Check for updates

ACM DIGITAL LIBRARY

acm open

Latest updates: https://dl.acm.org/doi/10.1145/3743672

PDF Download
3743672.pdf
23 February 2026
Total Citations: 1
Total Downloads: 1716

Published: 12 July 2025
Online AM: 06 June 2025
Accepted: 02 June 2025
Revised: 15 April 2025
Received: 08 February 2023

Citation in BibTeX format

SURVEY

A Survey on Event Prediction Methods from a Systems Perspective: Bringing Together Disparate Research Areas

JANIK VASILY BENZIN, Technical University of Munich, Munich, Bayern, Germany

STEFANIE RINDERLE-MA, Technical University of Munich, Munich, Bayern, Germany

Open Access Support provided by:

Technical University of Munich

A Survey on Event Prediction Methods from a Systems Perspective: Bringing Together Disparate Research Areas

JANIK-VASILY BENZIN, TUM School of Computation, Information and Technology, Department of Informatics, Technical University of Munich, Garching, Germany
STEFANIE RINDERLE-MA, TUM School of Computation, Information and Technology, Department of Informatics, Technical University of Munich, Garching, Germany

Event prediction is the ability of anticipating future events, i.e., future real-world occurrences, and aims to support the user in deciding on actions that change future events towards a desired state. An event prediction method learns the relation between features of past events and future events. It is applied to newly observed events to predict corresponding future events that are evaluated with respect to the user's desired future state. If the predicted future events do not comply with this state, actions are taken towards achieving desirable future states. Evidently, event prediction is valuable in many application domains such as business and natural disasters. The diversity of application domains results in a diverse range of methods that are scattered across various research areas which, in turn, use different terminology for event prediction methods. Consequently, sharing methods and knowledge for developing future event prediction methods is restricted. To facilitate knowledge sharing on account of a comprehensive integration and assessment of event prediction methods, we take a systems perspective to integrate event prediction methods into a single system, elicit requirements, and assess existing work with respect to the requirements. Based on the assessment, we identify open challenges and discuss future research directions.

CCS Concepts: • **Information systems** → **Data mining**; **Decision support systems**; • **Computing methodologies** → **Machine learning;**

Additional Key Words and Phrases: Event prediction, artificial intelligence, predictive process monitoring, anomaly prediction, systems engineering

ACM Reference Format:
Janik-Vasily Benzin and Stefanie Rinderle-Ma. 2025. A Survey on Event Prediction Methods from a Systems Perspective: Bringing Together Disparate Research Areas. *ACM Comput. Surv.* 57, 12, Article 301 (July 2025), 37 pages. https://doi.org/10.1145/3743672

1 Introduction

The rapid growth of data [67] and major advances in enabling technologies such as cloud computing [118], Internet of Things [126], and machine learning [43] drives **event prediction** (**EP**) research (cf. Table 2). EP is the ability of anticipating events, i.e., future real-world occurrences. To

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) under project number 514769482.
Authors' Contact Information: Janik-Vasily Benzin, TUM School of Computation, Information and Technology, Department of Informatics, Technical University of Munich, Garching, Germany; e-mail: janik.benzin@tum.de; Stefanie Rinderle-Ma, TUM School of Computation, Information and Technology, Department of Informatics, Technical University of Munich, Garching, Germany; e-mail: stefanie.rinderle-ma@tum.de.

This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2025 Copyright held by the owner/author(s).
ACM 0360-0300/2025/07-ART301
https://doi.org/10.1145/3743672

anticipate events, EP processes past events together with further relevant data for learning an EP method to map the processed input to predicted future events that are of interest to the respective application domain. Being able to anticipate events is highly beneficial to many application domains such as business, healthcare, transportation, crime, and natural disasters [161]. The value of knowing and ideally understanding predicted future events today lies in the ability of a user in the application domain to act on the predicted future events, e.g., by focusing police resources in areas with high crime risk. In order to unlock the potential of EP, several challenges have to be addressed [161], including the knowledge on the true relationships between causes and effects of events [161], heterogeneous multi-output predictions (e.g., predicting the time and location of future events) [117], complex dependencies among the predicted future events as they may interact with each other [95], real-time stream of past events that requires continuous monitoring [122, 128], and data quality challenges [122].

Existing work on EP has proposed a multitude of different solutions to those challenges. However, the solutions are often developed within one application domain, e.g., healthcare [125], and corresponding research area, e.g., **predictive process monitoring** (**PPM**). The type of event and the laws and rules governing the occurrence of events greatly differ among the application domains, e.g., large-scale weather events governed by the laws and rules of meteorology in the domain of natural disasters versus small-scale business events of a business process governed by the respective business regulations and internal procedures in the domain of business [5]. Thus, EP research typically approaches the problem in an event-type-and domain-dependent manner [122, 161]. Despite apparent advantages of this approach to EP, it hampers sharing EP methods and solutions across different application domains and corresponding research areas. Most importantly, the fragmentation in EP research significantly complicates any effort to draw a clear picture on where we are standing on our road towards applying EP methods to real-world problems and how we should proceed. In this article, we aim at providing a clear picture on EP research, its fragmentation, current status, and remaining challenges by taking a *systems perspective*. Taking a systems perspective results in a common architecture with requirements for EP methods that are generally valid across domains and that enable assessing EP methods for real-world applications.

To illustrate the architecture at a high-level, Figure 1 depicts an intersection example from the transportation domain. The intersection has a road segment s_1 that is the *monitored system*. s_1 is depicted at two points in time: the current point in time t_{now} and two different future states at a future point in time $t_{\text{now+1}}$. The first future state is labeled with both the "future and predicted event" to which s_1 will progress without an action. Hence, s_1 will and is predicted to face heavy traffic without an action. The second future state is labeled with "future event with action", Thus, s_1 will have little traffic after an action. Our *event prediction goal* (cf. goal in Figure 1) for the intersection with road segment s_1 is to maintain little traffic for s_1. Currently, event e_n ="Traffic is green at s_1" at t_{now} *complies with* our goal. Event e together with the goal is observed by the *observer*, e.g., a video camera directed at road segment s_1, and a user interface for setting goals. In general, the observer is an abstract interface between the monitored system in the real-world and the system for EP. Therefore, the observer typically consists of a set of sensors, user interfaces and/or application programming interfaces.

Then, the observer sends event e_n and the goal to the *input component* of the generalized **predictive compliance monitoring** (**PCM**) system (cf. Table 2). The input component processes the event and goal for subsequent prediction by the *prediction component*. *PCM* is concerned with continuously anticipating whether the monitored system complies to a goal or set of goals in the future and supporting the user in understanding and acting on the anticipated compliance status [122]. Thus, a generalized PCM system conceptualizes EP through its input component and prediction component, i.e., the EP method is conceptualized as the prediction component. After

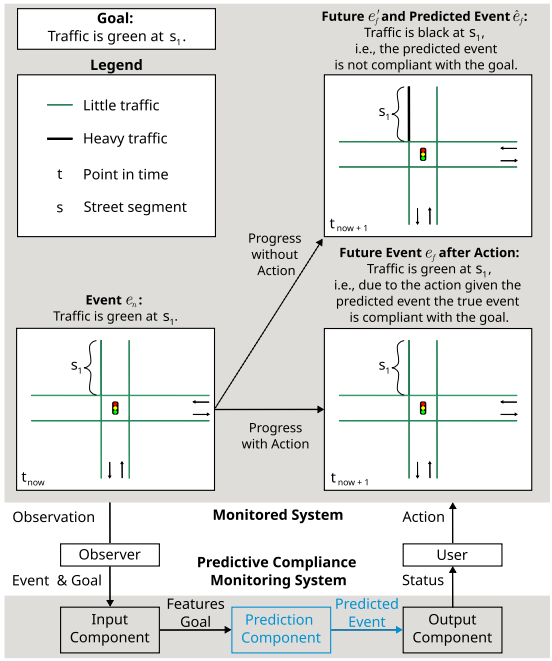

Fig. 1. EP as a component of a PCM system.

Table 1. Application Domains of Covered Articles

Domain	Short	Share	Count
Business	B	0.42	109
Engineering Systems	E	0.12	31
Healthcare	H	0.09	23
Political Events	P	0.09	23
Cyber Systems	I	0.08	22
Media	M	0.07	19
Crime	C	0.06	15
Natural Disasters	N	0.04	10
Transportation	T	0.03	8
$\sum$		1.00	260

Table 2. List of Abbreviations

Abbreviation	Meaning
EP	Event prediction
PPM	Predictive process monitoring
PCM	Predictive compliance monitoring
CM	Compliance monitoring

pre-processing and encoding the event in the input component, the prediction component correctly predicts $\hat{e}_f$ ="Traffic is black at s_1" at point in time $t_{\text{now}+1}$. Hence, the future progress of the monitored system is predicted to be non-compliant with our goal.

The *output component* of the generalized PCM system prepares a status for the user. Users are characterized by having the ability to take actions to steer the monitored system into a future direction compliant with the goal. For example, an urban traffic control center is a user that decides on the necessary action to change the future towards a compliant status. For road segment s_1, the chosen action is to reroute incoming traffic. Due to the action, the ground-true real event (true event) at $t_{\text{now}+1}$ is e_f ="Traffic is green at s_1". Thus, the PCM system supports the user to maintain a future that is compliant with the goal. To sum up, the user of the application domain is able to anticipate the future of the monitored system with respect to the goal and act accordingly on the account of the generalized PCM system.

Consider a practitioner that is faced with the problem of operationalizing the PCM system (cf. Figure 1). The practitioner is particularly interested in a quantification of how certain the method is in predicting the future event (*uncertainty quantification*). There are at least eight existing EP methods that focus on the transportation domain (cf. Table 1). Yet, there are at least 252 other EP methods that may be better suited for operationalizing the actual PCM system with uncertainty quantification in spite of the different domain. Because the research landscape is fragmented, it is very hard to understand whether an EP method with a specific property may exist and whether it is applicable to the PCM system. [98], for example, propose an EP method based on neural networks that predicts whether a logistics process will meet a delivery deadline. Does this EP method fit the needs of the practitioner? Giving an answer is difficult for several reasons. First, [98] is missing in existing surveys on EP methods such as [161]. Second, [98] conceptualize the logistics process in a specific modeling notation that the practitioner has to know in order to analyze to what extent the method is applicable to the real-world problem. Third, the title contains the term *business*

process. Hence, if the practitioner is a civil engineer, the EP method proposed in [98] may quickly be discarded. Lastly, [98] does not tell that the deadline is predicted through a binary answer and that reliability estimates refers to uncertainty quantification. The former can be deduced from the implicit assumption pervading related methods from the same area: simple yes/no answers are sufficient to represent future events [122].

To sum up, disparate research areas are difficult to navigate. Due to each area sharing its own conceptual models and implicit assumptions, the EP problem and the different solution methods seem incommensurable. Different terminology for EP such as PPM (cf. Table 2) in the field of business process management [33, 103, 143] or anomaly prediction in the field of engineering systems [85] exacerbates the problem of cumbersome method and knowledge sharing among various EP research areas. Here, taking a *systems perspective* on EP seems promising to place and integrate the diverse range of EP methods and, thus, foster knowledge and method sharing. The example in Figure 1 indicates how the system perspective benefits the integration of EP methods by giving the method a general, yet practical context that covers important parts such as goals and actions. Moreover, it gives a clear definition of the components, functionalities, and interfaces necessary to reap value from EP methods. Thus, we have a common anchor with which we can approach a given EP method during assessment. Some part of the PCM system and monitored system must ultimately be the cause for a design choice or assumption taken in an EP method. Hence, the system perspective allows us to integrate the apparently disparate research in EP methods. To integrate and assess existing work on EP methods with the aim of facilitating knowledge sharing and deriving open challenges and research directions, this survey is guided by the following research questions:

RQ1 How can existing EP methods be assessed in a general, consistent, and system-oriented manner?
RQ2 To what extent do existing EP methods fulfill assessment criteria?
RQ3 Which open challenges and research directions remain in the field of EP?

1.1 Research Methodology

The goal of answering **RQ1–RQ3** is to provide a comprehensive, consistent, and integrative overview of state-of-the-art EP across application domains and research areas and to facilitate knowledge sharing. Assume that a set of prediction goals $g_1, g_2, \ldots$ results from heterogeneous monitored systems and desired future states (cf. Figure 1). A set of different EP methods $m_1, m_2, \ldots$ is proposed for an EP goal. If EP method m_i can be applied to continuously anticipate prediction goal g_j, we say that method m_i is *applicable to* g_j. An ideal situation for knowledge sharing would be to have a common terminology for equivalent "applicable to" relations between methods and goals. The literature analysis in [161] underpins that currently we do not have an ideal situation for knowledge sharing. Consequently, the first challenge for the survey at hand is to collect and resolve different conceptual models, implicit assumptions, and terminology for equivalent prediction goals and corresponding methods. The second challenge is to integrate, assess, and present the vast amount of existing work such that our research questions are answered concisely.

To overcome the challenge of different conceptual models, implicit assumptions, and terminology, we take the survey in [161] as a baseline on existing work. We extend this baseline by searches in two directions: EP and PPM. The first direction aims to uncover EP method research that is missing in the baseline. As we have observed that EP research in the application domains engineering systems and cyber systems tend to use "indicator" or "anomaly" instead of "event", we have added both as keywords. The second direction aims to uncover missing EP method research mainly from the business application domain. The research area of business process management

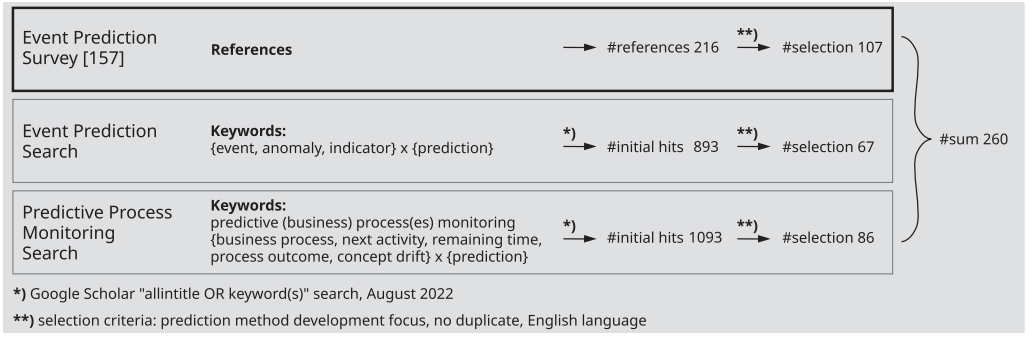

Fig. 2. Methodology for compiling literature.

uses different terminology for EP, e.g., PPM or next activity prediction. We include all keywords that we have observed during analysis of EP research in business process management. Also, we cross-check our keywords with the keywords used in a recent survey of EP in the business domain [122]. The methodology is illustrated in Figure 2 with keywords for searches and the results.

Given the 2202 search results, we apply the same filter denoted by **) in Figure 2 on each stream of literature resulting in selections that represent significant advancements in prediction method development. Here, significant advancement means work that focuses on novel EP methods instead of only applying a classical machine learning algorithm to a new application domain or focusing on domain-specific feature engineering. Lastly, we take the union of the three selections yielding 260 articles covered in this survey. The selection of 260 articles is classified into an integrated taxonomy and reported in Section B in the supplementary material.[1] These articles span a wide array of different application domains each having a large enough absolute article count for proper representation (cf. Table 1).

We combine an *integrative review* [155] with an *umbrella review* [155] and the systems perspective [12]. The integrative approach provides an overview of state-of-the-art EP through a qualitative research approach. The final selection of 260 articles contains 34 surveys from disparate research areas (cf. Table 3) which are taken as an initial yardstick for review and assessment in order to be consistent with existing work (cf. Figure 1 in the supp. material[1]). Synthesizing existing surveys from disparate research areas constitutes an umbrella review. Taking a systems perspective supports our mixed review by providing a mental [17] and reference model [145]. To this end, the PCM system explains the differences between EP methods and supports understanding different methods through a mental model. Also, it acts as a reference for aggregating requirements (cf. Section 3) and as a tool for applying EP methods in real-world (cf. Section 2.3).

For the general, consistent, and systematic assessment ($\mapsto$ RQ1), we identify existing requirements in the related surveys and merge them into an initial set of requirements (cf. Figure 1 in the supp. material[1]). Main challenges for merging requirements are different functional granularities and terminology for similar functionalities. We also encounter the challenge that identified requirements do not necessarily relate to the EP method itself, e.g., data requirements as identified in [122]. The system perspective helps us overcome this additional challenge. Similar to the initial set of requirements, the initial PCM system (cf. Figure 1 in the appendix) may not consistently, generally, and systematically represent all articles. Therefore, we apply the following analysis cycle

[1]The supplementary material consists of the supplemental online-only appendix Section A - F and our online repository https://gitlab.com/janikbenzin/event_prediction_survey for machine-readable material.

Table 3. Comparison of Our Survey with Related Surveys through Three Distinctive Survey Characteristics

Area	Related surveys	Assessing	System-integrative	Umbrella
EP	[13, 52, 65, 97, 101, 110, 161]	✗	✗	✗
PPM	[32, 33, 68, 76, 80, 103, 135, 138, 156]	✗	✗	✗
EP	[59, 62, 119, 140, 141]	✓	✗	✗
PPM	[57, 73, 75, 99, 105, 107, 116, 140, 142, 143, 147, 153]	✓	✗	✗
EP	Not available	✓	✓	✗
PPM	[122]	✓	✓	✗
Any	Survey at hand	✓	✓	✓

(cf. Figure 1 in the supp. material) for each of the 260 selected articles to overcome the challenge of uncovered requirements, different conceptual models and implicit assumptions.

