 2025 (SAC ' 25)*, 8 pages. https://doi.org/10.1145/3672608.3707826

SAC'25, March 31 – April 4, 2025, Catania, Italy

ACM ISBN 979-8-4007-0629-5/25/03
https://doi.org/10.1145/3672608.3707826

1 INTRODUCTION

Digital Twins (DTs) have gained attention over the last years, and are currently used in many domains, such as aerospace, smart cities, manufacturing and logistics [14]. With the capability to simulate, analyze and optimize [33], DTs allow a wide area of application and have the potential to enable, e.g., time and cost reductions in manufacturing [11].

The Industry 4.0 revolution, driven by technologies like the Internet of Things (IoT), has enabled data-driven DT model extraction approaches that aim for automation and utilize (real-time) IoT data [10, 14]. However, these data-driven DT model extraction processes often neglect valuable insights offered as expert knowledge (EK) from experts working with these systems [28]. Thus, systematic fusion of EK and IoT data for hybrid DT model extractions can result in more robust and better-informed DT models [16].

Hybrid model extractions that seamlessly and systematically integrate EK with IoT data for DT model extraction remain a highly complex task and a significant research gap [16]. In our earlier work [16], we initiated addressing this gap by proposing a Fusion DT Framework for extracting DT models from both EK and IoT data, while also identifying key challenges and opportunities for implementing the framework. Building on this foundation, in our subsequent work [17], we presented a Proof of Concept (POC) for a hybrid DT model extraction approach. This approach systematically fuses EK and IoT data to extract simulation models, focusing on the domain of reliability assessment in manufacturing.

EK from experts is, however, often formulated as ambiguous and complicated natural language expert knowledge statements (EKSs). EKSs can be incomplete, conflicting or divergent, e.g., when made by different individual experts [5, 8, 9, 29]. Consequently, EK inherently exhibits uncertainty [4, 5, 9, 15, 34] that should be integrated in DT models from fused EKSs and IoT data to accurately reflect real-world conditions.

To the best of our knowledge, no existing approaches in literature systematically and seamlessly fuse EK and IoT data for DT model extraction, while also integrating uncertainty inherent to EKSs in these extracted DT models. We address this research gap by extending on our previous approach to consider uncertainty during the fusion process. Specifically, we propose an approach that maps the uncertainty in expert knowledge to precise probabilities of occurrence during the integration of IoT data and EK. As a result, the extracted DT models embed uncertainty for each EKS, offering valuable information for simulation runs.

The remainder of this paper is organized as follows: In Section 2, we provide a brief overview of DTs and DT model extraction processes. We also discuss the concept of uncertainty in EK and review the literature on mapping uncertainty in EKSs to Fuzzy Logic. Additionally, we provide background on Fuzzy Petri nets, which we use to represent DT underlying models. We conclude Section 2 with a literature review that addresses uncertainty in EK for DT models. In Section 3, we propose our extended approach that captures uncertainty of EKSs in Fuzzy Production Rules and extracts DT Fuzzy Petri net models that incorporate uncertainty while fusing EKSs and IoT data. In Section 4, we demonstrate, and discuss our approach through a case study on reliability modeling. In Section 5, we conclude the paper.

2 BACKGROUND AND RELATED WORK

2.1 Digital Twins for Reliability Assessment

The concept of DTs was first introduced by Grieves in 2002 and gained attention in recent years in academia and industry [32]. However, there still does not exist one unified definition or framework [24, 32]. For this paper, we define a DT, based on the definition by VanDerHorn [32], as a DT being a digital representative of a physical entity that features a bi-directional communication between the digital representative and the physical entity. Whereas, the bi-directional communication link distinguishes a DT from the concepts of a Digital Model and a Digital Shadow [21, 31].

DTs can enable cost and risk reductions, efficiency and service improvements, as well as security, reliability and resilience increases [32]. These potentials are enabled by capabilities such as monitoring, simulation, analysis and optimization [24, 33]. Based on DTs' broad potential, DTs are applied in various areas such as aeronautics, healthcare, smart cities or manufacturing [14, 24].

In the manufacturing domain, reliability assessment is an important application area for DT models [3, 10]. Faults and failures in manufacturing systems can lead to high costs if they, e.g., cause production outages [35]. DTs can enhance reliability of systems by, e.g., simulations, monitoring, optimization and predictive maintenance [3]. Therefore, we focus our case study on reliability in the manufacturing domain. Furthermore, we focus on faults, which describe the event(s) that trigger the failure(s) of the manufacturing system or component [22].

2.2 Digital Twin Model Extractions

The literature identifies three main categories of techniques for DT model extraction: 1) knowledge-driven; 2) data-driven; and 3) hybrid [36]. In recent years, the Industry 4.0 revolution has fostered a change from knowledge-driven towards data-driven extractions [12, 13].

The knowledge-driven model development approach is based on EK and, thus, requires significant manual human input, making it unsuitable for automated adaption to frequent changes in the production environment [10]. Furthermore, developing or updating these models demands extensive effort from experts, leading to high costs and time inefficiencies [10]. On the other hand, data-driven model extraction seeks to automate the process by relying solely on sensor and IoT data [10]. However, this approach is often hindered by issues such as noise, errors and lack of data. On top of that, even with large datasets, data-driven methods may struggle to reliably detect certain events [36].

