BRINGING LIGHT INTO THE DARKNESS - A SYSTEMATIC LITERATURE REVIEW ON EXPLAINABLE PREDICTIVE BUSINESS PROCESS MONITORING TECHNIQUES

Research in Progress

Matthias Stierle, University of Erlangen-Nürnberg, Nürnberg, Germany, matthias.stierle@fau.de
Jens Brunk, University of Muenster - ERCIS, Münster, Germany, jens.brunk@ercis.uni-muenster.de
Sven Weinzierl, University of Erlangen-Nürnberg, Nürnberg, Germany, sven.weinzierl@fau.de
Sandra Zilker, University of Erlangen-Nürnberg, Nürnberg, Germany, sandra.zilker@fau.de
Martin Matzner, University of Erlangen-Nürnberg, Nürnberg, Germany, martin.matzner@fau.de
Jörg Becker, University of Muenster - ERCIS, Münster, Germany, joerg.becker@ercis.uni-muenster.de

Abstract

Predictive business process monitoring (PBPM) provides a set of techniques to perform different prediction tasks in running business processes, such as the next activity, the process outcome, or the remaining time. Nowadays, deep-learning-based techniques provide more accurate predictive models. However, the explainability of these models has long been neglected. The predictive quality is essential for PBPM-based decision support systems, but also its explainability for human stakeholders needs to be considered. Explainable artificial intelligence (XAI) describes different approaches to make machine-learning-based techniques explainable. To examine the current state of explainable PBPM techniques, we perform a structured and descriptive literature review. We identify explainable PBPM techniques of the domain and classify them along with different XAI-related concepts: prediction purpose, intrinsically interpretable or post-hoc, evaluation objective, and evaluation method. Based on our classification, we identify trends in the domain and remaining research gaps.

Keywords: business process, prediction, interpretability, explainable artificial intelligence.

1 Introduction

The highly volatile and uncertain digital economy increases the pressure on organizations to proactively manage their business processes (Poll et al., 2018). As a consequence, predictive business process monitoring (PBPM) is gaining momentum in information systems research (Breuker et al., 2016).
PBPM provides a set of techniques for predicting different targets in running business processes, such as process-related key performance indicators (KPIs) (i.e., process outcomes), future behavior (i.e., next activities), or remaining times (Marquez-Chamorro et al., 2018). A PBPM technique receives as input historical event log data extracted from information systems and constructs predictive models that generate predictions for different process stakeholders. While the first PBPM techniques used model-based approaches like annotated transition nets (van der Aalst et al., 2011) or stochastic Petri nets (Rogge-Solti and Weske, 2013), more recent techniques rely on machine learning (ML), in particular deep learning

(DL) algorithms for constructing more accurate predictive models to achieve adoption in practice. However, due to their complexity, these models often lack explainability for humans (Du et al., 2019).
Further, early PBPM techniques solely focused on the control-flow of a business process; generally, the structure of the process is defined by the activities and their relationship to each other. Driven by the need to improve the predictive models' quality to make them applicable in real-world business environments, researchers started to include additional process attributes (Brunk et al., 2020b). These process attributes are referred to as the context of the business process, and researchers used them as input for more sophisticated — often DL-based — techniques. Consequently, the complexity of the predictive models increased further.
Striving for more accurate predictive models, the explanation of the generated predictions has faded into the background. Nevertheless, Marquez-Chamorro et al. (2018) emphasized in their comprehensive review on PBPM that the predictive quality of PBPM techniques is not the only crucial driver for adoption in practice. When a decision support system (DSS) like a PBPM system is implemented in a real business setting where predictions and recommendations are provided to humans, they are often skeptical about the system-based decision (Kvamme et al., 2018; Wanner et al., 2020). In many cases, the reason for this reaction lies in the closed design — usually referred to as *black box* — of the ML-based PBPM systems. For example, a process analyst will be eager to discover new insights from a predictive model. However, it will often be difficult or even infeasible to understand how or why a PBPM technique reached its decision. In contrast, developers of such systems need to be able to understand why a particular decision was made for the purpose of validating and debugging predictive models. It is crucial to make ML-based decisions explainable to its different stakeholders (Mohseni et al., 2020) which for PBPM can range from *artificial intelligence (AI) novices* such as process managers over *data experts* such as process analysts up to *AI experts* such as data scientists (who are responsible for the system's development and maintenance), to support and ensure its adoption in practice.
To make ML-based process predictions explainable, business process management (BPM) researchers have begun to transfer and apply approaches from the field of Explainable Artificial Intelligence (XAI) to PBPM. XAI is a relatively new sub-field of artificial intelligence and represents a class of ML techniques that aims to enable humans to understand, trust and manage the advanced artificial DSSs by producing more explainable models while maintaining a high level of predictive quality (Gunning, 2017). As a consequence, the fast-developing research field of explainable predictive business process monitoring (XPBPM) has emerged in the last few years in the context of PBPM, or more generally BPM, with the goal to make PBPM techniques explainable. However, existing works in XPBPM currently draw a somewhat fragmented picture. They differ in purpose, prediction task, model type, evaluation strategy as well as applied XAI techniques.
Thus, our work seeks to shed light on the current state of research on XPBPM and to pave the way for future research in this field. Our research goal (RG), therefore, is: *Provide an overview and classification of existing ML-based XPBPM techniques.*
To reach our research goal, we perform a structured literature review following Webster and Watson (2002). We consider the emerging classification of this work as an initial step towards a comprehensive framework for evaluating existing XPBPM techniques. At this stage, the classification serves for BPM and IS researchers as a first reference point to identify future research needs in XPBPM.
Section 2 introduces the preliminaries of XAI and, subsequently, the concepts which serve as a basis for our systematic review. Section 3 describes the literature review process and Section 4 presents the results of this process. In Section 5, we draw conclusions from the results for XPBPM research. Section 6 concludes the paper with a brief summary.