The analysis cycle for current article i begins by its classification into the taxonomy. Then, the article is assessed with the current set of requirements. Potential issues during assessment, e.g., some characteristics of the article are not represented by the requirements yet or imply to split a requirement into two separate requirements, determine the refinement of requirements. As the PCM system shall represent all requirements and generalize over all EP methods and their applications, it is also refined for every article. By placing the EP method into the PCM system and iteratively comparing it to already analyzed methods, we are able to uncover implicit assumptions. Additionally, we are able to map components of different conceptual models into an equivalent system context. Last, the assessment of the previous articles is updated to represent the refined set of requirements. All in all, the analysis cycle for all 260 articles yields the PCM system and final set of requirements as presented in Section 3.

1.2 Related Surveys

The article selection contains 34 surveys. Some of these surveys focus on technical details of prediction methods in general and on specific problem formulations. [161] is taken as a yardstick for EP methods and presents the different methods in more detail when compared to the survey at hand. Table 3 compares the survey at hand with the 34 related surveys from a methodological perspective. This further differentiates the survey at hand from existing ones and aims at highlighting its uniqueness and comprehensiveness. 16 related surveys do not share any distinctive feature with the survey at hand. For example, [161] focuses more on classifying existing methods and gives a technical overview on approaches. Similar to 18 related surveys, this survey is *assessing* as it assesses the relative qualities of the EP methods either qualitatively or quantitatively. As outlined in Section 1.1, this survey also combines an *integrative review* [155] with an *umbrella review* [155] and a systems perspective [12]. We subsume the systems perspective and the integrative approach under *system-integrative* (cf. Table 3). A single related survey is additionally system-integrative. None of the surveys is an umbrella review. Hence, the survey at hand is the only one that aims at unifying existing EP method research in disparate areas by taking both research articles and surveys into account.

For synthesizing the 34 identified surveys from disparate areas, we start with the literature covered by EP survey [161] (cf. Figure 2). By contrast to [161], we focus only on the EP method instead of EP methods with encoding, preprocessing, and feature engineering (resulting in 107 of the 216 references) and extend the scope in the direction of uncovered EP method research (67 added articles) and uncovered PPM research (86 added articles). Aside from the scope, [161] focuses on the techniques to solve the identified EP research problems, e.g., regression. Hence, the existing

approaches are classified and summarized along inherent properties of the proposed solutions. This survey extends the classification in [161] by abstracting from the respective technique and integrating taxonomies proposed in other surveys [33, 52, 103, 122] (cf. Section B in supp. material[1]). With respect to summarizing existing approaches, our survey extends the summary in [161] by summarizing existing work along the requirements for EP methods that we elicit through a systems perspective, i.e., our summary takes an external, systems perspective on EP compared to the internal, methodological perspective of [161]. Furthermore, the requirements are used to assess existing EP methods, resulting in a comprehensive assessment that further extends [161].

Overall, we consider eleven EP surveys [13, 52, 59, 62, 65, 97, 101, 110, 119, 140, 141] (cf. Table 3). Each of them has a certain scope, e.g., [97, 101] focus on time-series data, [140, 141] focus on EP of business processes, and [13] on unstructured text events. When compared to the survey at hand, on top of a narrow scope, the eleven surveys neither take a system perspective to generally place EP method in a context, nor elicit requirements for assessing the methods, nor comprehensively assess existing work. [52] proposes a taxonomy for EP methods with a concept of domain-specificity that is integrated in our taxonomy.

We consider 22 PPM surveys [32, 33, 57, 68, 73, 75, 76, 80, 99, 103, 105, 107, 116, 122, 135, 138, 140, 142, 143, 147, 153, 156] (cf. Table 3). Similar to the reviewed EP surveys, the scope of the PPM surveys is limited. While the scope of EP surveys is heterogeneous, PPM surveys exclusively focus on next events (e.g., [140]) or key performance indicators of business processes (e.g., [147]). Only [122] takes a systems perspective for eliciting requirements and assessing existing PPM methods. Our survey generalizes the PCM concept to EP, develops a conceptual PCM system, and extends as well as integrates the requirements in [91, 122] with the challenges in [161] and the requirements contained in the 260 articles of the existing work selection. [33, 103] propose a similar concept of domain-specificity for classifying PPM methods that is generalized and integrated in our taxonomy. The distinction in [122] into methods that directly predict the goal of PPM and methods that indirectly predict the goal of PPM via an intermediate representation is generalized and integrated in our taxonomy. Predicting the late shipment status of an online shop delivery package, for example, is directly predicting the PPM goal and conceptually similar to the traffic status in the intersection example (cf. Figure 1). Predicting all future events pertaining to the delivery package and evaluating the shipment status on them is indirectly predicting the PPM goal and conceptually similar to predicting the number of vehicles on road segment s_1 to subsequently derive the traffic status in the intersection example.

If the time of the predicted event is of interest, EP becomes a classical, supervised classification problem for which methods such as naive Bayes or logistic regression can be applied. If the predicted events are represented as a time-series or necessitate time-series data (*discrete* or *continuous time* prediction), [11, 83] provide surveys based on deep learning, and [94] for other machine learning methods. If the predicted events are represented as spatio-temporal or necessitate spatio-temporal data (*location* prediction), [152] provides a survey of deep learning methods and [8] for other machine learning methods. If the predicted event is an anomaly, EP methods are typically coupled with anomaly detection methods for which [16] provides a general overview and [15] one for time-series data. [105, 116] provide surveys for categorical representation of the predicted event (*semantic prediction*).

1.3 Contribution and Structure

Taking a systems perspective on EP methods to integrate disparate areas of research results in the following four contributions of this survey:

— **Integration of disparate research areas through the generalized PCM system.** By taking a systems perspective, we identify an EP method as a component of the generalized

PCM system. The generalized PCM system further integrates contributions from the various research areas on EP methods. The systems perspective emphasizes integration challenges, e.g., regarding interoperability of system components. We scrutinise the integration challenges through two realizations of the PCM system in manufacturing and healthcare (the healthcare realization is reported in Section E of the appendix).

- **Structured requirement elicitation that encompasses the EP method within the PCM system.** The aim of integration through the PCM system is to elicit similar, yet broader requirements for EP methods. Integration through a set of general requirements facilitates the mutual exchange of method properties in general and of method properties that meet the requirements in particular. Without the broader view of the EP method in the PCM system, some, if not most, requirements are likely to be neglected or abstracted.
- **Qualitative assessment of existing methods with respect to the requirements and thorough analysis of current research status.** A qualitative assessment of all existing methods gives an in-depth view on the current research status. The current research status is illustrated with examples from the literature, analysed with respect to potential dependencies between requirements and discussed in detail.
- **Open challenges and discussion of future research directions.** The assessment result sheds light on blind spots that embody open challenges to be solved by future research. We identify five challenges and discuss the possible future research direction for each challenge.

Section 2 introduces a notion of an EP method and places EP methods into the generalized PCM system (↦ RQ1). Section 3 explains the requirements for assessing EP methods (↦ RQ1). Section 4 shows to what extent existing prediction methods fulfill requirements and pinpoints exemplary method properties that meet the respective requirement (↦ RQ2). Section 5 highlights open challenges and distills research directions (↦ RQ3). Lastly, Section 6 concludes, presents our survey's impact and identifies limitations of the survey. Note that we provide a supplemental online-only appendix and an online repository as supplementary material.[1] Amongst others, the supp. material features an integrated taxonomy of EP methods and a classification of all 260 EP methods into the integrated taxonomy.

2 A Notion and System Architecture for EP Methods

This section introduces a notion of the EP method (Section 2.1). It presents the rationale for the PCM system to embed EP methods into their system context (Section 2.2). It concludes with a demonstration of the PCM system in the manufacturing domain (Section 2.3).

2.1 EP Method Formulation

In the beginning, EP requires a goal or a set of goals that is either formulated by researchers during their motivation for advancement of EP methods or by practitioners during implementation and application of EP. Without the prediction goal, we do not know what we predict and lack a purpose for predicting. Prediction goals are as diverse as EP domains are. "The traffic is green at s_1" for the intersection example depicted in Figure 1 is an EP goal. "Reduce re-hospitalization and increase treatment outcome of heart failure patients" [36] is a set of EP goals. EP goals define the monitored system implicitly or explicitly and how we perceive various current and future states of the monitored system by differentiating desired or compliant from undesired or non-compliant states. In the intersection example, the goal implicitly defines the monitored system to be the depicted intersection and that green traffic is regarded as compliant, whereas black traffic is regarded as non-compliant. For hospital patients with previously diagnosed heart failures, the goal defines the monitored system to be hospital heart failure patients. Moreover, future states

with reduced re-hospitalization numbers and improved treatment outcome statistics are desired in contrast to, e.g., states with constant re-hospitalization.

Definition 2.1 (EP Goal). An EP goal g is a well-formed formula in some logic G with atomic propositions over the domain of the monitored system that evaluates to either true or false given some predicted instantiation of its variables. If the EP goal g evaluates to true, we say the monitored system future state is compliant, otherwise it is non-compliant.

For the sake of simplicity, we do not state the presented examples as formulas of a certain logic such as propositional logic or linear-time logic and point towards existing methods on translating natural language to formulas in, e.g., linear-time logic [20]. The two important points here are the existence of at least one EP goal and that it is stated in terms of atomic propositions over the domain of the monitored system (implicitly or explicitly).

Based on the EP goal and the monitored system, we identify two kinds of data used for EP,[2] i.e., historical event data $Y_0 \subseteq \mathcal{T}^- \times \mathcal{L} \times \mathcal{S}$ and historical indicator data $X \subseteq \mathcal{T}^- \times \mathcal{L} \times \mathcal{F}_I$, where $\mathcal{T}$ is the time domain, $\mathcal{T}^- \equiv \{t | t \leq t_{\text{now}}, t \in \mathcal{T}\}$ are all times up to the current time t_{now}, $\mathcal{T}^+ \equiv \{t | t > t_{\text{now}}, t \in \mathcal{T}\}$ are all times after the current time, $\mathcal{L}$ is the location domain, $\mathcal{S}$ is the event semantic domain and $\mathcal{F}_I$ is the indicator feature domain that does not include the time and location domain [161]. Data types Y_0 and X are generated by the observer (see Figure 1) through observing the real-world, e.g., the observer generates an event $y = (\text{“05.01.2023 13:00”}, \text{“}s_1\text{”}, \text{“Traffic = green”})$ by observing that two vehicles have stopped in front of the red light and an indicator $x = (\text{“05.01.2023 13:00”}, \text{“}s_1\text{”}, \text{“Workday = Yes”})$. This data is sent alongside the EP goal to the input component of the PCM system (cf. Figure 1) that maps the data to suitable features $F \subseteq \mathcal{T}^- \times \mathcal{F}_Y \times \mathcal{F}_X$, where $\mathcal{F}_Y$ is the domain of features based on historical event data and $\mathcal{F}_X$ is the domain of features based on historical indicator data, by means of preprocessing and encoding. The input component outputs the features and prediction goal to the prediction component that is realized by an EP method, e.g., $f = (\text{“05.01.2023 13:00”}, (1, 0), 1)$ for a one-hot encoded traffic value and identifying the workday by 1.

Definition 2.2 (EP Method [161]). Given features $F \subseteq \mathcal{T}^- \times \mathcal{F}_Y \times \mathcal{F}_X$ and an EP goal $g \in G$ or set of goals, an EP method outputs predicted future events $\hat{Y} \subseteq \mathcal{T}^+ \times \mathcal{L} \times \mathcal{S}$ such that for each future predicted event $\hat{y} = (t, l, s) \in \hat{Y}$ for $t > t_{\text{now}}$ it holds that Ⓒ either the goal g or the set of goals is itself a future predicted event or it can be evaluated on the future predicted events.

In the traffic example, let us assume that we learn a neural network model that takes the most recent feature f and outputs $\hat{y} = (\text{“05.01.2023 14:00”}, \text{“}s_1\text{”}, \text{“Traffic = black”})$ as the predicted future event (cf. Figure 1), i.e., the goal is itself the predicted event. Consequently, the future state of the traffic at road segment s_1 is predicted to be non-compliant such that an action by the user is required.

Following the scope and aim as set out in Section 1.1, we separate feature engineering from EP and focus on the EP method. We add condition Ⓒ on prediction goals. Ⓒ expresses the benefit of predicting the future events and how we perceive the possible predicted future events, namely as compliant with our desired future state or as non-compliant. Defining EP methods with an explicit goal has three benefits. First, all existing work in our selection state a goal or set of goals in their motivation that expresses the benefits to users and how a specific predicted future event is perceived, e.g., [36] predict future patient treatment outcomes in clinics for a proactive treatment effect analysis. Hence, including condition Ⓒ in the definition combines the *how* (i.e., the EP

[2]We adopt the existing definition of the EP problem and method in [161] as close as our different perspective and scope allows.

method) and the *why* (i.e., the prediction goal) for EP. Second, it links predicted future events to actions. Since an action aims to change a monitored system towards a desired state, linking actions to future predicted events requires a notion of a desired state. With prediction goals, the notion of desired states is defined by compliant or desired future states and by non-compliant or undesired future states. Third, condition © emphasizes how we can decide whether a specific EP method is successful in predicting future events.

If we want to fully understand EP in all of its heterogeneous aspects, multi-scale dimensions and overall complexity, we must achieve to grasp the set of systematic relationships of the system in which EP is conceptualized or implemented [40, 41, 54, 74, 120]. This feature of understanding emphasizes the need for a systems perspective on this topic. In general, considering the diverse nature of prediction goals throughout the various application domains, the type of system for EP must be a decision support system rather than an autonomous system, i.e., the user that decides what action is carried out is a human person (cf. Figures 1 and 3). For instance, deciding on the treatment plan of patients based on predicted future treatment outcomes must be done by physicians [29]. Furthermore, the utility of EP critically depends on the lead time, i.e., the time from the prediction of an event to the true event time. This property necessitates a prediction at run-time and, thus, an online, stream-based conceptualization of the EP system [128]. The source streams have to be continuously monitored such that the lead time is as large as possible [161]. Altogether, we need a decision support system that continuously monitors and predicts with the aim of distinguishing non-compliant, action-demanding from compliant future states.

To that end, we present a conceptual PCM system that entails these properties in the next section and operationalize the PCM system in Section 2.3. In Section 3, we elicit assessment requirements from the system architecture to asses existing EP methods in their "PCM system readiness".

2.2 Widening the Scope: EP Method in the PCM System

Before presenting the PCM system, we elaborate on how it applies to EP. This enables us to understand how multiple research areas work towards the same goal with more or less similar methods, yet different terminology. Moreover, we illustrate the community- and application-domain-spanning characteristic of EP.

In the field of business process management, a PCM system supports companies in monitoring and understanding the future compliance status of their business processes to act proactively towards a desired future compliance status [91, 122]. To that end, a PCM system integrates research on **compliance monitoring** (**CM**) and PPM. CM aims to evaluate compliance constraints on the current and future events or key performance indicators of a business process [91], while PPM aims to predict the future events or key performance indicators of a business process [33, 103]. By comparing CM, PPM, and the EP method definition (cf. Definition 2.2), we identify the following similarities: Compliance constraints are prediction goals, future key performance indicators of a business process correspond to future predicted events (cf. condition © in Definition 2.2), PPM is an EP method, and the business process is a monitored system. If we generalize

(1) the monitored business process and its events as a monitored system and its events,
(2) the prediction of the future compliance status as EP methods,
(3) the anticipated compliance status as the evaluation result of an EP goal,
(4) and the proactive actions as any action an user can take with the support of a decision support system,

then the generalized PCM system depicted in Figure 3 is a conceptual model of decision support systems featuring EP. Not only does this generalization place EP inside a generic system necessary for understanding its merits more deeply, it also represents an integration of research on the

various aspects of EP and monitoring into a common theme. Furthermore, the generalization of a PCM system from business process management to a decision support system for EP in general enables us to mutually transfer requirements and solutions between the respective research areas. A PCM system contains the input, prediction, and output component (cf. Figure 1) which are refined in Figure 3. Moreover, a feedback cycle is added from the user to the input component that contains the action as a more direct and efficient way of accounting for the effects of derived actions onto the monitored system in the PCM system [161].

2.2.1 Input Component. The input component *extracts* data from the observer, *preprocesses* the data and *encodes* it into the respective feature space required by the EP method (cf. Definition 2.2). In practice, the input component has to implement the various functionalities of *data collection* proposed by the data management community, in particular data acquisition and data labeling [124]. Within data acquisition, the tasks of augmentation through data integration/data fusion are particularly important. For the case of input event data lacking an event notion, the task of data programming or event extraction [35] become additionally relevant. Considering that the observer is only an abstraction for multiple, heterogeneous data sources (cf. Section 1), the input component's *extract* functionality has to deal with challenges of raw data ingestion from various data sources [96].