The afore-described challenges of knowledge-driven and data-driven approaches can be mitigated by combining them into a hybrid model extraction approach. This would allow IoT data to be leveraged, while filling gaps of missing information with EK, which provides context and experience [16, 28]. With this hybrid extraction approach, better-informed DT models can be achieved. However, achieving the seamless and systematic fusion of IoT data and EK for hybrid approaches remains a highly complex and unresolved research challenge [16, 17].

2.3 Uncertainty in Expert Knowledge

In our previous work [17], we introduced three evaluation metric dimensions to assess the difficulty of EKSs, which we further elaborate on in this work with a focus on uncertainty: 1) degree of belief; 2) correspondence to an IoT data level; and 3) time distance between fault and failure.

Degree of belief captures the inaccuracy and uncertainty in EKSs. *Correspondence to an IoT data level* refers to the correlation between IoT data and EK information. Thus, the criteria describes how many and which IoT data levels one EKS spans. *Time distance between fault and failure* captures how difficult the fault failure correlation is based on the amount of historic data needed.

In this paper, we focus on the evaluation dimension *degree of belief* to address the various levels of uncertainty that is included in EKSs [23]. As EKSs are made by individual experts, each featuring a different level of experiences, context understandings, judgements and decisions [5], EKSs can be highly diverge, complex, ambiguous or even incomplete [9]. Another characteristic leading to uncertainty is that experts often state EKSs in complex and ambiguous natural language [8]. Thus, working with EK is a highly challenging task.

Nevertheless, uncertainty contains valuable information and has, therefore, become an important subject in complex systems [5, 9]. In literature, uncertainty is defined as the deviation from a complete determinism [15] and has three aspects: 1) epistemic; 2) stochastic/ontological; and ambiguity [4, 5, 9, 15, 34], elaborated in the following.

1) Epistemic refers to a lack of knowledge or incomplete knowledge. Experts can have insufficient or incomplete knowledge about the system of interest, which leads to uncertainty [5]. This kind of uncertainty can happen if experts have a lack of understanding or ignore available facts [5], which leads to experts stating wrong or flawed information in the EKSs. Incomplete knowledge from experts can also lead to EKSs that are stated vaguely with approximations.

2) Stochastic/ontological refers to the unpredictability due to natural variability in the system.

3) Ambiguity refers to the fact that there are multiple valid frames of knowledge. Experts can understand the same system of interest in different forms, which can lead to conflicts [5] in EKSs. Furthermore, information in EKSs can be interpreted differently by experts [5]. Additionally, from a natural language processing (NLP) point of view, EKSs can be stated in various formulations and different words due to the ambiguity of natural language [8]. On top of that, if experts are uncertain about absolute figures or the correctness of their statement, they use terms from linguistic modifiers such as "almost certainly", "possible", "probably", "fairly" or "slightly" [19, 20]. However, every expert formulates knowledge differently, which means that, e.g., "probably" does not always express the same amount of uncertainty. This can lead to ambiguities in the meaning [2]. Therefore, for this paper, we assume that experts utilize linguistic modifier terms in the same ratio so that each linguistic modifier means the same uncertainty in EKSs independent of the expert stating it.

In this paper, we consider aspects 1) and 3) as we focus on uncertainty in EKSs and not uncertainty in the physical system behavior itself. We further split aspects 1) and 3) into two extraction categories for uncertainty information:

1) extracted directly from the EKSs; and
2) extracted from meta-information.

Meta-information contains additional properties about the expert that cannot be extracted from pure EKSs information. Thus, conclusions drawn from meta-information are associated with the expert itself and not the specific EKS. Meta-information from experts can be properties such as the experience, the job role or the years of company affiliation.

For capturing uncertainty from meta-information, we propose the term *"trustworthiness"*, where trustworthiness is calculated as the mean from the available meta-information. Based on our definition, an expert that is very experienced receives a higher trustworthiness than an inexperienced one. The same holds for the trustworthiness of an expert known for their precision, in comparison to an expert known for slips of the tongue. For this paper, we assume that the trustworthiness of each EKS is equal to the trustworthiness of the expert stating it. With this assumption, the EKS that is associated with the higher trustworthiness could be chosen in case of a contradiction between EKSs. This supports determining if an EKS is more correct than another EKS or correlated IoT data, which will be developed in our future work within a cross-validation. In further work, we will also consider that experts can have different trustworthiness levels dependent on the context domain. With this, the same expert can have a high trustworthiness for, e.g., machines, and a lower trustworthiness for, e.g., logistic processes. We focus this paper on uncertainty that can be directly extracted from EKSs natural language terms and reserve trustworthiness for our future work.

2.4 Mapping Uncertainty in Expert Knowledge Statements to Fuzzy Logic

In literature, there are multiple approaches to evaluate uncertainty of natural language expressions with probability values [2]. E.g., Fuzzy Logic (FL) has the ability to model uncertainty and vagueness and, thus, can be used for modeling of semantics from natural language [29]. However, it is difficult to standardize the probability of natural language terms [6] due to ambiguities [2]. Ambiguity is caused, e.g., as the meaning of terms such as "very likely" depends from human to human [6] and is highly dependent on the context [29]. In the following we focus our literature review on authors that propose firstly FL applications and secondly FL probability values for concrete natural language terms.