2 Related Concepts

XAI aims to provide explanations of ML models to make their behavior more intelligible for humans (Gunning et al., 2019), while at the same time maintaining high predictive quality (Barredo Arrieta et al., 2020). With XAI's evolution, some closely related terminologies to describe a concept of intelligibility have been asserted in research and are often used as synonyms (Adadi and Berrada, 2018; Barredo Arrieta et al., 2020; Lipton, 2018). *Explainability* is provided through an explanation, which functions as an interface between humans and decision-making instances. Therefore, it is on the one side intelligible to humans, while at the same time functions as an accurate proxy of the decision-maker (Guidotti et al., 2019). Concerning *interpretability*, there is no distinct definition, but a multitude of descriptions exists (Molnar, 2020; Murdoch et al., 2019), which mostly share one aspect: A ML model is interpretable if humans can understand it. Furthermore, interpretability cannot be seen as a binary concept but as the degree to which the ML model can be understood (Molnar, 2020). In other words, the higher the interpretability of an ML model, the easier a human understands it. Further, in the context of AI, *comprehensibility* describes the ability of a learning algorithm to represent its output in a way that it can be understood by humans (Craven, 1996; Gleicher, 2016).

In this paper, we use the term *explainability*. We consider *interpretability* and *comprehensibility* as synonyms in our literature search because they are often used interchangeably (Adadi and Berrada, 2018) and, currently, no consensus on their use exists (Lipton, 2018).

In their seminal paper on interpretable ML (i.e., XAI), Du et al. (2019) present various concepts for structuring the topic. As *purpose* for the application of interpretable ML, they mention model validation and debugging as well as knowledge discovery. In other words, the audience of the explanation is either the model developers (validation and debugging) or the end-users of the decision support system (knowledge discovery).

To classify interpretable ML techniques, Du et al. (2019) distinguish between two different concepts: *intrinsic* interpretable models and *post-hoc* explanations. Models that exhibit intrinsic interpretability - often due to their structure - are "self-explanatory" (Du et al., 2019). Common methods used in PBPM that provide intrinsic interpretability are either *rule-based*, *regression-based*, *tree-based*, or *probabilistic-based*.

Rule-based The rules learned from the data set are used for prediction, and, thus, the rule itself serves as an explanation.

Regression-based The coefficients learned for each feature indicate its influence, and, hence, the decision process for the prediction is transparent.

Tree-based A decision tree learned from the data set visually outlines the decision process and the split points. However, more advanced tree-based methods suffer from increased complexity. Tree-based ensemble learning techniques (e.g., random forest (RF) (Breiman, 2001) or extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016) combine various trees and, therefore, can often not be considered intrinsically explainable.