The input component's *preprocess* functionality covers the transformation of ingested data to a common representation based on a series of rules [66], data cleaning [64], schema matching for relational data [115], deduplication and record linkage [42, 50, 104], and, finally, canonical column selection for relational data [139]. In the case of non-relational data such as continuous, unlabelled sensor readings, schema matching is replaced by data programming or event extraction [35] to define an event notion. In case of low-level events coming from the event source, the task of event abstraction [146] may be necessary. Since events from multiple sources may actually be the same event, these events have to be matched to finalize the data integration. As a last step, the *encode* functionality maps the integrated data to the event $\mathcal{F}_Y$ and indicator feature space $\mathcal{F}_X$. Here, it is important to follow the principles of feature engineering [165], while the resulting features are typically domain- and/or goal-specific.

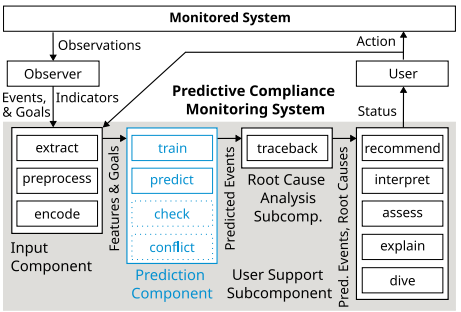

Fig. 3. Generalized PCM system—conceptual and structural view.

2.2.2 Prediction Component. The prediction component is realized by an EP method that *trains* a prediction model and *predicts* the future events. In Definition 2.2, the future event can either be (i) the prediction goal or (ii) an intermediate event on which the prediction goal can be evaluated. Case (i) does not require the additional functionalities *check* and *conflict*. For case (ii), the evaluation of the prediction goal is realized by *checking* the compliance of the intermediate event with respect to the prediction goal. If a monitored system is subject to a set of prediction goals, these goals may be in conflict [91]. The PCM takes this possibility into account when checking the set of prediction goals on the intermediate event and includes this information in its output.

2.2.3 Output Component. The output component is refined into two subcomponents: the **root cause analysis component** and the **user support component** (cf. Figure 3). The refinement underlines the need for model transparency and interpretability (in short explainability), accountability, and prescriptiveness [91, 122, 161]. The ability of the PCM to *trace* the non-compliance of

the future events *back* to the root cause(s) contributes to an explainable and actionable prediction. The prediction with root cause analysis is explainable, because a causal relationship between a root cause and its anticipated result exists [120]. Furthermore, root cause analysis makes the prediction actionable, since the user can derive countermeasures for non-compliant future events through acting on the root causes. In engineering and cyber systems, i.e., IT systems, root cause analysis is typically developed for explaining the root causes of machine faults or system anomalies to the operation manager [37, 92].

Finally, the PCM system supports the user in understanding the results and acting on them. To that end, the user support subcomponent *recommends* actions and countermeasures for non-compliant predicted future events [4, 109] and *interprets* the impact of the proposed actions on the predicted future events [55]. Furthermore, it *assesses* the overall uncertainty that remains throughout all prior functionalities [63]. Separately showing all the results to users may quickly overwhelm, distract from major points, and hinder understanding. Hence, the gist of the results are *explained* to them. In Section 5, we discuss a potential approach to this challenging functionality. If more detailed information is necessary for understanding or analysing, the user can *dive* into the respective results.

This new systems perspective encompasses a holistic view on the many facets of EP, yet is simple enough to guide us in our understanding of functionalities surrounding the actual predictions that are necessary to hone its value. Eventually, the PCM system depicted in Figure 3 can be seen as an approach and system blueprint for the realization of real-world applications, comparable to the *business model canvas* for business model creation [108]. Two PCM system realizations from manufacturing and healthcare are presented in the next section and in Section E of the supp. material.[1]

2.3 Realization of a PCM System in Manufacturing

In this section, we show how the PCM approach (cf. Figure 3) can be used in order to realize EP projects as the example of manufacturing. *"Manufacturers need to know what happens next and what actions to take in order to get optimal results. It is a challenge to develop advanced analytics techniques including machine learning and predictive algorithms to transform data into relevant information for gaining useful insights to take appropriate action"* [144]. This means that several event sources in a cyber-physical environment are to be leveraged to predict drifts and outcomes in a production process. Figure 4 illustrates how the PCM system can be applied to help achieve these goals in manufacturing. The described use case brings together requirements and settings from real-world projects as described in [38, 93, 111, 129, 130, 137].

The situation faced concerns the introduction of a cyber-physical system *"tracking new orders from an **Enterprise Resource Planning (ERP)** system, sourcing materials and scheduling production resources from a **Manufacturing Execution System (MES)**, and collecting data from all machines, robots, and other equipment used on the shop floor"* [111] (cf. Figure 4 ①). The integration is realized through process orchestration [121] acting as *observer process* (cf. Figure 4 ②) that connects to the different systems, machines, robots, and sensors using interfaces such as OPC UA and collects data through the Siemens S7 communication protocol [111]. The log snippet depicted in Figure 4 ③ shows process event `Fetch` and the associated sensor stream for keyence measurements in the DataStream XES format [93].

Prediction goals in manufacturing (cf. Figure 4 ④) comprise prediction of drifts in the prescribed process behavior [137] and in the process data [136]. In addition, the process outcome [38] such as the quality of the produced part and monitoring the decision rules that drive the routing of the process are typical prediction goals [129, 130]. For example, monitoring a temperature measurement exceeding a certain threshold for three times in a row can be crucial for quality assurance. Taking the prediction of concept drift [137] as another example, the approach combines predicting

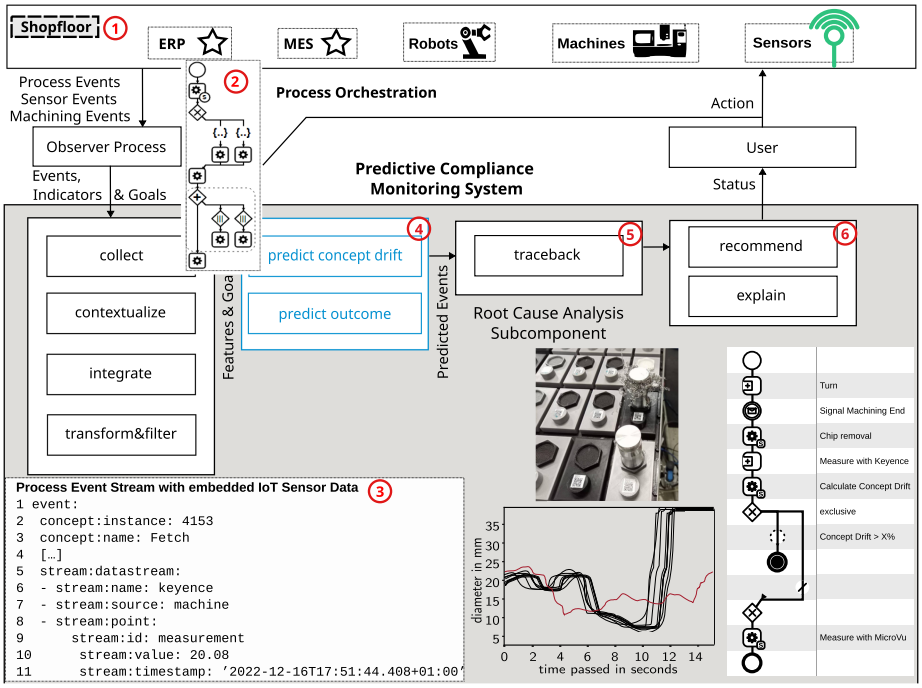

Fig. 4. PCM approach in manufacturing: Predicting and explaining concept drifts.

the next event in a process stream with the prediction of sensor data values, e.g., the data stream representing the diameter measurement of the produced parts. EP methods can predict drifts in the diameter measurements. These drifts can he analyzed and used as basis to recommend (cf. Figure 4 ⑥) an adaptation in the production process. The occurrence of chips on the part results in diameter drifts and, hence, a new task `chip removal` is executed. The approach can be also used for explanation (cf. Figure 4 ⑥), i.e., if a concept drift occurs, the root cause can be traced back to the diameter measurements (cf. Figure 4 ⑤).

Utilizing the PCM approach for describing the manufacturing scenario yields an overview of the entire pipeline: from the data sources to the prediction methods, insights, and actions to be taken. It also shows which challenges are still open in realizing the pipeline in its entirety. First of all, a systematic way of combining EP methods and PPM methods is missing. Moreover, the presentation and explanation of insights to the user is still in its infancy, requiring more research to keep users in and on the loop.

Summary: Using the PCM approach in the manufacturing case underpins the gap between requirements that are met by existing approaches and requirements that are not met, but highly relevant in EP projects, i.e., data quality, lifecycle support, and keeping humans in the loop by, i.a. advanced root cause analysis.

3 Comprehension of EP Methods through the Generalized PCM System

Based on the refined understanding of EP as a component of PCM systems, we take the systems perspective one step further by eliciting requirements for EP methods from existing surveys [122, 161] and iteratively refining them during the assessment of the 260 articles in our selection. Aside from advancing our account of EP, the requirements serve as an assessment scheme for assessing the selection of existing work. Following a concern-based taxonomy of requirements [51], the 16 EP

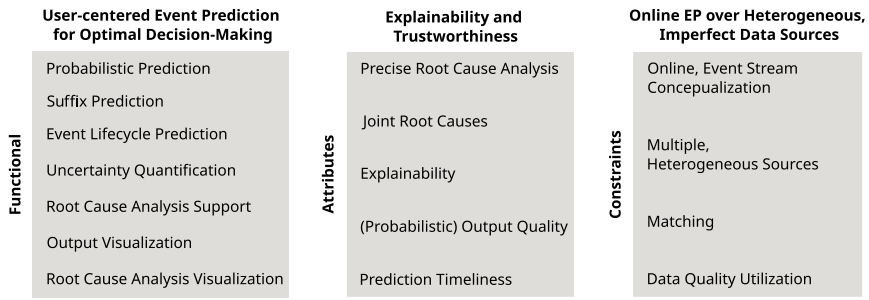

Fig. 5. Requirements engineering result for EP methods in the context of PCM systems categorized by functional requirements, attributes and constraints [51] and annotated by the underlying theme.

method requirements in Figure 5 are classified into seven functional and nine non-functional requirements in the form of five attributes and four constraints. Attributes specify qualities of the method, whereas constraints limit the solution space from which methods can be instantiated. To emphasize how meeting the requirements impacts the PCM system, we identify an underlying theme for each group of requirements. The seven functional requirements enable optimal decision-making [4] and center around the user through visualization [100]. The five attributes represent the need for an explainable and trustworthy PCM system [9, 21]. Lastly, the four constraints specify a PCM system with online EP that continuously predicts future events from heterogeneous, imperfect data sources [122].

Optimal decision making and user-centeredness. Considering that the real-world is complex, chaotic, and imperfect, a deterministic prediction of future events idealizes and keeps important nuances in the future events from the user [120, 122, 161]. Therefore, a *probabilistic prediction* that includes information on the likelihood of the predicted future events actually occurring is required. If we want to enable the user to optimize decisions with respect to a comprehensive anticipation of the future, predicting the next event is not enough. A *suffix prediction* specifies the need for anticipating more events following the next event, e.g., knowing of the next three heavy rainfalls instead of the next one, officials can more efficiently allocate resources. Instead of increasing the anticipated time horizon as suffix prediction does, *event lifecycle prediction* refines the granularity of the anticipated event time horizon. Although events are typically conceptualized as real-world occurrences happening at a certain point of time, each real-world event has a duration and, thus, at least a *start* and an *end* point of time [78, 91]. If we do not predict the lifecycle of events, we are, for example, not aware of the time period affected by the event. For example, our travel plans remain unchanged, if the traffic jam is likely to vanish by the time we arrive at its location.

Uncertainty quantification aims at reflecting the epistemic and aleatoric uncertainty of the method's output [63]. Epistemic uncertainty consists of model and approximation uncertainty and is the result of our lack of knowledge on the true relationships of the monitored system necessary for the design of a perfect prediction method. As such, it is reducible by further research. Aleatoric uncertainty is an irreducible uncertainty that captures the non-deterministic nature of real-world occurrences. Recently, method development acknowledged the existence of both types of uncertainty and incorporated their quantification in its output [154]. Uncertainty quantification can build trust for the prediction method by measuring confidence metrics [161] that accompany probabilistic outputs and improves the subsequent actions, as the user can decide to wait for more certain prediction results that are more likely to occur.

Despite the fact that root cause analysis and visualization are separated from prediction in the PCM system (cf. Figure 3), it is important that the method itself enables these functionalities or

already presents a proof of concept on how these functionalities can be done. Otherwise, it will be more difficult for researchers and practitioners to develop these functionalities given only the method itself. Thus, *root cause analysis support* (RCA support), *output visualization* and *root cause analysis visualization* (RCA visualization) indicate whether the proposed method clearly defines how to achieve root cause analysis or visualization on top of the method respectively or already presents a proof of concept. In particular, the visualization functionalities focus on the user in the PCM system [100].

Explainability and trustworthiness. The first set of attributes are the specific qualities *precise root cause analysis* (precise RCA), *joint root causes* (joint RCs), and *explainability* that are essential for an EP method, since this first set of attributes helps the user to quickly grasp the situation and act accordingly. The method should enable the precise identification of root causes and detect that multiple true root causes for different predicted future events may, in fact, be the same true root cause. As it is out of the scope of this survey to assess the full variety of how explainability can be achieved or conceptualized [9, 21] and there is no widely accepted definition [21, 47, 120], we limit the attribute to coarse-grained distinction. *Black box* models are intransparent, *white box* models are transparent, and a combination of both exhibit both transparent and intransparent components [21]. Taken together, these three qualities determine the accountability of the method to a large extent. If the user can explain the predicted future events as prerequisites to an optimal strategy for acting on the true root causes, the method and the user both should not be held accountable in case something goes wrong.

The second set of attributes are performance attributes that evaluate the EP method itself. [161] presents how the output quality of an EP method is evaluated. Considering the requirement on a *probabilistic output quality* (probabilistic quality), we further require a probabilistic evaluation, e.g., with the Brier score [10] or some other appropriate scoring rule [23]. *Prediction timeliness* pertains to the speed a method is able to predict. Prediction timeliness is particularly important in resource-constrained environments such as vehicle onboard systems [3] or multi-tenant cloud architectures with short customer response times [133]. It determines the lead time, but as a result of concept drift, the training time also has an effect on the lead time. Due to changing environments such as the COVID-19 pandemic concept drift occurs and leads to outdated historic data [122, 123, 161]. Consequently, the prediction model must be kept up-to-date. Having to update prediction models results in the training time of the method playing an additional role for a timely prediction, yet we do not know the significance of this role. Moreover, it is not clear what the optimal time for updating is. Aside from these unknowns, training and prediction time comparisons are only reasonable in a fair benchmark that goes beyond the scope of this survey. Hence, prediction timeliness is not assessed in Section 4.

Online EP over heterogeneous, imperfect sources. To place EP methods into the PCM system, the following four constraints must be met. First, *online, event stream conceptualization* (online conceptualization) is a consequence of the need for monitoring and timely prediction [122, 161]. A method may fail to show how it can be applied to an event stream at run-time. There is a difference between online prediction based on an event stream and offline prediction based on an ex-post dataset, e.g., in terms of efficiently storing and updating already received events from the event stream [25]. Thus, offline EP methods may be unsuitable for monitoring. Second, *multiple, heterogeneous sources* (multiple sources) limit the method solution space to methods that are aware of or actively deal with the challenges of data integration (cf. input component in Section 2.2). Although overcoming these challenges is the responsibility of the input component in a PCM system, the solution may affect the EP method, e.g., by necessitating multiple prediction models [117] or multivariate problem formulations [112].

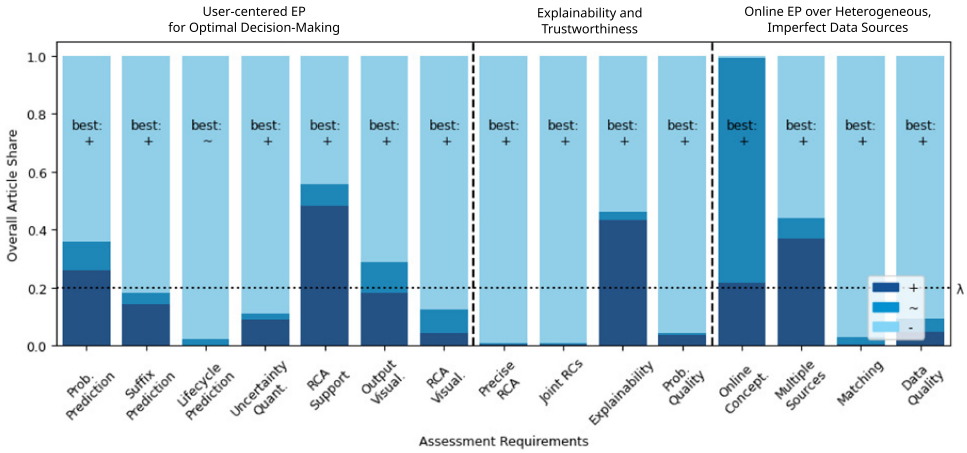

Fig. 6. Assessment results for all articles per requirement.