Authors such as Omoregbe et al. [30], Chowdhary [8], Zadeh [38], Novák [29] and Baaj and Poli [2] utilize FL in relation to NLP, but do not consider transforming concrete examples of natural language terms and linguistic modifiers into FL with concrete probabilities or uncertainties. Omoregbe et al. [30] use a fuzzy inference engine, where the degree of membership is manually set by a human expert. Chowdhary [8] mentions fuzzy matching. Zadeh [38] states that FL is able to generalize and compute probabilistic, possibilistic or elastic propositions from natural language. Novák [29] proposes an approach for fuzzy natural logic that fuses natural logic from the domain of linguistic and FL to capture semantics of natural language based on the work of Zadeh. Baaj and Poli [2] propose an approach that generates natural language expressions from fuzzy rules to enhance the understanding of a fuzzy inference system decision.

Several authors proposed methods to calculate uncertainty or vagueness from natural language terms and linguistic modifiers by assigning concrete values, as elaborated in the following. Kent [19] proposes a uncertainty scale for the domain of intelligence agencies. Budescu et al. [6] propose a uncertainty scale for the domain of climatology. While, Zadeh [37] shows the possibility of linguistic characterizations for numeric data inputs based on membership functions.

We selected the literature from Kent [19], Budescu et al. [6] and Zadeh [37] as foundation for this paper, as these are highly-cited, widely recognized in literature, and propose concrete probability values for terms and linguistic modifiers. Based on these sources, we summarized and created a mapping of terms and probability of occurrence values, shown in Table 1. For a better overview, we grouped terms and linguistic modifiers with similar meanings and probability of occurrence values into same uncertainty categories. This resulted with seven categories, each containing terms with similar meanings and associated probabilities.

Table 1: Uncertainty in EK based on terms

Terms	Probability of Occurrence
Virtually certain [6] ≈ Certain [19]	≈ >99% [6] ≈ 100% [19]
Very likely [6, 37] ≈ Very probable [37] ≈ Almost certain [19]	≈ >90% [6] ≈ 87% to 99% [19] ≈ 0.25/0.6 + 0.49/0.7 + 0.81/0.8 + 1/0.9 + 1/1 [37]
Likely [6, 37] ≈ Probable [19]	≈ >66% [6] ≈ 63% to 87% [19] ≈ 9.5/0.6 + 0.7/0.7 + 0.9/0.8 + 1/0.9 + 1/1 [37]
About as likely as not [6] ≈ More or less likely [37] ≈ Not likely [37] ≈ Neither very probable nor very improbable [37] ≈ Chances about even [19]	≈ 33% to 66% [6] ≈ 40% to 60% [19] ≈ 1 / (0 + 0.1 + 0.2 + 0.3 + 0.4 + 0.5) + 0.5/0.6 + 0.3/0.7 + 0.1/0.8 [37]
Unlikely [6, 37] ≈ Improbable [37] ≈ Probably not [19]	≈ <33% [6] ≈ 20% to 40% [19] ≈ 1/0 + 1/0.1+ 0.9/0.2 + 0.7/0.3 + 0.5/0.4 [37]
Very unlikely [6, 37] ≈ Very improbable [37] ≈ Almost certainly not [19]	≈ <10% [6] ≈ 2% to 13% [19]
Exceptionally unlikely [6] ≈ Impossible [19]	≈ <1% [6] ≈ 0% [19]

In this paper, for our approach and case study, we focus on the items summarized in Table 1. Additional terms, linguistic modifiers and qualifiers, such as "rather", "fairly", "slightly" [20] are part of our future work.

2.5 Fuzzy Petri nets as Modeling Formalism for Integrating Uncertainty

DT models can be represented with different modeling formalism such as Markov chains, Fault Trees or Petri nets (PNs). As PNs can be directly extracted from data-driven process mining and are widely used to represent DT models, also in the area of reliability assessment [18], we choose PNs and their variants as modeling formalism.

As we aim to represent uncertainty in EKSs from human experts in our DT models, we utilize a variant of PNs called Fuzzy Petri nets (FPNs) for this paper. This variant, introduced by Looney [25], has the capability to represent knowledge and reasoning in expert systems [23]. Based on this capability, knowledge that contains imprecision, vagueness or fuzziness can be captured in FPNs [23].

FPNs utilize common elements of PNs, i.e., places, transitions and arcs. Places are drawn as cycles, transitions as bars and directed arcs describe the relationship between places and transitions [23]. FPNs contain the parameter μ ($\mu \in [0,1]$), associated with a transition, that represents the certainty factor and indicates how certain the rule is true, i.e., the degree of belief [23]. The transition firing process represents the knowledge reasoning process. The arcs connecting places and transitions, show the interconnection between propositions (places) and reasoning rules [23].

For this paper, we utilize Fuzzy Production Rules (FPRs) as general knowledge formalization, as they play an essential part in storing and representing expert knowledge that is stated in a vague way [23]. FPRs are mostly presented in a fuzzy if-then rule in the form of IF *a* THEN *c* (μ), where both the *if* and the *then* part are expressed through FL terms. For each rule, *a* is the antecedent part and *c* is the consequent part [23]. Both parts can comprise multiple propositions containing fuzzy variables joined with AND/OR [7, 23].