Probabilistic-based Probabilistic models are based on learned conditional probability tables that represent the models' decision-making process. As such, the impact of each feature on the prediction is transparent. Furthermore, they can be visualized to indicate the relationship between features.

Post-hoc explanations, on the other hand, are created after model training. Post-hoc methods can be further divided into *model-agnostic* and *model-specific*. As the names suggest, the former work independently of the ML model's intricate structure and its concrete implementation, while the latter is specifically designed for a certain model type (Du et al., 2019). Post-hoc explanations can also be created for intrinsically interpretable models. Last, post-hoc explanations can be either *local* (i.e. for single instances) or *global* (i.e. the overall model) (Du et al., 2019).

An important aspect of designing interpretable ML techniques is the evaluation — both the *objective* and the *method*. In regards to the objective, ML techniques are commonly evaluated by their *predictive quality*, which is not suitable for assessing the model's interpretability. Assessing the degree of interpretability or the quality of an explanation is a challenging endeavor, and especially a quantitative evaluation is difficult to achieve. Commonly, authors try to demonstrate the *applicability* of the explanation by providing an

example. Especially for post-hoc explanations, the *faithfulness* is an important aspect to evaluate as the information retrieved from the model is subjected to a certain degree of uncertainty (Du et al., 2019).
Hevner et al. (2004) outline various evaluation methods for artifacts. For ML techniques, a common approach is to develop an instantiation of the technique for training and *testing* it with data. Moreover, a *descriptive* evaluation through demonstration of the technique — and most importantly, its ability to provide interpretability — can be achieved by using the instantiation or by creating a scenario. For *observational* evaluation, case studies can be used to apply the technique in a real-life setting. Lastly, an *experimental* evaluation, for example, through user studies, could be applied to assess specific aspects of the techniques.

3 Research Method

This work employs a systematic literature review that follows the concept-centric notion of Webster and Watson (2002). Our overall goal is to identify the status quo of explainable techniques in the PBPM domain. Therefore, we identify existing techniques, which employ XAI methods or are explainable by nature. Thus, this literature review is a descriptive review (Paré et al., 2015). In the following, we describe the characteristics of our literature review further (Cooper, 1988).
The *focus* of our systematic literature review is on PBPM techniques that were developed and or applied in the broader domain of BPM. We also identify if and how the type of PBPM prediction goal is related to the application of XAI methods. Since we perform a descriptive review, the *perspective* is neutral and the *organization* is conceptual. The *audience* of this publication is primarily oriented towards scholars. However, practitioners who intend to apply explainable PBPM techniques can also benefit from the results. Further, we structure our systematic literature review and the respective tasks based on the four primary phases proposed by Okoli and Schabram (2010).

Planning Phase. In the first phase, we established the literature review's purpose by developing our research goal, as described in Section 1. Additionally, we defined a protocol to carry out the literature search. For the literature search process, we chose the databases *Scopus*, *EbscoHost*, and *Web of Science*, which together cover a wide range of academic journals and conferences relevant in our domain. Next, we developed a search string based on precise keywords and their synonyms. Moreover, in iterative pilot search processes, we revised the keywords relevant for our literature review. We combined the final set of keywords with boolean operators as follows:

TITLE-ABS-KEY(”business process” AND ”predict” AND (”interpretable” OR ”explainable” OR ”comprehensible”))*

Selection Phase. In the second phase, we conducted the search and selected the relevant papers. The conducted search process in the selection phase is depicted in Figure 1. We performed the searches in October 2020, which resulted in 81 hits. After filtering these for relevant papers and removing duplicates, twelve papers remained. After reading all of them in detail, we performed backward and forward reference searches as suggested by Webster and Watson (2002), which resulted in four additional relevant hits. Lastly, we screened the proceedings of highly relevant conferences in the domain of PBPM. These are the International Conference on Business Process Management (BPM) and the International Conference on Process Mining (ICPM). From those, we retrieved three additional relevant papers. In the end, we identified 19 PBPM techniques, which either apply XAI to make their predictions explainable or are explainable by design.

Extraction Phase. In the third step, we extracted the applicable data from the relevant sources. Based on our research goal and the theoretical background, we derived initial concepts using a deductive approach. Resting on this, we created a concept matrix according to Webster and Watson (2002) based on the concepts described in Section 2. During this process, the concepts were revised and successively refined.