Third, the challenge of *matching* events from multiple data sources is emphasized as a separate constraint (cf. input component in Section 2.2). The matching strategy for the input should be consistent with the strategy for matching events in prediction goals to predicted future events and the strategy for matching events for evaluation purposes. Multiple matching strategies exist [161], but may have to employ different equivalence notions that underlie the matching of events [122]. For example, matching can be based on equivalent machine identifiers for multiple sensors attached to a machine in a production facility (cf. Section 2.3). Fourth, *data quality utilization* (data quality) limits the solution space to methods that not only acknowledge the existence of data quality issues such as missing data in their design, but also improve the prediction by actively exploiting the existence of data quality issues. For example, [90] develops a self-supervised technique to fully utilize electronic health records that inherently lack a certain type of label.

Conjointly, 15 of the 16 discussed requirements (excluding the *prediction timeliness* requirement) serve as the assessment scheme for existing work on EP methods in Section 4.

4 Assessment of Existing EP Methods

This section presents the analysis of 260 articles from the literature compilation in Figure 2. Each article is analyzed with respect to the 15 requirements contained in the assessment scheme introduced in Section 3, resulting in an overview for each requirement depicted in Figure 6. The goal of the analysis is to understand the current status of EP methods from a systems perspective (cf. Sections 1 and 3) and to pinpoint open research challenges with their corresponding research directions (cf. Section 5).

The assessment of the requirements uses an ordinal scale: + means that the requirement is met, ~ means that the requirement is partly met and − means that the requirement is not met. In the following, we illustrate assessment criteria for one selected requirement of each of the 3 requirement groups (functional, attributes, and constraints). The selection is made due to space limitations (for the remaining requirements, see Section C in the appendix) and aims at conveying the general idea of the assessment criteria.

(1) **Probabilistic Prediction**: + corresponds to directly providing probabilities of occurrence for predicted events. Ref. [106], for example, propose a multi-task spatio-temporal event forecasting framework to predict the probability of large-scale societal events like civil unrest through news articles. By explicitly predicting the probability P_i^k of the ith-societal event

occurring in city k given the news articles X_i^k published H days before the event, [106] meets the requirement. ∼ corresponds to an EP method that does not directly provide occurrence probabilities of predicted events, but can easily be extended to do so. For example, [87] estimates a probability density function for event occurrences across time and space given past event occurrences to assess the crime risk in a monitored region. Evidently, the density function does not constitute an event probability yet, but we can easily compute a probability for a concrete event. Lastly, − corresponds to a predicted event without an occurrence probability and no easy extension. Ref. [22], for example, predict the remaining time that an ongoing process instance takes to complete. Consider a loan application process in which the bank is interested in the remaining times of ongoing applications. For a given application, an LSTM model ensemble is learned from similar past applications. Thus, the approach in [22] is not easily extensible to predict the remaining time's probability.

(2) **Explainability**: + corresponds to an approach being "white box" (cf. Section 3). For example, [159] propose a temporal extension to association rules. ∼ corresponds to "white box" and "black box". Ref. [6], for example, combine transition systems and LSTM to predict remaining times of ongoing process instances. Lastly, − corresponds to "black box". For example, [88] propose a multi-channel fusion LSTM to predict medical events.

(3) **Online Conceptualization**: + corresponds to an EP method that is defined on an event stream such that the model and predictions are continuously updated and the updating times are known. For example, [70] proposes EP within the PCM system (cf. Section 2.2) to predict large-scale societal events. Hence, periodic data ingest through Apache Spark Jobs is coupled with daily predictions, a lead time of 5 days (cf. Section 2.1), and a weekly update of all prediction models. Next, ∼ corresponds to an EP method that is defined on an event stream such that the model and predictions are continuously updated, but the updating times are not known. For example, [127] proposes to predict system performance and errors for a large computing cluster. Their proposal takes the 5 minute interval for new events and performance measurements into account, but the updating of the prediction models is unknown. Lastly, − corresponds to an EP method that is defined on a set of historical events with a single training and prediction for evaluation. For example, [132] proposes to predict security-breaking attacks from past attacks without continuous updates.

4.1 Overall Results

Detailed assessment results are reported in our online repository.[1] The results for each requirement in Figure 6 give a nuanced view on research progress in the field of EP methods, as the overall article shares vary across the requirements. By applying a threshold of $\lambda = 20\ \%$ (corresponds to 52 articles) on the ∼ assessment result, we obtain two sets of requirements: One set that contains requirements with article shares above the threshold (i) and another set that contains requirements below the threshold (ii). In set (i), existing work has proposed methods that meet the requirements *probabilistic prediction*, *RCA support*, *output visualization*, *explainability*, *online conceptualization,* and *multiple sources*. The majority of methods meeting those requirements emphasize the need for monitoring, causality, explanation, and visualization and acknowledge that relevant data is typically distributed over multiple data sources. In set (ii), existing work misses requirements *suffix prediction*, *lifecycle prediction*, *uncertainty quantification*, *RCA visualization*, *precise RCA*, *joint RCs*, *probabilistic quality*, *matching* and *data quality* to a large extent. Despite the very low shares of articles meeting requirements *precise RCA*, *joint RCs*, and *matching*, still some

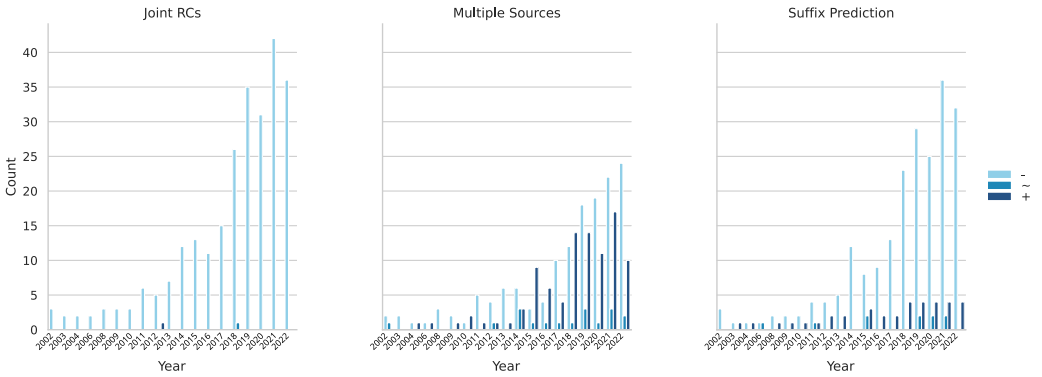

Fig. 7. Meeting criteria for requirements *joint RC*, *multiple sources*, and *suffix prediction* (+, ~, -) over time.

approaches fully meet them (+). *Lifecycle prediction* is the only requirement with a ~ as the best assessment.

The distribution of the + assessment count for each article is presented as Figure 3 in the supp. material[1]. It reveals that the majority of articles focus on one to three of the assessment requirements, i.e., the design of EP methods is typically aimed towards a few, particular properties. Although a method design with a clear focus is beneficial during method design, the result leaves the problem of system design to the practitioner. If an article had the best assessment of meeting all requirements, then the boxplot in the appendix would show a count of 15. As the highest count is an 8, the boxplot shows that the mostly positive best assessment in the requirements view (cf. Figure 6) is due to various articles, i.e., there is no single article that is close to the best possible assessment.

We also systematically analyze the dynamics of EP research based on different timeline-based charts. The distribution of all selected publications over time shows the ongoing and still increasing research output in EP and PPM. We plot the development of the publications regarding their fulfillment of the 15 analyzed requirements per requirement over time. Figure 7 depicts the plots for requirements *joint RC*, *multiple sources*, and *suffix prediction* as representatives for different patterns/trends.

Joint RCs (left side) can be seen as representative distribution for requirements *lifecycle prediction*, *precise RCA*, *prob. quality*, and *matching*. For all of these requirements, $\lambda < 0.2$ and even more precisely, $\lambda \leq 0.04$ holds and the requirement is met (+) or partly met (~) only selectively. *Multiple sources* (middle part) can be seen as representative distribution for requirements *prob. prediction*, *RCA support*, *output visualization*, *explainability*, and *online concept.*. For all of these requirements, $\lambda > 0.2$ holds and the distribution of criteria met (+), partly met (~), and not met (-) is evenly distributed with steady increase. *Suffix prediction* (right part) can be seen as representative distribution for requirements *uncertainty quant.*, *RCA visual.*, and *data quality* for which $0.04 < \lambda < 0.2$ holds. Here, the number of publications with criterion not met (-) increases whereas the other two assessments as + and ~ show a plateau distribution, i.e., the number of publications remains roughly the same over time. The distribution of publications categorized by *model type* (cf. Section B in the suppl. material[1]) over the considered timeline reveals two additional interesting observations (cf. Figure 4 in the supp. material). First, publications featuring *inferential* prediction models that are "black box" follow the overall increasing trend. Interestingly, for *hybrid* models that combine "black" and "white box", it seems that a peak was reached in 2019, i.e., the number of publications featuring hybrid models has been decreasing since 2019.

In the following, we aim to support a more holistic development of EP methods meeting more requirements by illustrating a method's properties[3] for each requirement with an example from existing work on each requirement. The examples are selected based on how straightforward, concise and representative their properties for meeting the requirements are and how well they reflect the diversity of prediction goals and application domains. The illustration shows how key design aspects are responsible for properties of the method that are, in turn, responsible for meeting the requirement. The presentation of examples is structured along the three themes underlying requirements (cf. Section 3). It is accompanied by an analysis of how the requirement may be related to other requirements and a discussion of the respective approaches.

4.2 User-centered EP for Optimal Decision-Making

Seven requirements (cf. Figure 6) center around the user and the support for optimal decision-making given the status of the PCM system (cf. Section 3). While probabilistic, detailed, and certain predictions across the whole future time horizon are crucial for optimal decision-making, advanced visualization functionalities are insightful to the user.

Probabilistic prediction. [44] predicts the time and event failure type of high-performance computing cluster nodes by extracting event graphs from system logs. To that end, [44] compute the confidence of the correlation between events, the average time delay between events and its standard deviation. Figure 8 shows a small, exemplary event graph with three events y_0, y_1 and $\hat{y}$. For a *probabilistic prediction*, the EP method in [44] predicts the probability of event $\hat{y}$ that follows event y_0 and y_1 with Bayes law to be 50% * 50% = 25% and the time of event $\hat{y}$ to be 25 minutes +/- (2 * 4 minute + 2 * 3 minute) = 14 minutes with 95% probability after y_0 has occurred.

Leveraging the event graph, Bayes law and the assumption of normally distributed time delays leads to a straightforward, probabilistic conceptualization of EP. For the same reasons, the prediction can be traced back to one or a set of root causes in the event graph such that the method meets the requirements of *RCA support*. Furthermore, the simplicity of the method helps domain experts to understand its output. Yet, it may fail to perform well when facing complex dependencies between events in multiple dimensions, e.g., a patient's clinical history and future treatment outcome. This indicates a tradeoff between simplicity of the probabilistic conceptualization and its applicability to complex monitored systems, their application domains and corresponding prediction goals.

Methods with a probabilistic prediction all conceptualize the method with conditional probabilities, Bayesian or Markovian settings, but differ in their degree of statistical rigor. Statistical rigor captures the number of assumptions on the distribution of the input data, how restraining the distribution assumptions are and how formal the derivation of the EP method is given those assumptions. For example, [7] exhibits a high degree of statistical rigor, as it features a high number of assumptions on the distribution of the data that restrain it to a large degree (e.g., the input data is the result of a Gaussian autoregressive process of the second order). Then, the whole EP method is formally derived from those assumptions. [44] exhibit a medium degree of statistical rigor, as they assume normally distributed time delays and apply Bayes law, but do not formally derive the overall EP method. [113] exhibits a low degree of statistical rigor, as they explicitly aim at designing the EP method to have no assumptions and do not formally derive the method.

Suffix prediction. [69] predict the full sequence of future events for an ongoing manufacturing cycle. Thus, [69] predicts all the events that are likely to occur until the production is complete.

[3]Some properties are illustrated with definitions or equations from the original work cited at the beginning of the example, i.e., if the illustration states a definition or equation, then it is taken from the original work and its key part for illustration is extracted and adapted for consistent presentation w.r.t to the survey's notation.

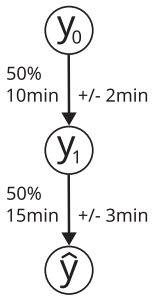

Fig. 8. An event graph for events y_0, y_1, and $\hat{y}$ that follow each other in system logs of a supercomputing cluster node [44]. The probability of following, the average time delay and the standard deviation of time delay is added for each relation between events.

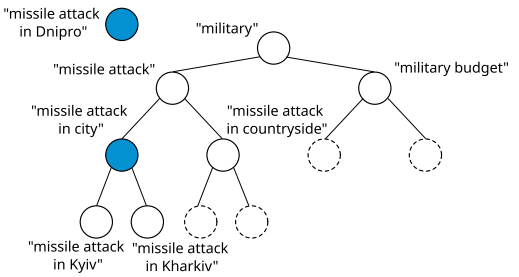

Fig. 9. A hierarchy of most general to most specific observed events [114]. On receiving a current event of a missile attack in Dnipro, the most specific event in the hierarchy that semantically matches the current event is used for prediction.

The prediction goal is a set of key performance indicators such as product quality. [69] assume that the future progress of the ongoing cycle s with length n can be similar to the progress of k historical cycles $l_1, \ldots, l_k$ after n events of each historical cycle l_i have been observed. Based on this assumption, [69] apply k-Nearest-Neighbor clustering with Euclidean distance on the n-length prefix of historical cycles and the ongoing cycle resulting in k suffixes for the ongoing trace. Then, [69] apply a support vector machine classifier on the k suffixes to predict the key performance indicator. At last, the k predicted key performance indicators are aggregated into a single indicator. Although this indirect basic technique can show the operations manager in the production facility additional information on how the current cycle will progress, this method likely fails to accurately predict for a production line with hundreds of thousands of different cycle variations. In contrast to [69], many other methods such as [84] or [34] predict the suffix by repeatedly applying a next EP method. Although this can be done for any EP method, the major challenge of looping through events that are likely has to be tackled by methods repeatedly applying the EP method for *suffix prediction*.

Lifecycle prediction. [84] predict the next events of business processes by learning a **recurrent neural network** (**RNN**) that takes multiple event attributes into account. Accordingly, the EP problem is stated for an ongoing process instance $y = \langle (e_1, a_1^1, \ldots, a_1^m), \ldots, (e_n, a_n^1, \ldots, a_n^m) \rangle$ with event activities $e_1, \ldots, e_n$ and event attributes $a_1^1, \ldots, a_n^m$ as predicting the next event $(e_{n+1}, a_{n+1}^1, \ldots, a_{n+1}^m)$ of the ongoing process instance s. Hence, [84] can predict lifecycles of events, since the lifecycle of an event is an attribute of the event. Nevertheless, lifecycle carries special semantics for EP. For example, an *abort* lifecycle event for the delivery activity of a logistics company signifies that the *end* will not occur at all and, therefore, the delivery will fail [122]. Also, the lifecycle determines the duration of the activity recorded through multiple events. For instance, an activity paused once for 5 minutes has a duration of 5 minutes lower than the time between its *start* and *end*. For prediction goals that are evaluated on the duration, e.g., efficiency of the resource executing the activity, the special semantics of the lifecycle has to be taken into account. Consequently, the prediction method should not stop at predicting the lifecycle as an attribute, but incorporate its semantics.

Similarly, the other proposals partly meeting lifecycle prediction conceptualize EP with multiple, unspecified event attributes and do not incorporate lifecycle semantics. If the data source for the evaluation does not provide lifecycle information, this should not be taken as the reason for excluding it in a method design. Exclusion would not allow the method to be generally applied to

data sources with event lifecycle information. Due to the importance of lifecycle information for prediction, researchers should recommend the additional recording of this information to a data source, if the data source does not provide lifecycle information.

Uncertainty quantification. [158] predict the time-to-event for treatment outcomes of patients in clinics by extracting features with RNNs and learning a **Parametric Predictive Gaussian Process** (**PPGP**) Regressor in a survival analysis setting. The combination of the PPGP Regressor and survival analysis results in the survival function:

$$S(y_i|\mathbf{h}_i) = 1 - F\left(\frac{y_i - \mu_{\mathbf{f}}(\mathbf{h}_i)}{\sigma_{\mathbf{f}}(\mathbf{h}_i) + \sigma_{\text{obs}}}\right),$$

where S is a survival function, y_i the logarithm of the time-to-event for patient i, $\mathbf{h}_i$ the latent representation extracted by the RNN, F the normal cumulative distribution function, $\mathbf{f} \in \mathbb{R}^n$ a vector of Gaussian process function values, $\sigma^2_{\mathbf{f}}(\mathbf{h}_i)$ the input-dependent predictive variance and σ^2_{obs} the input-independent observational noise for the logarithmic time-to-event y_i. The corresponding objective that [158] propose puts more weight on the input-dependent predictive variance $\sigma^2_{\mathbf{f}}(\mathbf{h}_i)$ than earlier proposals that combine Gaussian Processes and survival analysis. Hence, the formulation of survival analysis with PPGP Regressors and the corresponding uncertainty-aware objective quantifies predictive variance, i.e., epistemic uncertainty, and observational noise, i.e., aleatoric uncertainty. As well, the formulation also strengthens the positive relationship between predictive variance and prediction error. Consequently, the less confident the uncertainty-aware EP method is, the less accurate its predictions are.

In addition to quantifying the uncertainty of the prediction, appropriately integrating it to the objective of the EP method yields a desirable relationship between uncertainty and prediction quality. The information that more confident predictions correspond to more accurate predictions is valuable to practitioners [63]. This information likely results in more trust towards the method. By further differentiating between epistemic and aleatoric uncertainty, the quantified uncertainty also shows whether a currently low confidence could be improved with more data.