2.6 Existing Approaches on Addressing Uncertainty from Expert Knowledge in Digital Twin Models

In our literature review, we focus on approaches within the domain of DTs that use FPNs or FL to integrate EK into their concepts and implementations. We present some of these works as follows. Manocha et al. [26] propose an approach inspired by DTs and FL for flood prediction in the smart city domain and propose to integrate their solution into a DT in further work. Alves de Araujo Junior et al. [1] present an approach for the electric sector and power plants in Brazil. They extract fuzzy rules and develop a DT for a water-cooling system that calculates the number of efficiently operated fans, thus, supporting the decision-making process. Similarly, Monek and Fischer [27] propose an Expert Twin system that reduces reaction times and improves the efficiency in decision support. They introduce a DT framework in which human EK is integrated through FL in the decision-making module.

The works by Manocha et al. [26], Alves de Araujo Junior et al. [1] and Monek and Fischer [27] have in common that they do not detail on EKSs and their formalization by using FL. The authors also do not focus on the extraction of DT models from both EKS and IoT data and do not utilize FPNs as modeling formalism simultaneously.

Thus, to the best of our knowledge, no existing approaches integrate uncertainty from EKSs in the fusion of EKSs and IoT data for DT model extractions. Similarly, there are no existing approaches that extract DT models as FPNs generated by IoT data and natural language EKSs in the field of reliability assessment for manufacturing. Therefore, we aim to address this research gap by incorporating uncertainty in DT model extractions.

3 ADDRESSING UNCERTAINTY IN EXPERT STATEMENTS FOR MODEL EXTRACTION

To address the identified research gap of considering uncertainty in EK within the fusion of EK and IoT data for DT model extractions, we extend on our previously proposed hybrid fused model extraction approach for EK and IoT data from [17]. As noted, we focus our approach and use case on reliability assessment of manufacturing systems. With the extension of our approach both expert knowledge and its associated uncertainty are considered through a probability of occurrence. This ensures that not all EKSs are equally weighted during simulations.

To consider uncertainty in DT model extractions, we use FPRs as formalization for EKSs. We, further, add an IN part for stating the component affected by the fault. We define the formalization for each EKS and its correlated uncertainty as:

$$\textit{IF}\ \text{antecedent}\ \textit{THEN}\ \text{consequent}\ \textit{IN}\ \text{component}\ (\mu), \quad (1)$$

where μ is a certainty factor, and equal to the probability of occurrence and degree of belief derived from literature. The certainty factor is calculated as:

$$\mu = (100\%\ \text{certainty}) - (\text{mean of uncertainty in EKS in }\%).$$

In [17], we proposed four data-EK fusion algorithms in two fusion strategies to extract stochastic Petri nets from fused EKSs and IoT data. In this paper, we take our previously proposed PRE-FA and POST-FA data-EK fusion algorithms as the foundation, as the algorithms focus on fault information by extracting guard functions for the transitions. In this paper, we extend both algorithms to extract FPNs from EKSs that contain uncertainty expressed through terms using natural language. We refer to these extended algorithms as FUZZY-PRE-FA and FUZZY-POST-FA algorithms. Pseudocode extracts of both algorithms are shown in Fig. 1. Both algorithms have in common that they take formalized EKSs (FEKSs) as input, formalized as shown in (1), and extract FPNs as output. The extracted FPNs then both contain FEKSs fault information and uncertainty as guard functions and certainty factors.

Algorithm 1: FUZZY-PRE-FA extract.

```
syndata = synthetic state condition log;
sv = sensor value; lsv = lowest sv; t= transition
for FEKS in list of FEKSs do
    filter syndata for resource equal to component;
    for entry in filtered syndata do
        get sv based on FEKS condition
        add sv to sv list
    end
    find lsv out of sv list;
    find assigned transition t;
    set guard function to larger or equal to lsv in t;
    set certainty factor from FEKS as µ in t;
end
for t in transitions of model do
    if t does not have guard function then
        set certainty factor of t to 0.995;
    end
end
```

Algorithm 2: FUZZY-POST-FA extract.

```
c = condition; t = transition; s = statement;
for s in list of FEKS do
    set c to s between "IF" and "THEN";
    add c to c list;
    add µ to µ list;
end
for t in transitions of model do
    if t is assigned to c in c list then
        modify t to type immediate;
        set guard function of t to assigned c;
        set certainty factor of t to assigned µ;
    end
    else
        set certainty factor of t to 0.995;
    end
end
```

Figure 1: Fusion algorithm pseudocode extracts.

The FUZZY-PRE-FA algorithm is part of the *a priori* strategy (FEKSs fusion during FPN extraction) that requires synthetic state condition IoT data, as additional input. The synthetic state condition IoT log is a merge of the IoT condition monitoring log data and IoT state log data. Based on synthetic data and FEKSs, the algorithm finds guard function values and assigns certainty factors to transitions.

The FUZZY-POST-FA algorithm is part of the *a posteriori* strategy (FEKSs fusion after FPN extraction) and assigns guard functions and certainty factor values, directly extracted from FEKSs, to the assigned transitions.