Existing proposals mostly quantify uncertainty by means of a probabilistic conceptualization that allows to inherently estimate uncertainty as some form of variance in the method and, thus, focuses on the method, e.g., [158]. Depending on the degree of statistical rigor (cf. *probabilistic prediction*), these estimates are more or less justified by statistical theory. Another option is to extrinsically quantify uncertainty by approximating it using, e.g., characteristics of the input data. [26] propose an uncertainty metric that also considers how many activities of a business process instance have been observed and what time has elapsed since the last activity has been observed. The rationale for their extrinsic uncertainty quantification is that more and more recent data on the monitored system relate to a more certain prediction.

RCA support, output visualization and RCA visualization. While [44] apply Bayes law on descriptive statistics of the data without a statistical analysis formulation, [81] combine decision rules with Bayesian analysis into **Bayesian Rule Lists** (**BRL**):

if antecedent$_1$ **then** $y \sim \text{Multinomial}(\boldsymbol{\Theta}_1)$, $\boldsymbol{\Theta}_1 \sim \text{Dirichlet}(\boldsymbol{\Phi} + \mathbf{N}_1)$

$\vdots$

else if antecedent$_m$ **then** $y \sim \text{Multinomial}(\boldsymbol{\Theta}_m)$, $\boldsymbol{\Theta}_m \sim \text{Dirichlet}(\boldsymbol{\Phi} + \mathbf{N}_m)$

else $y \sim \text{Multinomial}(\boldsymbol{\Theta}_0)$, $\boldsymbol{\Theta}_0 \sim \text{Dirichlet}(\boldsymbol{\Phi} + \mathbf{N}_0)$

where antecedents are well-formed formulas with c_i, $i \in \{1, \ldots, m\}$ atomic propositions over the feature vector that do not overlap, y is the prediction goal value, Θ_i are parameters of the respective distribution, Θ_0 a default rule parameter for observations that do not satisfy any of the

antecedents, Φ a prior of the prediction goal values and $\mathbf{N}_i$ the respective observation counts satisfying antecedent i. By limiting the parameters m and c_i to small integers as well as deriving a mean point estimate and credible interval for the prediction goal value y, BRLs represent a simple, interpretable prediction model. Potential root causes are represented as antecedents of the BRL (*RCA support*). The prediction output and the RCA are visualized by showing the BRL and traversing its rules from top to bottom given an observation. [81] compares this BRL to a medical score that is frequently used to predict the stroke risk by physicians. They identify antecedents with known root causes or risk factors and conclude that the BRL is as interpretable as the medical score, while improving the accuracy.

Methods meeting all three requirements *RCA support*, *output visualization*, and *RCA visualization* are designed in their functionality to support the user in understanding the method and its prediction. A simple prediction model as in [44, 81] underlying the EP method supports the user in tracing the prediction back to its root causes. This meaning of simplicity is captured by the "white box" assessment of *explainability*, but it is not the only sufficient condition for meeting *RCA support*. For instance, a hybrid basic approach can still support RCA in spite of a "black box" model due to the "white box" part of the model [86]. Due to a lack of a widely accepted definition on *explainability* (cf. Section 3), it is not clear whether further relationships with root causes exist. Considering visualization, some methods present a straightforward way of visualizing the output or RCA as in [81], while others need specifically designed visualization techniques, e.g., neural networks.

4.3 Explainability and Trustworthiness

Four requirements (cf. Figure 6) aim towards explainable and trustworthy EP through appropriate quality metrics and root-cause analysis.

Precise RCA, joint RCs and explainability. [114] predict future events that can be caused by a current news event. To that end, they choose an aggregate semantic event representation for an event $e = (P, O_1, \ldots, O_4, t)$. Here, P is a state or temporal action exhibited by the event's objects, O_1 a set of users, O_2 a set of objects on which the action P was performed, O_3 a set of instruments utilized by the action, O_4 a set of locations and T a timestamp. Given a set of historic events $\{\langle e_1, g(e_1)\rangle, \ldots, \langle e_n, g(e_n)\rangle\}$ for an unknown causality function g mapping a historic cause event to its effects, [114] aim to approximate g by learning a causality graph of events. The causality graph is learned through existing ontologies on actions such as VerbNet [71] and on objects such as the LinkedData ontology [14].

Using existing ontologies, the method can evaluate the precision of identified root causes (*precise RCA*). For identifying *joint RCs*, the method proposes a generalization event by generalizing over its actions and objects. The result is a hierarchy that detects that similar root causes are in fact joint root causes. Figure 9, for example, depicts a generalization hierarchy of observed events that identifies similar root causes as joint, more general root causes. The prediction is computed solely on the causality graph through matching the current event with an existing event in the causality graph (denoted in green in Figure 9) and applying a prediction rule associated with the event. Thus, a user can traverse the predicted event back to its historic event. Finally, the method employs a "white box" model (*explainability*).

Despite their existence, ontologies are rarely used in EP. With the help of ontologies, the method can establish a concept of causality and meaning. Moreover, EP methods rarely exploit that events are related by a semantic generalization relationship. However, other semantic relationships between events are exploited by recent methods, e.g., the disease co-occurrence [89] or compatible related treatment outcomes relationships [46].

Probabilistic Quality. [151] predict crime frequencies in the spatio-temporal domain by combining variational autoencoders and sequence generative neural networks. Hence, a predicted crime frequency is $\hat{y}^{(t)}_{i,j,k}$ for the geographical grid indexes i, j, crime type k in the time slot t. To evaluate these probabilistic predictions, [151] first normalize the true and predicted crime frequencies:

$$p^{(t)}_{i,j,k} = \frac{y^{(t)}_{i,j,k}}{\sum_{i=1}^{M}\sum_{j=1}^{N} y^{(t)}_{i,j,k}}, \quad \hat{p}^{(t)}_{i,j,k} = \frac{\hat{y}^{(t)}_{i,j,k}}{\sum_{i=1}^{M}\sum_{j=1}^{N} \hat{y}^{(t)}_{i,j,k}}$$

and then compute the Jensen-Shannon divergence with Laplacian smoothing:

$$D_{\mathrm{JS}}(P, \hat{P}) = \frac{1}{2} D_{\mathrm{KL}}(P \parallel Q) + \frac{1}{2} D_{\mathrm{KL}}(\hat{P} \parallel Q)$$

where $Q = \frac{1}{2}(P + \hat{P})$ is the average of true and predicted probability. Next to the Jensen–Shannon divergence, there exist at least 200 further scores for evaluating probability estimates [23]. Thus, analysing its properties with respect to EP and proposing a standardized set of scores goes beyond the scope of this survey. For this survey, the important point is that the evaluation is done using a metric that scores probability estimates, as is the case for [151].

For evaluating probabilistic predictions, the predictions must come with probability estimates. Thus, meeting the requirement of *probabilistic quality* implies meeting *probabilistic prediction*. Yet, these two requirements are not equivalent, because many articles do not report results of scoring probability estimates. While roughly 25% of the 260 articles propose methods meeting *probabilistic prediction*, less than 5% evaluate the predicted probabilities with an appropriate score (cf. Figure 6).

4.4 Online EP over Heterogeneous, Imperfect Data Sources

Four requirements (cf. Figure 6) represent the need for continuous EP given heterogeneous data sources with dynamically-changing data quality.

Online conceptualization. [117] predict civil unrest events across ten Latin American countries through a continuous, automated system. The system runs 24/7 and constantly processes events coming from various open data sources such as social media. Consequently, the EP method consisting of an ensemble of methods is proposed to continuously learn from daily data. For example, a logistic regression model is learned using daily tweet counts while a rule-based method detects keywords contained in social media posts. The system has a throughput of 200–2,000 messages/sec, predicts roughly 40 events/day and can ingest up to 15 GB of messages on a given day. Therefore, it features an *online conceptualization* and can monitor the ten Latin American countries for social unrest events. Given its daily cycle and the ensemble of prediction methods, a daily, full retraining of the prediction models is feasible. However, a full retraining may become infeasible for EP methods in light of updating cycles not longer than a few minutes. For this reason, [123] investigate strategies for updating the prediction model that go beyond full retraining: `Do nothing`, i.e., no update; `retraining with no hyperparameter optimization`, i.e., a lightweight retraining; `full retraining`, i.e., train on all available events; and `incremental update`, i.e., applying incremental learning algorithms. The results show that `full retrain` and `incremental update` are the best strategies for predicting business process events.

Still, not all prediction models allow for incremental updates through incremental learning algorithms. Thus, [102] investigates six strategies to handle updates to the prediction model by means of data selection strategies. First, the `baseline` strategy means no update to the training data. Second, the `cumulative` strategy means "update training data on every new event". Third, the `non-cumulative` strategy means "update training data on every new event by keeping the k most

recent events". The `ensemble` strategy means "non-cumulative and keeping all models for ensemble prediction". Next, the `sampling` strategy means "update training data by sampling all available events". Finally, the `drift` strategy is non-cumulative with drift detection such that only training data after a drift is selected. The results show that the `ensemble` strategy performs best for predicting business process events.

Despite these pointers to suitable strategies for updating prediction models, a best strategy for all use cases cannot be determined due to a case-dependent tradeoff between the resources required to improve the quality of prediction through updating and its benefits.

Multiple sources. [2] predict the time-to-event for future failures of engineering system devices. Sensor data $X_p \in \mathbb{R}^{d \times c_p}$ from d sensors for each device p observed until time c_p and event data $y_p \in \{0, 1\}$ is used to predict the time-to-event T for the future failure $y_p = 1$ of device p. For prediction, [2] propose a multi-task learning framework based on neural networks with a task p for each device. Hence, p data sources are used in the EP method (*multiple sources*) and the integration of these data sources is implicitly learned by the neural network.

We can distinguish two approaches to data integration. The *implicit* approach to data integration combines the data integration with the learning of a prediction model in the EP method such that the data integration is tightly coupled with the EP method. For example, the data integration of the p devices is implicitly learned by the neural network in [2]. The *explicit* approach to data integration separates data integration from the EP method such that both of them become interchangeable. To that end, the explicit approach implements necessary functionality for data integration (cf. input component in Section 2.2) before the EP method is applied to the data. [117] not only illustrates the *online conceptualization*, but also the explicit approach to data integration. Many different sources such as NASA satellite meteorological data, Bloomberg financial news and X's (formerly known as Twitter) public API are used. Each source is ingested by a specialized routine to convert the data input to JSON and add identifiers. Before the EP method is applied, the JSON input is enriched through, e.g., geocoding, data normalization, and entity extraction. Consequently, the EP method does not need to learn how it can integrate the data. In comparison, the implicit approach intertwines the input component of a PCM system (cf. Section 2.2) with the prediction component, whereas the explicit approach keeps them separated. The former has the advantage that it does not need to design data integration functionalities, but may experience worse performance for heterogeneous data sources and consuming more resources for training. The latter has the advantage, that it has direct control and knowledge on the specific data integration routines, but comes at the cost of more design and maintenance effort.

Matching. [61] predict a future event in a time-series sequence, e.g., a future anomalous watthour meter clock event of the State Grid of China in weekly sensor readings. To that end, a future event $y_{t+1} \in \mathbb{Z}$ is predicted based on historic events y_t and T segments $\mathbf{X}_t \in \mathbb{R}^{\tau \times d}$ of the time-series sequence $\mathbf{X} \in \mathbb{R}^{N \times d}$ with length τ:

$$\langle \mathbf{X}_{1:T}, y_{1:T} \rangle = \{(\mathbf{X}_1, y_1), \ldots, (\mathbf{X}_T, y_T)\}$$

Before feeding the historic data $\langle \mathbf{X}_{1:T}, y_{1:T} \rangle$ into a neural network, the proposed EP method recognizes representative time-series patterns called *states* and denoted as $\Theta_v \in \mathbb{R}^{\tau \times d}$ for segments of the data (cf. Figure 10). Hence, redundant or very similar time-series patterns from multiples sources are recognized by the same state (*matching*). [61] is the only method proposed that features an explicit matching mechanism for data coming from multiple sources. Other proposals for matching [1, 60, 70, 79, 82, 117, 150, 160] are not explicitly including the mechanism in their method.

Data quality. [49] predict the location of event subtypes. Therefore, instead of predicting the air pollution of cities as future events, the respective air pollutant subtype (Carbon monoxide,

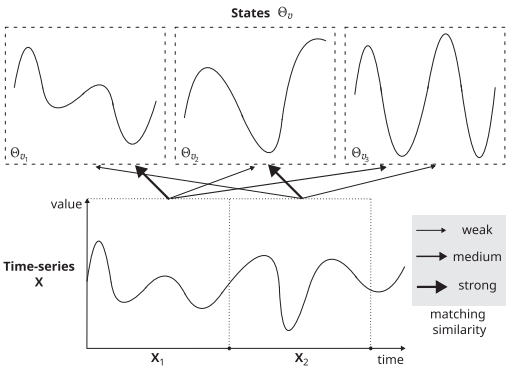

Fig. 10. Time-series $\mathbf{X}$ is segmented into $\mathbf{X}_1$ and $\mathbf{X}_2$. Each time-series segment is recognized with states Θ_{v_1}, Θ_{v_2} and Θ_{v_3} [61]. Recognized states and the similarities are used in the EP method proposed in [61] instead of the original time-series.

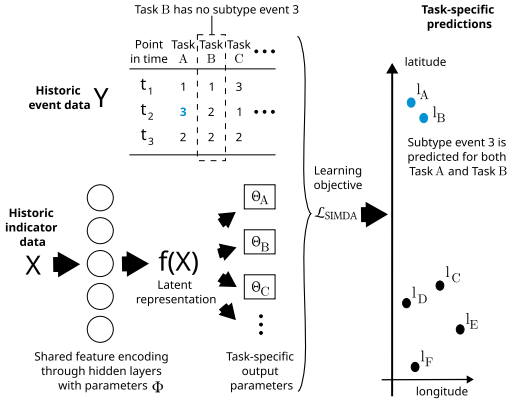

Fig. 11. SIMDA Framework [49]. Despite no recorded subtype event 3 in the historic event data Y for task B, SIMDA allows to predict subtype event 3 in task B that corresponds to location l_B, since location l_A is geographically close to location l_A and location l_A has recorded subtype event 3 in the historic event data of task A.

nitrogen dioxide, fine dust pollution, etc.) is predicted. The predicted event subtype allows for more fine-grained actions, e.g., allocating appropriate resources per air pollutant subtype. However, incomplete, historical data for event subtypes in many locations becomes a major data quality challenge for predicting event subtypes for all locations.

Figure 11 depicts the data quality challenge for predicting event subtypes and the proposed method, **Spatial Multi-task Deep leArning (SIMDA)**, to overcome the challenge. Each depicted prediction task A, B and C corresponds to a geographical location $L = \{l_A, l_B, l_C, \ldots\}$. Historic event data $Y_{l,t} \in S = \{1, 2, 3, \ldots\}$ and d-dimensional, historic indicator data $X_{l,t} \in \mathbb{R}^{1\times d}$ for date t is available for each location $l \in L$. However, for the location l_B corresponding to task B, no event subtype $s = 3$ was recorded, e.g., a city in the past has not recorded fine dust pollution (cf. Figure 11). Hence, the specific model for task B will not predict event subtype $\hat{y}_{l_B,\tau} = 3$ at location l_B for $\tau = t + p, p > 0$, although the geographically close location l_A in task A shows that event subtype $s = 3$ has occurred. Since geographical heterogeneity across all locations does not allow for a single task approach, learning from other locations such as l_A for location l_B has to be achieved differently. Here, the first law of geography stating that geographically closer locations will be more similar to each other than farther locations [27] is exploited. Reference [49] state that two close locations $i, j \in L$ have similar conditional probabilities for an event subtype s:

$$P(Y_{i,t} = s|X_{i,t}) \approx P(Y_{j,t} = s|X_{j,t})$$

This relationship is exploited to formulate a SIMDA objective that enforces event subtype patterns to be similar for geographically close tasks:

$$\mathcal{L}_{\text{SIMDA}} = \mathcal{L}_d(\Phi, \Theta) + \frac{\beta}{2} \sum_{l}^{L} \sum_{s,r}^{S^2} \left\| f(X_l)(\Theta_{l,s} - \Theta_{l,r})^T - \frac{1}{N_l} \sum_{c}^{L} adj(l,c) f(X_c)(\Theta_{c,s} - \Theta_{c,r})^T \right\|_2^2$$

where $\mathcal{L}_d(\Phi, \Theta)$ is the general multi-task deep learning objective with d dimensions, Φ the weight parameters of the shared hidden layer, Θ the weights of the task-specific output layer, β a hyperparameter, $f(\cdot)$ the computation of the shared hidden layers, $\Theta_{l,s}$ the weights of task

l for predicting event subtype s, $N_l = \sum_c^L adj(l, c)$ the normalization term for location l, and $adj(\cdot, \cdot)$ some physical distance function. Hence, [49] propose to add a regularization term to the general multi-task deep learning objective $\mathcal{L}_d(\Phi, \Theta)$ that enforces geographically close locations i, j to have similar event subtype patterns $X_i(\Theta_{i,s} - \Theta_{i,r})^T \approx X_j(\Theta_{j,s} - \Theta_{j,r})^T$. For the example as depicted in Figure 11, this allows SIMDA to predict fine dust pollution $s = 3$ for location l_B through task B, as the geographically close location l_A exhibits a lot of fine dust pollution.

Analogous to [49], methods that actively utilize data quality issues exploit the following semantic similarity relationships between components of the monitored system (close locations, comparable patients, etc.): Semantic similarity based on proximity [48, 49, 163, 164], hierarchy of geographical regions [163, 164], logging granularity of cyber system host sources [157], patient disease histories [90, 113, 134], hierarchy of medical codes [90], hierarchy of mobile health study participants [31], individual units of engineering systems [30], feature missingness patterns [149] and structural tweet content [162]. Exploiting semantic similarity relationships aims to improve the prediction while keeping the component-level granularity in the prediction model (e.g., a task for each location). In contrast to preprocessing techniques such as imputation for missingness, these proposals make the method aware of the data quality and base the awareness on the semantic similarity relationships.

Since data quality issues should lead to more uncertain predictions, this requirement is related to *uncertainty quantification*. For example, [134] propose a method that abstains from predicting in case of high uncertainty of the prediction due to missingness of data. Hence, the missingness of data has a direct impact on the prediction in their proposed method. All in all, the link between data quality and uncertainty can be used to guide the prediction (e.g., no prediction in case of high uncertainty due to data quality issues) or the data quality issue can be alleviated by exploiting semantic relationships leading to more certain predictions.

4.5 Summary

Overall, EP methods broadly support the requirements for building a PCM system concerning the input and output of the prediction method as well as its evaluation and understanding (cf. Figure 5). Looking more closely at the requirements support, the following observations can be made:

OB1 EP supports a selection of requirements through a combination of the prediction goal and EP method.

OB2 The number of approaches for EP has been and is steadily increasing.

OB3 None of the approaches supports all requirements; most of the approaches support one to three requirements and at most 8 out of 15 requirements.

OB4 Approaches supporting requirements with $\lambda \leq 0.04$, $0.04 < \lambda < 0.2$ and $\lambda > 0.2$ can be distinguished with distribution patterns quite clearly, i.e., selective occurrence, plateau, and steady increase. An interpretation here is that requirements that are already addressed, remain being addressed with increasing interest. Requirements that have not been addressed or only in a limited way remain being addressed with low intensity.

OB5 None of the approaches supports event life cycle prediction.

OB6 Although the support of root cause analysis (*RCA support*) is fully met by the highest number of articles in the article selection, other root cause related requirements are only supported by very few approaches, i.e., *RCA visualization*, *precise RCA*, and *joint RC*. Hence it seems that the importance of root cause analysis is acknowledged and addressed, but advanced root cause related requirements remain still open.

OB7 Although the existence of multiple event sources is supported by a quite high number of approaches, other data-related requirements such as *matching* and *data quality* are only addressed by very few approaches. Hence, it seems that the existence of multiple and possibly heterogeneous

data sources is acknowledged and addressed, but further aspects connected to the quality of data remain still open.

OB8 Predicted events without an occurrence probability (*probabilistic prediction*) and confidence (*uncertainty quantification*) can give a false sense of urgency to act on low probability or confidence events. It remains to be open how the various EP methods can be advanced to account for probabilities and confidences.

OB9 Despite the importance of visualizing predicted events and root causes to the user, *output visualization* and *RCA visualization* are only (partly) met by roughly 80 articles and 30 articles respectively. As EP is formulated through a heterogeneous range of conceptualizations (cf. taxonomy in Figure 2 of the supp. material[1]), it is not straightforward to deduce visualizations from the article's EP method and evaluation.

OB10 Suitable evaluation metrics for an EP method are a condition for understanding the method's qualities. Nevertheless, *probabilistic quality* is only successfully addressed by less than 5% of the articles.

Overall, OB1 to OB10 point to an ongoing research effort in EP with a focus on selected requirements which results in a rather static and fragmented research landscape. Hence, the integrative effort to combine EP methods from different areas and application domains as well as taking a system perspective offers the opportunity to consider EP under a broader set of requirements. Observations OB1 and OB2 are understood as confirming our EP definition (cf. Section 2) and the importance of EPs. Based on OB3 to OB10, we formulate challenges and research directions in Section 5.

5 Open Challenges and Research Directions

In order to counteract the current fragmentation in EP research, we derive open challenges from the assessment in Section 4, in particular observations OB3 to OB10.

C1 - Holistic EP Method

Challenge: A direct consequence of OB3 is that no EP method simultaneously meets all requirements. Hence, selecting a single EP method leaves us with the necessity of trading off various sets of requirements with one another. In particular, some requirements like *joint RCs* remain a blind spot for EP research, while other requirements like *suffix prediction* are still underrepresented (cf. OB4). The challenge is a development of future holistic EP methods that support more or all requirements and, thus, represent the manifold nature of EP. Moreover, the value of EP is only determined by accuracy of predicted events, as indicated by the prominent focus of EP research on accuracy. However, many more factors determine the value of EP in the PCM system as the assessment and related surveys emphasize [122, 161]. Aside from accuracy, the event resolution (e.g., hospital vs. ward utilization), likelihood (i.e., *probabilistic prediction* in Section 4.2), confidence (i.e., event uncertainty), prediction timeliness, event intensity (e.g., how bad is the occurrence), explainable events (e.g., root causes can be determined) and further quality dimensions that result from the requirements in Section 3 are important. Additionally, to what degree the event is actionable can play a significant role, as healthcare professionals do not act, if a serious condition is predicted that require aggressive interventions [39].

Research direction: The challenge of developing holistic EP methods requires **interdisciplinary research**. For instance, explainable events in the domain of manufacturing can be defined through a collaboration of EP researchers and mechanical engineers. By gradually increasing the number of researchers and application domains in the collaboration, we may be able to push the definition of explainable events to a generic formulation. EP is a **multi-objective** endeavor and must consider the abundance of regulatory requirements on monitored systems such as banks [45], clinics [19, 24] or manufacturing facilities [58] to define **appropriately**

complex and realistic prediction goals. A promising future research avenue is to deduce a ***markov decision process*** (**MDP**) [4] from the prediction output. By meeting all requirements for **optimal decision-making** (cf. Section 3), the prediction output is rich enough for constructing a MDP that models the future and its corresponding decisions in a statistical process. Then, the multi-objective optimization problem of EP together with formalized constraints from regulatory constraints can be combined into a goal for optimal *policy synthesis* [77]. As a result, EP can benefit from the numerous solution techniques to the **optimal policy synthesis** problem, which asks for a sequence of decisions that are most likely to guarantee the specified goal given the MDP.

In general, for a **holistic approach to method development**, future research can either combine existing properties from various methods into a single method, develop new solutions to the existing problems for a single method, or combine existing methods and, thus, their properties in a generic framework. Independent of the chosen option, a conceptual model of EP that identifies the systematic relationships between application domains, domain knowledge, prediction goals, input data, feature engineering, prediction methods and the action space is desirable. Such a model extends the structural system perspective on EP and its surrounding components embodied by the PCM system (cf. Figure 3) and the methodological perspective on EP in [161] with a **behavioral and operational system** perspective.

C2 - Semantics, Knowledge, and Explainability

Challenge: EP aims at anticipating the future of the monitored system by means of algorithms. Monitored systems come from a diverse range of application domains and, thus, have a diverse range of semantics. At some point in the PCM system, the algorithms used to develop EP should take the semantics, e.g., of the event lifecycle (cf. OB5), into account. The same applies to existing domain knowledge. Aside from potentially positive effects on accuracy, the way semantics and knowledge are taken into account determines the explainability of the method and its generality: If the method is very abstract and generic, it is typically not explainable to the domain expert, whereas a very concrete and domain-specific method is typically explainable, but does not generalize. The challenge is to shift the tradeoff between explainability and generality and develop new ways to integrate semantics and knowledge in the method for an improved explainability while maintaining generality (cf. OB6).

Research direction: Starting with a clear definition of explainability for EP, e.g., is it rather a counterfactual or the more strict causal definition [120], and the necessary content type, communication means and target group of the explanation [21], future research can evaluate the correctness of explanations. Explanation-guided learning [47] demonstrates promising approaches to explainability and correctness evaluation, yet it pre-determines the way of how the method integrates semantics and domain knowledge by stating them as explanation supervision and regularization in the learning objective. Other or additional ways of integration could be parametrization, feature engineering or a certain hypothesis space for explainable models. It is highly beneficial to explore various combinations of integration and test for generality of the method such that we can improve our understanding on how semantics and domain knowledge can be used for explainability while maintaining generality.

Since the quality of predicted events is also affected by the degree to which existing mechanistic knowledge [161] is integrated within the prediction model, **hybrid models** are beneficial. Hybrid models combine expressive, even **executable models** that represent the known causal relationships on a given monitored system with the generally well-performing machine learning models [53, 121]. The result benefits from superior prediction performance and maintains transparency. In particular, hybrid models are preferred that do not simply apply ensemble learning-based techniques, but integrate the knowledge and rules on a more fundamental level into the prediction

model similar to physics-informed neural networks [28]. Here, EP research can catch up from the related research stream that combines operations research with machine learning [53]. Yet, the count of EP methods with hybrid models is decreasing (cf. Section 4.1). Nevertheless, a general framework that represents a solution of how to integrate mechanistic knowledge across domains would be very beneficial. Since it is likely that there exist tradeoffs between the quality dimensions, user studies on typical tradeoffs in domains or in general help to integrate real-world needs to future research.

C3 - Flexibility and Robustness

Challenge: Considering the heterogeneity of application domains, data sources and prediction goals, EP methods must find a balance between flexibility and robustness (cf. OB7 and OB8). On the one hand, they have to be flexible with respect to three aspects. First, domain-specific properties and requirements (i) captures the effects of the respective application domain and environment. For example, compare a very dynamic, clinical environment in the healthcare domain with a particularly strong requirement on uncertainty vs. a relatively stable environment of a standardized goods manufacturing facility with a relatively weak requirement on uncertainty. Second, the data properties and quality issues of multiple data sources (ii) captures the effects of the available data sources, e.g., different granularities of the data such as city- vs. country-level and missing data. Third, prediction goal properties (iii) captures the effects of how the future should look like, e.g., a single, stable prediction goal such as "treatment outcome of a patient is positive" vs. multiple, dynamic prediction goals such as the set of regulatory constraints a bank's business processes have to adhere to. In addition to flexibility, EP methods have to be robust with respect to adversarial attacks and/or noisy input data [26, 72, 161], since EP in a PCM system has a direct real-world consequence through the action that the user chooses.

Research direction: One way to deal with the flexibility challenge is to limit the scope of the method to a certain problem space for which flexibility is not required anymore. So far, this approach dominates research on EP. Although this approach can solve the challenge by removing it from the equation to a large extent, it results in a very fragmented set of EP methods and transfers the challenge to the user. Future methods should abstain from definitions that simplify the problem in this way. For (i) domain-specific properties and requirements and (ii) data properties and quality issues existing work supports the required flexibility (cf. assessment for *probabilistic prediction*, *uncertainty quantification*, *online conceptualization*, *matching*, *multiple sources* in Section 4). This combination is promising for future work on meeting flexibility for (i) and (ii) together. While [91] called for flexibility regarding various sets of prediction goals, flexibility for (iii) prediction goal properties is usually still abstracted from or left unmentioned. Considering the abundance of regulatory requirements on monitored systems (cf. **C1**), EP methods that are flexible with respect to prediction goal heterogeneity are highly beneficial. Once again, **hybrid models** present themselves to separate predicting future events from the subsequent matching of prediction goals on the predicted sequence of future events [122]. As the hybrid model does not directly predict the prediction goal, it is significantly more flexible towards changing goals. Yet, a more flexible EP leaves more room for adversarial attacks. Consequently, research on flexible EP should study how these mechanisms interact with adversarial attack models. Considering noise, methods are desirable that can reliably distinguish variances in the input data that require flexibility from noise that should be captured as uncertainty.

C4 - User Interface

Challenge: As EP research focuses on the method and its rather technical evaluation, it often does not provide a simple and clear user interface of relevant output (cf. OB9). Moreover, probabilistic predictions with uncertainty estimates, multiple prediction goals, root causes and potential actions aggravate the challenge to present the output in a simple and clear way to the user. Although the

user interface corresponds to the output component of a PCM system that is separated from the prediction component, the method design should already have the subsequent user interface in mind to ease its design and implementation, e.g., by suggesting certain visualization techniques. Lastly, some methods have started proposing the integration of prediction with decision theory to include actions and their impact [18, 134, 148], but these are goal-dependent and domain-specific and, thus, cannot be easily transferred to other goals and application domains. Furthermore, [18, 134, 148] lack to provide how the additional information should be integrated to present the output to the user.

Research direction: By adopting principles of **user-centered design** [56], future research should explore means of presenting and visualizing the complex output of EP and the PCM system to the user. This challenge should not be taken lightly, as it significantly determines the adoption of the method by users and may affect the user's trust in the system. Research that incorporates actions through a decision theoretic framework further supports the user in deciding what to do based on the anticipated future. By allowing for parameterizing the action space (cf. **C1**), future work on EP can increase the flexibility of the method (cf. **C3**) and enables the user in setting the action space. The development of a **comprehensive user interface** for a PCM system that streamlines the relevant information to the user and takes subsequent actions into account is key for building trust and increasing the accessibility for users.

C5 - Evaluation

Challenge: [161] presents the evaluation design with respect to matching predicted with real events and metrics of effectiveness for EP, but neither considers probabilistic metrics of effectiveness and other dimensions of EP nor the performance in terms of training and prediction time. However, both are crucial for a standardized and valuable evaluation of EP. Without suitable evaluation metrics (cf. OB10), we cannot ascertain how successful an EP method is in achieving its design goals and how it compares to its peers. Besides, optimizing the prediction model with respect to unsuitable evaluation metrics skews the method's predicted events in the wrong direction such that we cannot derive practically meaningful conclusions from the evaluation results. Additionally, it is not clear to what extent evaluation metrics are linked to actual user value and currently popular evaluation metrics fail to measure how well a visualization presents predicted events and root causes (cf. **C4**). Hence, the challenge is to carefully design the evaluation, ideally, with users, and new metrics such that we can draw reliable conclusions for the state-of-the-art and the application.

Research direction: [23] reviewed 201 probabilistic metrics of effectiveness that are candidates for EP. Both an analysis as well as standard proposal for EP is needed. However, probabilistic metrics only account for the probabilistic nature of EP. To also account for the other dimensions such as event resolution, timeliness, confidence, intensity, explainability, and flexibility (cf. **C1** - **C3**), new metrics have to be developed that build on the standard probabilistic metrics. One promising direction is to formalize the desired qualities like timeliness as **real-time constraints** in a continuous-time logic [77]. Both evaluating the satisfiability given the prediction model and computing a likelihood for satisfaction should be achievable in a controlled experimental setting. During the development of new metrics, **interdisciplinary work** is desired such that the metric design reflects perspectives from, e.g., lawyers for explainable events. The resulting new metrics can be directly applied in a comprehensive, empirical benchmark of existing methods within comparable classes of our taxonomy for EP methods (cf. Section B in the supp. material[1]). An empirical benchmark that allows for **simple extension** with newly developed EP methods is beneficial to support future evaluation of methods similar to [131]. For evaluating visualization (cf. **C4**) and the design of a user interface, user studies have to be carried out.

6 Conclusion, Impact, and Outlook

This work presents a comprehensive survey on EP methods spanning the disparate research areas of EP and PPM. Taking this umbrella review approach allows us to draw novel challenges and research directions from a set of observations that we take on the result of assessing existing EP methods with an integrated set of requirements. Moreover, taking a system perspective lifts the survey from (technical) details of single EP methods to an integrated view in a PCM system. This system approach can be taken as a blueprint to assess the realization of real-world projects that aim to apply EP for decision support. The holistic approach with a systems perspective reveals challenges regarding the joint application of EP methods, their ability to deal with uncertainty in highly flexible environments, the evaluation of EP quality, and the inclusion of users based on advanced root cause analysis, visualization, and user interfaces.

This survey creates particular impact by knowledge sharing (cf. Section F in the appendix for background). First, this work positively mediates mechanisms by means of a shared mental model and language. The PCM approach integrates the analysed EP methods, and the assessment pertains to a shared set of requirements. Second, this work establishes a tool for several knowledge sharing mechanisms, both for academia and organizations. The classification and assessment is available in our online repository and can be used for, e.g., creating instructional material. Moreover, the survey at hand acts as a starting point to access the vast literature on EP methods in its application domains and disparate research areas. Additionally, it gives a comprehensive, state-of-the-art overview such that further research directions can be easily derived. For practitioners, the PCM system can be used as a blueprint to analyze their specific requirements; this is illustrated for two application domains in Section 2.3 and Section E of the supp. material.

In future work, we plan for applying the PCM approach in additional application scenarios, together with practitioners, and for conducting user studies for EP in order to validate research goals with real-world value and to create extended datasets that, for example, include actions. Moreover, we plan to undertake interdisciplinary work in order to overcome the complex nature of EP across domains and research areas such as law for explainability, design for the output component and visualization, as well as decision theory for prescriptive and counterfactual analysis.

References

[1] Saurav Acharya, Byung Suk Lee, and Paul Hines. 2017. Causal prediction of top-k event types over real-time event streams. *Comput. J.* 60, 11 (2017), 1561–1581.