As IoT data contains noise and errors, we assume that it does not possess a 100% certainty. Further, we assume that the certainty factor of transitions extracted from IoT data contains the same uncertainty as if an expert states "certain" in their EKS. Therefore, we assume a IoT data certainty factor of 99.5%, transferred from Table 3. This certainty factor is added to every transition not annotated with FEKSs. In the following, we demonstrate our approach.

4 CASE STUDY: INTEGRATION OF UNCERTAINTY INTO SIMULATION MODELS

4.1 Case Study in Reliability Modeling

With this case study, we show that uncertainty can be effectively captured when extracting DT models from a fusion of EKS and IoT data. For this, we extend our case study from [17], where we fused EKSs and IoT data to extract DT PNs within a POC in the reliability assessment domain.

The case study is based on a manufacturing system with two manufacturing cells, Cell1 and Cell2, and one Automated Guided Vehicle (AGV) that delivers material to Cell2. We utilize the simulation model, created IoT data and *ddra* library based PN data-EK fusion algorithms from our previous work [17] as foundation and modify them further in this case study to integrate uncertainty. In our case study, we focus on EKSs containing faults describing the manufacturing system.

Our case study consists of five steps: 1) design four EKSs containing fault information and uncertainty; 2) map EK terms to certainty factor probabilities; 3) formalize EKSs into FPRs; 4) extract FPNs using FUZZY-PRE-FA and FUZZY-POST-FA algorithms; and 5) validate and discuss results.

4.2 Step 1: Expert Knowledge Statements

We create four EKSs, shown in Table 2, that contain information about the case study manufacturing system. The EKSs include additional information, faults consisting of multiple conditions and uncertainty. We highlighted the terms that show uncertainty in Table 2 by underlining.

Table 2: EKSs containing uncertainty

	Expert Knowledge Statements
EKS1	If Sensor1 of manufacturing Cell1 reaches a value of <u>probable</u> 4 and Sensor2 a value of above <u>probable</u> 2, the <u>chances are about even</u> that Cell1 needs repair.
EKS2	Cell2 fails <u>almost certainly</u> within the next hour as soon as there is a sharp loud noise from the machine in Cell2. With a <u>high likelihood</u> the noise is somewhere above 90 decibels.
EKS3	It is <u>impossible</u> that the AGV continues its operation, if Material2 is delivered twisted from the AGV.
EKS4	I am <u>certain</u> that if the AGV runs more than 2 months without maintenance, there is the <u>possibility</u> that Cell1 fails and needs repair.

4.3 Step 2: Translation of Expert Knowledge Statements into Certainty Factors

Based on our literature review in Subsection 2.4, we assume the certainty factors for natural language terms in EKSs for this paper. The concluding terms and values are displayed in Table 3. Each listed term, e.g., "certain" is a representant for the terms within the same meaning category established in Table 1. To calculate concrete certainty factors for our FPN, we took the highest and lowest value for the probability of occurrence from literature and calculated the mean from the upper and lower bound. With this, e.g., "very probable" gets the certainty value of 92% assigned. This mapping is then the basis to transform EKSs into FPRs in the next subsection.

Table 3: Quantification of uncertainty

Term	Literature Range	Certainty Factor
Certain	99% - 100%	99.5%
Very probable	85% - 99%	92%
Likely, Possible	66% - 90%	78%
Chances about even	33% - 66%	49.5%
Improbable	10% - 40%	25%
Very unlikely	1% - 13%	7%
Impossible	0% - 1%	0.5%

4.4 Step 3: Formalization of Expert Knowledge Statements into Fuzzy Production Rules

In this step, we manually formalize EKS1-4 into FPRs, as shown in (1). The resulting EKSs are shown in Table 4.

Table 4: Formalization Approach of EKSs into FPRs

	Formalized EKSs	Certainty Factor
FEKS1	*IF* Cell1_sensor1 == 4 *AND* Cell1_sensor2 >= 2 *THEN* failure *IN* Cell1	μ = 0.685
FEKS2	*IF* Cell2_sensor >= 90 *THEN* failure *IN* Cell2	μ = 0.92
FEKS3	*IF* Material2 == twisted *THEN* failure *IN* AGV	μ = 0.005
FEKS4	*IF* maintenance > 60 days *THEN* failure *IN* AGV	μ = 0.8875

The certainty factor μ is assigned to the FEKSs based on the probability of occurrence values in Table 3. If one EKS contains multiple linguistic modifier terms related to the same circumstance, we calculate the mean of the probabilities. In EKS1 there are the linguistic modifiers "probable", "probable" and "chances are about even" involved. From Table 3, "probable" is associated with 78% and "chances are about even" is associated with 49.5%. Thus, we calculate the certainty factor value with (0.78 + 0.78 + 0.495) / 3 = 68.5 %.

4.5 Step 4: Extraction of Fuzzy Petri nets Including Uncertainty

To extract the FPN shown in Fig. 2, we execute our two proposed fusion algorithms, FUZZY-PRE-FA and FUZZY-POST-FA, on FEKS1-4. In Fig. 2, we highlighted the extracted guard functions and certainty factors for the respective transitions and annotated, which algorithm extracted which FEKS.