[2] Karan Aggarwal, Onur Atan, Ahmed K. Farahat, Chi Zhang, Kosta Ristovski, and Chetan Gupta. 2018. Two birds with one network: Unifying failure event prediction and time-to-failure modeling. In *IEEE BigData*. 1308–1317.

[3] Nay Abi Akl, Jawad El Khoury, and Charbel Mansour. 2021. Trip-based prediction of hybrid electric vehicles velocity using artificial neural networks. In *IEEE IMCET*. 60–65.

[4] Oguzhan Alagoz, Heather Hsu, Andrew J. Schaefer, and Mark S. Roberts. 2010. Markov decision processes: A tool for sequential decision making under uncertainty. *MDM* 30, 4 (2010), 474–483.

[5] James Allan, Ron Papka, and Victor Lavrenko. 1998. On-line new event detection and tracking. In *Proceedings of SIGIR*. 37–45.

[6] Renato Marinho Alves, Luciana Barbieri, Kleber Stroeh, Sarajane Marques Peres, and Edmundo Roberto Mauro Madeira. 2022. Context-aware completion time prediction for business process monitoring. In *Information Systems and Technologies*. 355–365.

[7] M. Antunes, M. A. Amaral Turkman, and K. F. Turkman. 2003. A bayesian approach to event prediction. *J. Time Ser. Anal.* 24, 6 (2003), 631–646.

[8] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2019. Spatio-temporal data mining: A survey of problems and methods. *ACM Comput. Surv.* 51, 4 (2019), 1–41.

[9] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[10] Riccardo Benedetti. 2010. Scoring rules for forecast verification. *Mon. Weather Rev.* 138, 1 (2010), 203–211.
[11] Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, François-Xavier Aubet, Laurent Callot, and Tim Januschowski. 2023. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput. Surv.* 55, 6 (2023), 1–36.
[12] Ludwig von Bertalanffy. 1968. *General System Theory: Foundations, Development, Applications*. G. Braziller.
[13] Krishnanjan Bhattacharjee, S. ShivaKarthik, Swati Mehta, Ajai Kumar, Rushikesh Kothawade, Paritosh Katre, Piyush Dharkar, Neil Pillai, and Devika Verma. 2021. Survey and gap analysis on event prediction of English unstructured texts. In *Machine Learning for Predictive Analysis*. Springer, 491–501.
[14] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2011. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI global, 205–227.
[15] Ane Blázquez-García, Angel Conde, Usue Mori, and Jose A. Lozano. 2021. A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.* 54, 3 (2021), 56:1–56:33.
[16] Azzedine Boukerche, Lining Zheng, and Omar Alfandi. 2021. Outlier detection: Methods, models, and classification. *ACM Comput. Surv.* 53, 3 (2021), 1–37.
[17] Gordon H. Bower and Daniel G. Morrow. 1990. Mental models in narrative comprehension. *Science* 247, 4938 (1990), 44–48.
[18] Zahra Dasht Bozorgi, Irene Teinemaa, Marlon Dumas, Marcello La Rosa, and Artem Polyvyanyy. 2021. Prescriptive process monitoring for cost-aware cycle time reduction. In *ICPM*. 96–103.
[19] Jeffrey Braithwaite, Kate Churruca, and Louise A. Ellis. 2017. Can we fix the uber-complexities of healthcare? *J. R. Soc. Med.* 110, 10 (2017), 392–394.
[20] Andrea Brunello, Angelo Montanari, and Mark Reynolds. 2019. Synthesis of LTL formulas from natural language texts: State-of-the-art and research directions. In *TIME*. 17:1–17:19.
[21] Nadia Burkart and Marco F. Huber. 2021. A survey on the explainability of supervised machine learning. *J. Artif. Intell. Res.* 70 (2021), 245–317. https://doi.org/10.1613/jair.1.12228
[22] Rui Cao, Weijian Ni, Qingtian Zeng, Faming Lu, Cong Liu, and Hua Duan. 2021. Remaining time prediction for business processes with concurrency based on log representation. *China Communications* 18, 11 (2021), 76–91.
[23] Arthur Carvalho. 2016. An overview of applications of proper scoring rules. *Decis. Anal.* 13, 4 (2016), 223–242.
[24] Center for Regulatory Stategy Americas. 2020. *2020 Health Care Regulatory Outlook*. Technical Report.
[25] Chung-Wen Cho, Yi-Hung Wu, Show-Jane Yen, Ying Zheng, and Arbee L. P. Chen. 2011. On-line rule matching for event prediction. *VLDB J.* 20, 3 (2011), 303–334.
[26] Marco Comuzzi, Alfonso E. Marquez-Chamorro, and Manuel Resinas. 2019. Does your accurate process predictive monitoring model give reliable predictions?. In *Service-Oriented Computing – ICSOC 2018 Workshops*. 367–373.
[27] Noel Cressie. 2015. *Statistics for Spatial Data*. John Wiley and Sons.
[28] Salvatore Cuomo, Di Cola, Vincenzo Schiano, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. 2022. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *J. Sci. Comput.* 92, 3 (2022), 88.
[29] Matthew DeCamp and Jon C. Tilburt. 2019. Why we cannot trust artificial intelligence in medicine. *Lancet Digit. Health* 1, 8 (2019), e390.
[30] Akash Deep, Dharmaraj Veeramani, and Shiyu Zhou. 2020. Event prediction for individual unit based on recurrent event data collected in teleservice systems. *IEEE Trans. Rel.* 69, 1 (2020), 216–227.
[31] Walter H. Dempsey, Alexander Moreno, Christy K. Scott, Michael L. Dennis, David H. Gustafson, Susan A. Murphy, and James M. Rehg. 2017. iSurvive: An interpretable, event-time prediction model for mHealth. In *PMLR*. 970–979.
[32] Chiara Di Francescomarino and Chiara Ghidini. 2022. Predictive process monitoring. In *Process Mining Handbook*. Springer, 320–346.
[33] Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi, and Fredrik Milani. 2018. Predictive process monitoring methods: Which one suits me best?. In *BPM*. 462–479.
[34] Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi, Giulio Petrucci, and Anton Yeshchenko. 2017. An eye into the future: Leveraging A-priori knowledge in predictive business process monitoring. In *BPM*. 252–268.
[35] Kiarash Diba, Kimon Batoulis, Matthias Weidlich, and Mathias Weske. 2020. Extraction, correlation, and abstraction of event data for process mining. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 10, 3 (2020), e1346.
[36] Huilong Duan, Zhoujian Sun, Wei Dong, Kunlun He, and Zhengxing Huang. 2020. On clinical event prediction in patient treatment trajectory using longitudinal electronic health records. *IEEE J. Biomed. Health Inform.* 24, 7 (2020), 2053–2063.
[37] Eduardo e Oliveira, Vera L. Miguéis, and José L. Borges. 2022. Automatic root cause analysis in manufacturing: An overview and conceptualization. *J. Intell. Manuf.* 34, 5 (2022), 2061–2078.

[38] Matthias Ehrendorfer, Juergen Mangler, and Stefanie Rinderle-Ma. 2021. Assessing the impact of context data on process outcomes during runtime. In *ICSOC*. 3–18.
[39] Bar Eini-Porat, Ofra Amir, Danny Eytan, and Uri Shalit. 2022. Tell me something interesting: Clinical utility of machine learning prediction models in the ICU. *Journal of Biomedical Informatics* 132 (2022), 104107. https://doi.org/10.1016/j.jbi.2022.104107
[40] Catherine Elgin. 2007. Understanding and the facts. *Philos. Stud.* 132, 1 (2007), 33–42.
[41] Catherine Elgin. 2017. *True Enough*. MIT Press.
[42] Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, and Vassilios S. Verykios. 2007. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.* 19, 1 (2007), 1–16.
[43] Ioannis Flouris, Nikos Giatrakos, Antonios Deligiannakis, Minos Garofalakis, Michael Kamp, and Michael Mock. 2017. Issues in complex event processing: Status and prospects in the big data era. *J. Syst. Softw.* 127 (2017), 217–236.
[44] Joshi Fullop, Ana Gainaru, and Joel Plutchak. 2012. Real time analysis and event prediction engine. *Cray User Group* (2012). Cray User Group, Cray Inc., https://cug.org/proceedings/attendee_program_cug2012/includes/files/pap155.pdf
[45] Prasanna Gai, Malcolm Kemp, Antonio Sánchez Serrano, and Isabel Schnabel. 2019. *Regulatory Complexity and the Quest for Robust Regulation*. Technical Report. ESRB: Advisory Scientific Committee Reports 2019/8.
[46] Yan Gao and Yan Cui. 2022. Clinical time-to-event prediction enhanced by incorporating compatible related outcomes. *PLOS Digit. Health* 1, 5 (2022), e0000038.
[47] Yuyang Gao, Siyi Gu, Junji Jiang, Sungsoo Ray Hong, Dazhou Yu, and Liang Zhao. 2024. Going beyond XAI: A systematic survey for explanation-guided learning. *ACM Comput. Surv.* 56, 7 (2024), 188:1–188:39.
[48] Yuyang Gao and Liang Zhao. 2018. Incomplete label multi-task ordinal regression for spatial event scale forecasting. In *AAAI*. 2999–3006.
[49] Yuyang Gao, Liang Zhao, Lingfei Wu, Yanfang Ye, Hui Xiong, and Chaowei Yang. 2019. Incomplete label multi-task deep learning for spatio-temporal event subtype forecasting. In *AAAI*. 3638–3646.
[50] Lise Getoor and Ashwin Machanavajjhala. 2012. Entity resolution: Theory, practice and open challenges. *Proc. VLDB Endow.* 5, 12 (2012), 2018–2019.
[51] Martin Glinz. 2007. On non-functional requirements. In *IEEE RE*. 21–26.
[52] Fatma Ezzahra Gmati, Salem Chakhar, Wided Lejouad Chaari, and Mark Xu. 2019. A taxonomy of event prediction methods. In *Advances and Trends in Artificial Intelligence. From Theory to Practice*. Springer, 12–26.
[53] Andrew Greasley and John Steven Edwards. 2021. Enhancing discrete-event simulation with big data analytics: A review. *J. Oper. Res. Soc.* 72, 2 (2021), 247–267.
[54] Stephen R. Grimm. 2006. Is understanding a species of knowledge? *Br. J. Philos. Sci.* 57, 3 (2006), 515–535.
[55] Thomas Grisold, Jan Mendling, Markus Otto, and Jan vom Brocke. 2020. Adoption, use and management of process mining in practice. *Bus. Process Manag. J.* 27, 2 (2020), 369–387.
[56] Jan Gulliksen, Bengt Göransson, Inger Boivie, Stefan Blomkvist, Jenny Persson, and Åsa Cajander. 2003. Key principles for user-centred systems design. *Behav. Inf. Technol.* 22, 6 (2003), 397–409.
[57] Nitin Harane and Sheetal Rathi. 2020. Comprehensive survey on deep learning approaches in predictive business process monitoring. In *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI*. Springer, 115–128.
[58] Germán Herrera Vidal and Jairo Rafael Coronado Hernández. 2021. Complexity in manufacturing systems: A literature review. *Prod. Eng.* 15, 3-4 (2021), 321–333.
[59] Rachel Heyard, Jean-François Timsit, Leonhard Held, and Combacte-Magnet Consortium. 2020. Validation of discrete time-to-event prediction models in the presence of competing risks. *Biom. J.* 62, 3 (2020), 643–657.
[60] Andrew Hoegh, Scotland Leman, Parang Saraf, and Naren Ramakrishnan. 2015. Bayesian model fusion for forecasting civil unrest. *Technometrics* 57, 3 (2015), 332–340.
[61] Wenjie Hu, Yang Yang, Ziqiang Cheng, Carl Yang, and Xiang Ren. 2021. Time-series event prediction with evolutionary state graph. In *ACM WSDM*. 580–588.
[62] Kristian Wang Høiem, Vemund Santi, Bendik Nybakk Torsæter, Helge Langseth, Christian André Andresen, and Gjert H. Rosenlund. 2020. Comparative study of event prediction in power grids using supervised machine learning methods. In *SEST*. 1–6.
[63] Eyke Hüllermeier and Willem Waegeman. 2021. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* 110, 3 (2021), 457–506.
[64] Ihab F. Ilyas and Xu Chu. 2015. Trends in cleaning relational data: Consistency and deduplication. *Found. Trends Databases* 5, 4 (2015), 281–393.
[65] Fakhruzy Shakirin Jamaludin and R. Kanesaraj Ramasamy. 2020. Systematic review on event prediction models. *Indones. J. Electr. Eng. Comput. Sci.* 19, 3 (2020), 1490.

[66] Zhongjun Jin, Michael R. Anderson, Michael Cafarella, and H. V. Jagadish. 2017. Foofah: Transforming data by example. In *SIGMOD*. 683–698.
[67] John Rydning. 2022. *Worldwide IDC Global DataSphere Forecast, 2022–2026*. Technical Report.
[68] B. Kamala and B. Latha. 2022. Process mining and deep neural network approach for the prediction of business process outcome. In *IC3IoT*. 1–4.
[69] Bokyoung Kang, Dongsoo Kim, and Suk-Ho Kang. 2012. Periodic performance prediction for real-time business process monitoring. *Ind. Manag. Data Syst.* 112, 1 (2012), 4–23.
[70] Wei Kang, Jie Chen, Jiuyong Li, Jixue Liu, Lin Liu, Grant Osborne, Nick Lothian, Brenton Cooper, Terry Moschou, and Grant Neale. 2017. Carbon: Forecasting civil unrest events by monitoring news and social media. In *Advanced Data Mining and Appl.* 859–865.
[71] Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *LREC*. 1027–1032.
[72] Christopher Klinkmüller, Nick R. T. P. van Beest, and Ingo Weber. 2018. Towards reliable predictive process monitoring. In *CAiSE*. 163–181.
[73] Wolfgang Kratsch, Jonas Manderscheid, Maximilian Röglinger, and Johannes Seyfried. 2021. Machine learning in business process monitoring: A comparison of deep learning and classical approaches used for outcome prediction. *Business and Inf. Syst. Engineering* 63, 3 (2021), 261–276.
[74] Jonathan L. Kvanvig. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.
[75] Martin Käppel, Stefan Jablonski, and Stefan Schönig. 2021. Evaluating predictive business process monitoring approaches on small event logs. In *QUATIC*. 167–182.
[76] Zineb Lamghari, Maryam Radgui, Rajaa Saidi, and Moulay Driss Rahmani. 2019. Predictive process monitoring related to the remaining time dimension: A value-driven framework. In *ICSSD*. 1–6.
[77] Abolfazl Lavaei, Sadegh Soudjani, Alessandro Abate, and Majid Zamani. 2022. Automated verification and synthesis of stochastic hybrid systems: A survey. *Automatica* 146 (2022), 110617. https://doi.org/10.1016/j.automatica.202
[78] Sander J. J. Leemans, Sebastiaan J. van Zelst, and Xixi Lu. 2023. Partial-order-based process mining: A survey and outlook. *Knowledge and Inf. Syst.* 65, 1 (2023), 1–29.
[79] Lei Lei, Xuguang Ren, Nigel Franciscus, Junhu Wang, and Bela Stantic. 2019. Event prediction based on causality reasoning. In *ACIIDS*. 165–176.
[80] Anna Leontjeva, Raffaele Conforti, Chiara Di Francescomarino, Marlon Dumas, and Fabrizio Maria Maggi. 2015. Complex symbolic sequence encodings for predictive monitoring of business processes. In *BPM*. 297–313.
[81] Benjamin Letham, Cynthia Rudin, Tyler H. McCormick, and David Madigan. 2015. Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *Ann. Appl. Stat.* 9, 3 (2015), 1350–1371.
[82] Zhongyang Li, Xiao Ding, and Ting Liu. 2018. Constructing narrative event evolutionary graph for script event prediction. In *AAAI*. 4201–4207.
[83] Bryan Lim and Stefan Zohren. 2021. Time-series forecasting with deep learning: A survey. *Philos. trans., Math. Phys. Eng. Sci.* 379 (2021), 20200209. https://doi.org/10.1098/rsta.2020.0209
[84] Li Lin, Lijie Wen, and Jianmin Wang. 2019. MM-Pred: A deep predictive model for multi-attribute event sequence. In *ICDM*. 118–126.
[85] Xin-Xue Lin, Phone Lin, and En-Hau Yeh. 2021. Anomaly detection/prediction for the internet of things: State-of-the-art and the future. *IEEE Network* 35, 1 (2021), 212–218.
[86] Benjamin Lindemann, Nasser Jazdi, and Michael Weyrich. 2020. Anomaly detection and prediction in discrete manufacturing based on cooperative LSTM networks. In *IEEE CASE*. 1003–1010.
[87] Hua Liu and D. E. Brown. 2004. A new point process transition density model for space-time event prediction. *IEEE Trans. Syst., Man, Cybern. C* 34, 3 (2004), 310–324.
[88] Sicen Liu, Xiaolong Wang, Yang Xiang, Hui Xu, Hui Wang, and Buzhou Tang. 2022. Multi-channel fusion LSTM for medical event prediction using EHRs. *J. Biomed. Inform.* 127 (2022), 104011. https://doi.org/10.1016/j.jbi.2022.104011
[89] Chang Lu, Tian Han, and Yue Ning. 2022. Context-aware health event prediction via transition functions on dynamic disease graphs. *AAAI* 36, 4 (2022), 4567–4574.
[90] Chang Lu, Chandan K. Reddy, and Yue Ning. 2021. Self-supervised graph learning with hyperbolic embedding for temporal health event prediction. *IEEE Trans. Cybern.* 53, 4 (2021), 2124–2136.
[91] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil M. P. van der Aalst. 2015. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Inf. Syst.* 54 (2015), 209–234. https://doi.org/10.1016/j.is.2015.02.007
[92] Qiuping Ma, Hongyan Li, and Anders Thorstenson. 2021. A big data-driven root cause analysis system: Application of Machine Learning in quality problem solving. *Comput. Ind. Eng.* 160 (2021), 107580. https://doi.org/10.1016/j.cie.2021.107580

[93] Juergen Mangler, Joscha Grüger, Lukas Malburg, Matthias Ehrendorfer, Yannis Bertrand, Janik-Vasily Benzin, Stefanie Rinderle-Ma, Estefania Serral Asensio, and Ralph Bergmann. 2023. DataStream XES extension: Embedding IoT sensor data into extensible event stream logs. *Fut. Int.* 15, 3 (2023), 109.