For this case study, we decided which algorithm to use based on the availability of IoT and sensor data. As there is synthetic state condition IoT data available for FEKS1-2 from our case study in [17], we executed FUZZY-PRE-FA on FEKS1-2. This is why, e.g., for FEKS2, the guard function gets 93.47 decibel assigned and not 90 as stated in FEKS2. As there are no real IoT data available for twisted materials and maintenance cycles in this case study setup, we are not able to generate synthetic IoT data for FEKS3-4. Thus, the FUZZY-PPRE-FA algorithm cannot be applied. Therefore, we use the FUZZY-POST-FA algorithm for FEKS3-4 to extract the guard function values and certainty factors directly from the FEKSs.

FUZZY-PRE-FA, executed on FEKS1, adds one guard function to the transition "fail_cell1" with two fault conditions connected over an AND as well as the certainty factor of 68.5%. As FEKS3 and FEKS4 describe both a fault of the AGV, the transition "fail_agv" is annotated with two independent guard functions and two different certainty factors (0.5% and 88.75%) for each fault. All transitions without an assigned guard function from EKSs get the certainty factor of 99.5% assigned independent of which algorithm we choose.

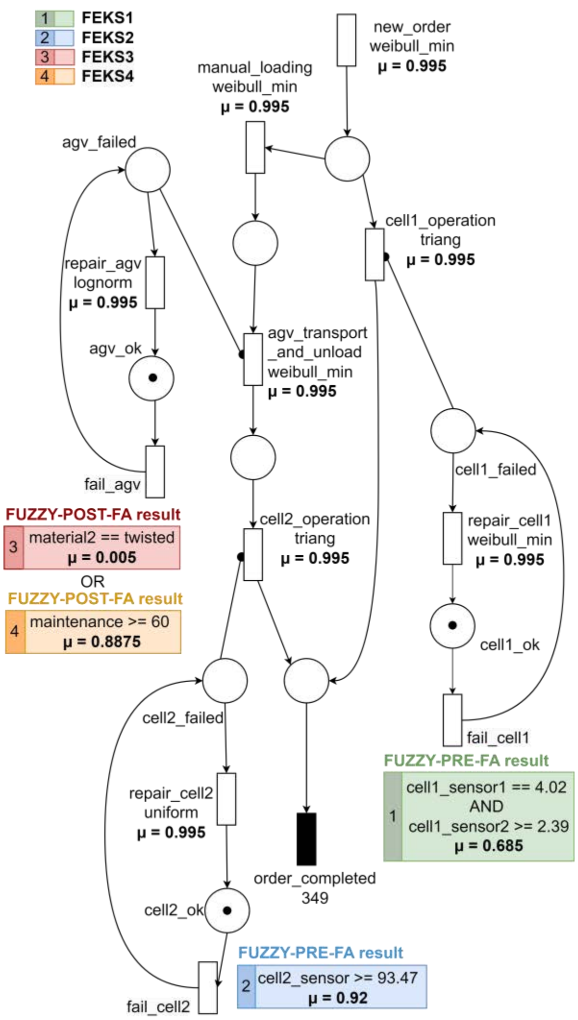

Figure 2: Fuzzy Petri net containing uncertainty of FEKSs extracted from FUZZY-PRE-FA and FUZZY-POST-FA.

4.6 Step 5: Result Validation and Discussion

We conclude that our proposed approach effectively captures uncertainty of EKSs during the extraction of DT models, based on a face validity and manual comparison approach [39]. Specifically, our approach extracts FPNs from both IoT data and EKSs, incorporating uncertainty in the information extracted from EKSs. This process involves extracting places and transitions, as well as integrating certainty factors and guard functions to the transitions. The face validity assessment confirms that all guard functions and certainty factors are correctly extracted and assigned to the respective transitions with our case study examples. As part of our future work, we aim to further refine validation processes.

When comparing both algorithms in their functionalities, we conclude that the FUZZY-PRE-FA extracts more concrete guard function values, as it does not depend on vague values stated in the FEKSs. Instead, the algorithm finds the concrete sensor fault values in the synthetic IoT data, which includes real sensor values from the system of interest. Thus, we chose this algorithm whenever IoT data is available from which then synthetic data can be merged. However, as the FUZZY-POST-FA algorithm solely depends on FEKSs, it can be used in more cases, especially when IoT data is unavailable or insufficient.

As future refinement, we aim to enhance validation by automating the checking and refinement of certainty factors and automating the formalization approach using NLP techniques. We, furthermore, aim to develop cross-validation methods for handling complementing or conflicting EKSs and IoT data.

5 CONCLUSIONS AND OUTLOOK

In this study, we developed an approach to incorporate uncertainty in expert knowledge statements during the fusion of expert knowledge and data for the extraction of Digital Twin models. Our approach includes a formalization technique and two fusion algorithms, which we, subsequently, demonstrated in a case study focused on reliability assessment. For future work, we aim to refine and extend our approach by, e.g., automating algorithm execution and decision-making processes, as well as implementing a cross-validation between data and expert knowledge. Additionally, we aim to extend our approach to Weighted Fuzzy Petri nets, enabling integration of uncertainty derived from expert trustworthiness into Digital Twin models.

ACKNOWLEDGMENTS

The authors extend their thanks for the funding received from the DMaaST project funded by the European Commission, Horizon Europe Programme under the Grant Agreement No. 101138648.

REFERENCES

[1] Carlos A. Alves de Araujo Junior, Juan M. Mauricio Villanueva, Rodrigo J. S. de Almeida, and Isaac E. Azevedo de Medeiros. 2021. Digital Twins of the Water Cooling System in a Power Plant Based on Fuzzy Logic. *Sensors*

21, 20 (January 2021), 6737. https://doi.org/10.3390/s21206737

[2] Ismaïl Baaj and Jean-Philippe Poli. 2019. Natural Language Generation of Explanations of Fuzzy Inference Decisions. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, June 2019. 1–6. https://doi.org/10.1109/FUZZ-IEEE.2019.8858994

[3] Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. 2019. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* 7, (2019), 167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499

[4] M Brugnach, A Dewulf, C Pahl-Wostl, and T Taillieu. 2007. Towards a relational concept of uncertainty: Incorporating the human dimension. (2007).

[5] Marcela Brugnach, Art Dewulf, Claudia Pahl-Wostl, and Tharsi Taillieu. 2008. Toward a Relational Concept of Uncertainty: about Knowing Too Little, Knowing Too Differently, and Accepting Not to Know. *Ecology and Society* 13, 2 (2008). Retrieved August 2, 2024 from https://www.jstor.org/stable/26267972

[6] David V. Budescu, Han-Hui Por, and Stephen B. Broomell. 2012. Effective communication of uncertainty in the IPCC reports. *Climatic Change* 113, 2 (July 2012), 181–200. https://doi.org/10.1007/s10584-011-0330-3

[7] Shyi-Ming Chen. 1996. A fuzzy reasoning approach for rule-based systems based on fuzzy logics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 26, 5 (October 1996), 769–778. https://doi.org/10.1109/3477.537318

[8] K. R. Chowdhary. 2020. Natural Language Processing. In *Fundamentals of Artificial Intelligence*, K.R. Chowdhary (ed.). Springer India, New Delhi, 603–649. https://doi.org/10.1007/978-81-322-3972-7_19

[9] A. Dewulf, M. Craps, R. Bouwen, T. Taillieu, and C. Pahl-Wostl. 2005. Integrated management of natural resources: dealing with ambiguous issues, multiple actors and diverging frames. *Water Science and Technology* 52, 6 (September 2005), 115–124. https://doi.org/10.2166/wst.2005.0159

[10] Jonas Friederich, Deena P. Francis, Sanja Lazarova-Molnar, and Nader Mohamed. 2022. A framework for data-driven digital twins of smart manufacturing systems. *Computers in Industry* 136, (April 2022), 103586. https://doi.org/10.1016/j.compind.2021.103586

[11] Jonas Friederich, Sune C. Jepsen, Sanja Lazarova-Molnar, and Torben Worm. 2021. Requirements for Data-Driven Reliability Modeling and Simulation of Smart Manufacturing Systems. In *2021 Winter Simulation Conference (WSC)*, December 2021. 1–12. https://doi.org/10.1109/WSC52266.2021.9715410

[12] Jonas Friederich and Sanja Lazarova-Molnar. 2021. Towards Data-Driven Reliability Modeling for Cyber-Physical Production Systems. *Procedia Computer Science* 184, (2021), 589–596. https://doi.org/10.1016/j.procs.2021.03.073

[13] Jonas Friederich and Sanja Lazarova-Molnar. 2022. Data-Driven Reliability Modeling of Smart Manufacturing Systems Using Process Mining. In *2022 Winter Simulation Conference (WSC)*, December 2022. 2534–2545. https://doi.org/10.1109/WSC57314.2022.10015301

[14] Aidan Fuller, Zhong Fan, Charles Day, and Chris Barlow. 2020. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* 8, (2020), 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358

[15] J. A. E. B. Janssen, M. S. Krol, R. M. J. Schielen, A. Y. Hoekstra, and J. -L. de Kok. 2010. Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. *Ecological Modelling* 221, 9 (May 2010), 1245–1251. https://doi.org/10.1016/j.ecolmodel.2010.01.011

[16] Michelle Jungmann and Sanja Lazarova-Molnar. 2024. Towards Fusing Data and Expert Knowledge for Better-Informed Digital Twins: An Initial Framework. In *Procedia Computer Science*, 2024. .

[17] Michelle Jungmann and Sanja Lazarova-Molnar. 2024. Fusing Expert Knowledge and Data for Simulation Model Discovery in Digital Twins: A Case Study from Reliability Modeling. In *2023 Winter Simulation Conference (WSC). IEEE*, 2024. .

[18] Sohag Kabir and Yiannis Papadopoulos. 2019. Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. *Safety Science* 115, (June 2019), 154–175. https://doi.org/10.1016/j.ssci.2019.02.009

[19] Sherman Kent. 1964. Words of Estimative Probability. Studies in intelligence, 8.4 (1964), 49–65.