[94] Ricardo P. Masini, Marcelo C. Medeiros, and Eduardo F. Mendes. 2023. Machine learning advances for time series forecasting. *J. Econ. Surv.* 37, 1 (2023), 76–111.

[95] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. 2012. Fast mining and forecasting of complex time-stamped events. In *SIGKDD*. 271–279.

[96] John Meehan, Cansu Aslantas, Stan Zdonik, Nesime Tatbul, and Jiang Du. 2017. Data ingestion for the connected world. In *CIDR*. 8–11.

[97] Soheila Mehrmolaei and Mohammad Reza Keyvanpourr. 2015. A brief survey on event prediction methods in time series. In *Artificial Intelligence Perspectives and Applications*. Springer, 235–246.

[98] Andreas Metzger and Felix Föcker. 2017. Predictive business process monitoring considering reliability estimates. In *CAiSE*. 445–460.

[99] Andreas Metzger, Philipp Leitner, Dragan Ivanović, Eric Schmieders, Rod Franklin, Manuel Carro, Schahram Dustdar, and Klaus Pohl. 2015. Comparing and combining predictive business process monitoring techniques. *IEEE Trans. Syst., Man, Cybern., Syst.* 45, 2 (2015), 276–290.

[100] Luay Thamer Mohammed, AbdAllah A. AlHabshy, and Kamal A. ElDahshan. 2022. Big data visualization: A survey. In *IEEE HORA*. 1–12.

[101] Soheila Mehr Molaei and Mohammad Reza Keyvanpour. 2015. An analytical review for event prediction system on time series. In *IEEE IPRIA*. 1–6.

[102] Alfonso E. Márquez-Chamorro, Isabel A. Nepomuceno-Chamorro, Manuel Resinas, and Antonio Ruiz-Cortés. 2022. Updating prediction models for predictive process monitoring. In *CAiSE*. 304–318.

[103] Alfonso Eduardo Márquez-Chamorro, Manuel Resinas, and Antonio Ruiz-Cortés. 2018. Predictive monitoring of business processes: A survey. *TSC* 11, 6 (2018), 962–977.

[104] Felix Naumann and Melanie Herschel. 2010. *An Introduction to Duplicate Detection*. Springer, Cham.

[105] Dominic A. Neu, Johannes Lahann, and Peter Fettke. 2022. A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artif. Intell. Rev.* 55, 2 (2022), 801–827.

[106] Yue Ning, Rongrong Tao, Chandan K. Reddy, Huzefa Rangwala, James C. Starz, and Naren Ramakrishnan. 2018. STAPLE: Spatio-temporal precursor learning for event forecasting. In *Proceedings of SDM. SIAM*. 99–107.

[107] Niyi Ogunbiyi, Artie Basukoski, and Thierry Chaussalet. 2021. Comparative analysis of clustering-based remaining-time predictive process monitoring approaches. *Int. J. Bus. Process Integr. Manag.* 10, 3-4 (2021), 230–241.

[108] Alexander Osterwalder and Yves Pigneur. 2010. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley and Sons.

[109] Gyunam Park and Wil M. P. van der Aalst. 2022. Action-oriented process mining: Bridging the gap between insights and actions. *Prog. Artif. Intell.* (2022), 1–22. https://doi.org/10.1007/s13748-022-00281-7

[110] Seo Young Park, Ji Eun Park, Hyungjin Kim, and Seong Ho Park. 2021. Review of statistical methods for evaluating the performance of survival or other time-to-event prediction models (from conventional to deep learning approaches). *Korean J. Radiol.* 22, 10 (2021), 1697–1707.

[111] Florian Pauker, Juergen Mangler, Stefanie Rinderle-Ma, and Matthias Ehrendorfer. 2021. Industry 4.0 integration assessment and evolution at EVVA GmbH: Process-driven automation through *centurio.work*. In *Business Process Management Cases Vol. 2*. 81–91.

[112] Francisco José García Peñalvo, Tamanna Maan, Sunil K. Singh, Sudhakar Kumar, Varsha Arya, Kwok Tai Chui, and Gaurav Pratap Singh. 2022. Sustainable stock market prediction framework using machine learning models. *IJSSCI* 14, 1 (2022), 1–15.

[113] Kaci L. Pickett, Krithika Suresh, Kristen R. Campbell, Scott Davis, and Elizabeth Juarez-Colunga. 2021. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Med. Res. Methodol.* 21, 1 (2021), 216.

[114] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning causality for news events prediction. In *ACM WWW*. 909–918.

[115] Erhard Rahm and Philip A. Bernstein. 2001. A survey of approaches to automatic schema matching. *VLDB J.* 10, 4 (2001), 334–350.

[116] Efren Rama-Maneiro, Juan Vidal, and Manuel Lama. 2021. Deep learning for predictive business process monitoring: Review and benchmark. *TSC* 16, 1 (2021), 739–756.

[117] Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua, Feng Chen, Chang Tien Lu, Bert Huang, Aravind Srinivasan, Khoa Trinh, Lise Getoor, Graham Katz, Andy Doyle, Chris Ackermann, Ilya Zavorin, Jim Ford, Kristen Summers, Youssef Fayed, Jaime Arredondo, Dipak Gupta, and David Mares. 2014. 'Beating the news' with EMBERS: Forecasting civil unrest using open source indicators. In *SIGKDD*. 1799–1808.

[118] Aaqib Rashid and Amit Chaturvedi. 2019. Cloud computing characteristics and services: A brief review. *Int. J. Comput. Sci. Eng.* 7, 2 (2019), 421–426.
[119] Jonathan Rebane, Isak Karlsson, and Panagiotis Papapetrou. 2019. An investigation of interpretable deep learning for adverse drug event prediction. In *IEEE CBMS*. 337–342.
[120] Collin Rice. 2021. *Leveraging Distortions: Explanation, Idealization, and Universality in Science*. The MIT Press.
[121] Stefanie Rinderle-Ma, Jürgen Mangler, and Daniel Ritter. 2024. *Fundamentals of Information Systems Interoperability*. Springer.
[122] Stefanie Rinderle-Ma, Karolin Winter, and Janik-Vasily Benzin. 2023. Predictive compliance monitoring in process-aware information systems: State-of-the-art, functionalities, research directions. *Inf. Sys.* 115 (2023), 102210. https://doi.org/10.1016/j.is.2023.102210
[123] Williams Rizzi, Chiara Di Francescomarino, Chiara Ghidini, and Fabrizio Maria Maggi. 2022. How do I update my model? On the resilience of predictive process monitoring models to change. *Knowledge and Inf. Syst.* 64, 5 (2022), 1385–1416.
[124] Yuji Roh, Geon Heo, and Steven Euijong Whang. 2021. A survey on data collection for machine learning: A big data - AI integration perspective. *TKDE* 33, 4 (2021), 1328–1347.
[125] Massimiliano Ronzani, Roger Ferrod, Chiara Di Francescomarino, Emilio Sulis, Roberto Aringhieri, Guido Boella, Enrico Brunetti, Luigi Di Caro, Mauro Dragoni, Chiara Ghidini, and Renata Marinello. 2022. Unstructured data in predictive process monitoring: Lexicographic and semantic mapping to ICD-9-CM codes for the home hospitalization service. In *AIxIA 2021—Advances in Artificial Intelligence*. 700–715.
[126] Mohammed Mohammed Sadeeq, Nasiba M. Abdulkareem, Subhi R. M. Zeebaree, Dindar Mikaeel Ahmed, Ahmed Saifullah Sami, and Rizgar R. Zebari. 2021. IoT and cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal* 1, 2 (2021), 1–7.
[127] Ramendra K. Sahoo, Adam J. Oliner, Irina Rish, Manish Gupta, José E. Moreira, Sheng Ma, Ricardo Vilalta, and Anand Sivasubramaniam. 2003. Critical event prediction for proactive management in large-scale computer clusters. In *SIGKDD*. 426–435.
[128] Felix Salfner, Maren Lenk, and Miroslaw Malek. 2010. A survey of online failure prediction methods. *ACM Computing Surveys* 42, 3 (2010), 1–42.
[129] Beate Scheibel and Stefanie Rinderle-Ma. 2022. Decision mining with time series data based on automatic feature generation. In *CAiSE*. 3–18.
[130] Beate Scheibel and Stefanie Rinderle-Ma. 2023. An end-to-end approach for online decision mining and decision drift analysis in process-aware information systems. In *CAiSE*. 17–25.
[131] Zezhi Shao, Fei Wang, Yongjun Xu, Wei Wei, Chengqing Yu, Zhao Zhang, Di Yao, Tao Sun, Guangyin Jin, Xin Cao, Gao Cong, Christian S. Jensen, and Xueqi Cheng. 2025. Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis. *TKDE* 37, 1 (2025), 291–305.
[132] Siva S. Sivatha Sindhu, S. Geetha, Siva. S. Sivanath, and A. Kannan. 2006. A neuro-genetic ensemble short term forecasting framework for anomaly intrusion prediction. In *IEEE AdCom*. 187–190.
[133] Sweta Singh, Rakesh Kumar, and Udai Pratap Rao. 2022. Multi-objective adaptive manta-ray foraging optimization for workflow scheduling with selected virtual machines using time-series-based prediction. *IJSSCI* 14, 1 (2022), 1–25.
[134] Hossein Soleimani, James Hensman, and Suchi Saria. 2018. Scalable joint models for reliable uncertainty-aware event prediction. *TPAMI* 40, 8 (2018), 1948–1963.
[135] Florian Spree. 2020. Predictive process monitoring - a use-case-driven literature review. In *EMISA Forum: Vol. 40, No. 1*. 43–54.
[136] Florian Stertz and Stefanie Rinderle-Ma. 2019. Detecting and identifying data drifts in process event streams based on process histories. In *CAiSE*. 240–252.
[137] Florian Stertz, Stefanie Rinderle-Ma, and Juergen Mangler. 2020. Analyzing process concept drifts based on sensor event streams during runtime. In *BPM*. 202–219.
[138] Matthias Stierle, Jens Brunk, Sven Weinzierl, Sandra Zilker, Martin Matzner, and Jörg Becker. 2021. *Bringing Light Into the Darkness - A Systematic Literature Review on Explainable Predictive Business Process Monitoring Techniques*.
[139] Michael Stonebraker and Ihab F. Ilyas. 2018. Data integration: The current status and the way forward. *IEEE Data Eng. Bull.* 41, 2 (2018), 3–9.
[140] Bayu Adhi Tama and Marco Comuzzi. 2019. An empirical comparison of classification techniques for next event prediction using business process event logs. *Expert Syst. Appl.* 129 (2019), 233–245. https://doi.org/10.1016/j.eswa.2019.04.016
[141] Bayu Adhi Tama, Marco Comuzzi, and Jonghyeon Ko. 2020. An empirical investigation of different classifiers, encoding, and ensemble schemes for next event prediction using business process event logs. *ACM Trans. Intell. Syst. Technol.* 11, 6 (2020), 1–34.

[142] Niek Tax, Irene Teinemaa, and Sebastiaan J. van Zelst. 2020. An interdisciplinary comparison of sequence modeling methods for next-element prediction. *Softw. Syst. Model.* 19, 6 (2020), 1345–1365.
[143] Irene Teinemaa, Marlon Dumas, Marcello La Rosa, and Fabrizio Maria Maggi. 2019. Outcome-oriented predictive process monitoring: Review and benchmark. *ACM Trans. Knowl. Discov. Data* 13, 2 (2019), 17:1–17:57.
[144] Stefan Thalmann, Juergen Mangler, Tobias Schreck, Christian Huemer, Marc Streit, Florian Pauker, Georg Weichhart, Stefan Schulte, Christian Kittl, Christoph Pollak, Matej Vukovic, Gerti Kappel, Milot Gashi, Stefanie Rinderle-Ma, Josef Suschnigg, Nikolina Jekic, and Stefanie Lindstaedt. 2018. Data analytics for industrial process improvement a vision paper. In *Business Informatics Workshops*. 92–96.
[145] Oliver Thomas. 2005. Understanding the term reference model in information systems research: History, literature analysis and explanation. In *BPM*. Springer, 484–496.
[146] Sebastiaan J. van Zelst, Felix Mannhardt, Massimiliano de Leoni, and Agnes Koschmider. 2021. Event abstraction in process mining: Literature review and taxonomy. *Granul. Comput.* 6, 3 (2021), 719–736.
[147] Ilya Verenich, Marlon Dumas, Marcello La Rosa, Fabrizio Maria Maggi, and Irene Teinemaa. 2019. Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring. *ACM Trans. Intell. Syst. Technol.* 10, 4 (2019), 34:1–34:34.
[148] Jingwen Wang, Jingxin Liu, Juntao Pu, Qinghong Yang, Zhongchen Miao, Jian Gao, and You Song. 2019. An anomaly prediction framework for financial IT systems using hybrid machine learning methods. *J. Ambient Intell. Humaniz. Comput.* 14, 11 (2019), 15277–15286.
[149] Junxiang Wang, Yuyang Gao, Andreas Züfle, Jingyuan Yang, and Liang Zhao. 2018. Incomplete label uncertainty estimation for petition victory prediction with dynamic features. In *IEEE ICDM*. 537–546.
[150] Lihong Wang, Juwei Yue, Shu Guo, Jiawei Sheng, Qianren Mao, Zhenyu Chen, Shenghai Zhong, and Chen L. 2021. Multi-level connection enhanced representation learning for script event prediction. In *Web Conference 2021*. 3524–3533.
[151] Qi Wang, Guangyin Jin, Xia Zhao, Yanghe Feng, and Jincai Huang. 2020. CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale. *Knowl. Based. Syst.* 189 (2020), 105120. https://doi.org/10.1016/j.knosys.2019.105120
[152] Senzhang Wang, Jiannong Cao, and Philip S. Yu. 2022. Deep learning for spatio-temporal data mining: A survey. *TKDE* 34, 8 (2022), 3681–3700.
[153] Hans Weytjens and Jochen De Weerdt. 2020. Process outcome prediction: CNN vs. LSTM (with attention). In *BPM Workshops*. 321–333.
[154] Hans Weytjens and Jochen De Weerdt. 2022. Learning uncertainty with artificial neural networks for predictive process monitoring. *Appl. Soft Comput.* 125 (2022), 109134. https://doi.org/10.1016/j.asoc.2022.109134
[155] Robin Whittemore, Ariana Chao, Myoungock Jang, Karl E. Minges, and Chorong Park. 2014. Methods for knowledge synthesis: An overview. *Heart and Lung* 43, 5 (2014), 453–461.
[156] Frederik Wolf, Jens Brunk, and Jörg Becker. 2021. A framework of business process monitoring and prediction techniques. In *Innovation Through Information Systems*. Springer, 714–724.
[157] Songyun Wu, Bo Wang, Zhiliang Wang, Shuhan Fan, Jiahai Yang, and Jia Li. 2022. Joint prediction on security event and time interval through deep learning. *Comput. Secur.* 117 (2022), 102696. https://doi.org/10.1016/j.cose.2022.102696
[158] Zhiliang Wu, Yinchong Yang, Peter A. Fashing, and Volker Tresp. 2021. Uncertainty-aware time-to-event prediction using deep kernel accelerated failure time models. In *Proceedings of the 6th Machine Learning for Healthcare Conference*. PMLR, 54–79.
[159] Qiang Yang, Hui Wang, and Wei Zhang. 2002. Web-log mining for quantitative temporal-event prediction. *IEEE Comput. Intell. Mag.* 1, 1 (2002), 10–18.
[160] Hongming Zhang, Muhao Chen, Haoyu Wang, Yangqiu Song, and Dan Roth. 2020. Analogous process structure induction for sub-event sequence prediction. In *EMNLP*. 1541–1550.
[161] Liang Zhao. 2022. Event prediction in the big data era: A systematic survey. *ACM Comput. Surv.* 54, 5 (2022), 1–37.
[162] Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Spatiotemporal event forecasting in social media. In *SIAM SDM*. 963–971.
[163] Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2016. Multi-resolution spatial event forecasting in social media. In *IEEE ICDM*. 689–698.
[164] Liang Zhao, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2016. Hierarchical incomplete multi-source feature learning for spatiotemporal event forecasting. In *Proceedings of SIGKDD*. 2085–2094.
[165] Alice Zheng and Amanda Casari. 2018. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. "O'Reilly Media, Inc.".

Received 8 February 2023; revised 15 April 2025; accepted 2 June 2025