[20] Etienne E. Kerre and Martine De Cock. 1999. Linguistic Modifiers: An Overview. In *Fuzzy Logic and Soft Computing*, Guoqing Chen, Mingsheng Ying and Kai-Yuan Cai (eds.). Springer US, Boston, MA, 69–85. https://doi.org/10.1007/978-1-4615-5261-1_5

[21] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 51, 11 (2018), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474

[22] Sanja Lazarova-Molnar, Nader Mohamed, and Hamid R. Shaker. 2017. Reliability modeling of cyber-physical systems: A holistic overview and challenges. In *2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, April 2017. 1–6. https://doi.org/10.1109/MSCPES.2017.8064536

[23] Hu-Chen Liu, Jian-Xin You, ZhiWu Li, and Guangdong Tian. 2017. Fuzzy Petri nets for knowledge representation and reasoning: A literature review. *Engineering Applications of Artificial Intelligence* 60, (April 2017), 45–56. https://doi.org/10.1016/j.engappai.2017.01.012

[24] Xin Liu, Du Jiang, Bo Tao, Feng Xiang, Guozhang Jiang, Ying Sun, Jianyi Kong, and Gongfa Li. 2023. A systematic review of digital twin about physical entities, virtual models, twin data, and applications. *Advanced Engineering Informatics* 55, (January 2023), 101876. https://doi.org/10.1016/j.aei.2023.101876

[25] C.G. Looney. 1988. Fuzzy Petri nets for rule-based decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics* 18, 1 (January 1988), 178–183. https://doi.org/10.1109/21.87067

[26] Ankush Manocha, Sandeep Kumar Sood, and Munish Bhatia. 2023. Digital Twin-assisted Fuzzy Logic-inspired Intelligent Approach for Flood Prediction. *IEEE Sensors Journal* (2023), 1–1. https://doi.org/10.1109/JSEN.2023.3322535

[27] Gergő D. Monek and Szabolcs Fischer. 2025. Expert Twin: A Digital Twin with an Integrated Fuzzy-Based Decision-Making Module. *Decision Making: Applications in Management and Engineering* (2025). https://doi.org/10.31181/dmame8120251181

[28] Parisa Niloofar and Sanja Lazarova-Molnar. 2021. Fusion of Data and Expert Knowledge for Fault Tree Reliability Analysis of Cyber-Physical Systems. In *2021 5th International Conference on System Reliability and Safety (ICSRS)*, November 2021. 92–97. https://doi.org/10.1109/ICSRS53853.2021.9660717

[29] Vilém Novák. 2017. Fuzzy logic in natural language processing. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2017. 1–6. https://doi.org/10.1109/FUZZ-IEEE.2017.8015405

[30] Nicholas A. I. Omoregbe, Israel O. Ndaman, Sanjay Misra, Olusola O. Abayomi-Alli, and Robertas Damaševičius. 2020. Text Messaging-Based Medical Diagnosis Using Natural Language Processing and Fuzzy Logic. *Journal of Healthcare Engineering* 2020, 1 (2020), 8839524. https://doi.org/10.1155/2020/8839524

[31] Mariana Segovia and Joaquin Garcia-Alfaro. 2022. Design, Modeling and Implementation of Digital Twins. *Sensors* 22, 14 (July 2022), 5396. https://doi.org/10.3390/s22145396

[32] Eric VanDerHorn and Sankaran Mahadevan. 2021. Digital Twin: Generalization, characterization and implementation. *Decision Support Systems* 145, (June 2021), 113524. https://doi.org/10.1016/j.dss.2021.113524

[33] Birgit Vogel-Heuser, Felix Ocker, Iris Weiß, Robert Mieth, and Frederik Mann. 2021. Potential for combining semantics and data analysis in the context of digital twins. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379, 2207 (August 2021), 20200368. https://doi.org/10.1098/rsta.2020.0368

[34] W.E. Walker, P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, and M.P. Krayer von Krauss. 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment* 4, 1 (March 2003), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466

[35] Heiko Webert, Tamara Döß, Lukas Kaupp, and Stephan Simons. 2022. Fault Handling in Industry 4.0: Definition, Process and Applications. *Sensors* 22, 6 (January 2022), 2205. https://doi.org/10.3390/s22062205

[36] Andrew Wunderlich, Kristen Booth, and Enrico Santi. 2021. Hybrid Analytical and Data-Driven Modeling Techniques for Digital Twin Applications. In *2021 IEEE Electric Ship Technologies Symposium (ESTS)*, August 03, 2021. IEEE, Arlington, VA, USA, 1–7. https://doi.org/10.1109/ESTS49166.2021.9512364

[37] L. A. Zadeh. 1975. The concept of a linguistic variable and its application to approximate reasoning-III. *Information Sciences* 9, 1 (January 1975), 43–80. https://doi.org/10.1016/0020-0255(75)90017-1

[38] Lotfi A. Zadeh. 2010. A Summary and Update of "Fuzzy Logic." In *2010 IEEE International Conference on Granular Computing*, August 2010. 42–44. https://doi.org/10.1109/GrC.2010.144

[39] Ashkan Zare and Sanja Lazarova-Molnar. 2024. Validation of Digital Twins in Labor-intensive Manufacturing: Significance and Challenges. *Procedia Computer Science* 238, (January 2024), 623–630. https://doi.org/10.1016/j.procs.2024.06.070