Execution Phase. Four researchers analyzed all of the identified publications by reading the full-text and, if possible, reviewing implementations of the techniques to ensure the reliability of the results. Building on the created concept matrix, we derived different findings. Finally, we diffuse our results in this literature review paper.

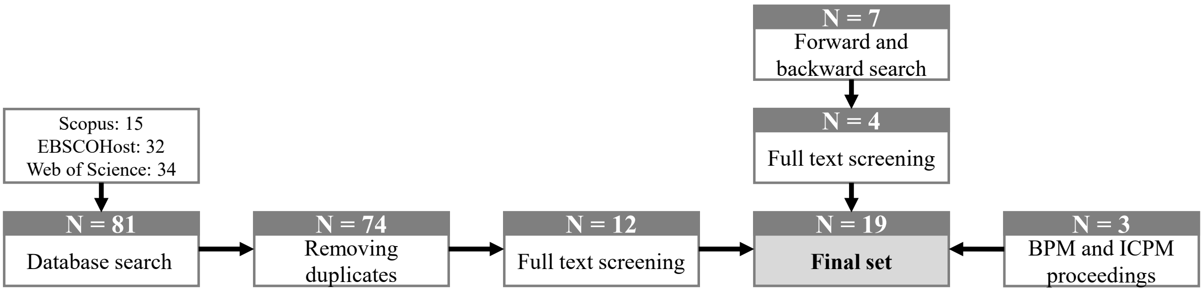

Figure 1. Overview of the Search Process.

4 Results

Table 1 displays a concept matrix with all identified PBPM techniques that position themselves as explainable. The respective concepts are purpose, intrinsic, post-hoc, evaluation objective, and evaluation method as presented in Section 2.

Maggi et al. (2014) presented one of the first PBPM techniques in 2014. It works on traditional *tree-based* methods which are intrinsically interpretable for humans. Explainability is not mentioned as a design criterion, but an example of the resulting decision tree is demonstrated. Breuker et al. (2016) published the first *probabilistic-based* next activity prediction technique based on a probabilistic finite automaton. They enable the visual pruning of the generated probabilistic model to make the control-flow more comprehensible to the business user. Senderovich et al. (2017) presented the first *regression-based* technique for remaining time prediction. They use both simple linear regression models, which are intrinsically interpretable, and ensemble learning methods such as RF and XGBoost. For the latter, they do not provide (post-hoc) explanation, and, in general, they do not define explainability as a design criterion.

Even though these early techniques can be considered *intrinsically* interpretable based on their relatively simple methods, the PBPM domain only started considering explainability as a design criterion for prediction techniques more recently. The remaining 15 techniques were all published within the last two years, 2019 and 2020, respectively.

Bukhsh et al. (2019) developed a *tree-based* technique to identify the maintenance-need of railway switches. They implemented it with simple decision trees, RF, and XGBoost on top of which they applied local interpretable model-agnostic explanation (LIME) (Ribeiro et al., 2016) for post-hoc explainability. Another outcome-oriented technique was presented by Rehse et al. (2019); they apply a *rule-based* technique (model-agnostic) in conjunction with a connection-weight *post-hoc* approach (model-specific). The technique outputs the confidence of the identified rules as a measure for the explanation's faithfulness. Verenich et al. (2019) presented another *regression-based* remaining time prediction technique that provides explanation both through the intrinsic structure and added post-hoc explainability with augmented control-flow models. Böhmer and Rinderle-Ma's (2020) technique called *LoGo* combines global *rule-based* explanations with local *probabilistic-based* interpretability, however, without mentioning the purpose of the explanations. Sindhgatta et al. (2020b) used feature importance values from XGBoost to provide global explanations and applied LIME for local, post-hoc explainability. Another outcome-oriented technique by Sachan et al. (2020) also applies a rule-based method and its usefulness is evaluated through a case study. Mehdiyev and Fettke (2020b), on the other hand, provide *post-hoc* interpretability by using *model-agnostic* partial dependence plots (PDPs).

	Purpose			Intrinsic				Post-hoc				Evaluation objective			Evaluation method			
Concept / Articles	Model validation/debugging	Knowledge discovery	unknown/ not specified	Rule-based	Regression-based	Tree-based	Probabilistic-based	Model-agnostic	Model-specific	Local	Global	Predictive quality of technique	Faithfulness of explanation	Applicability of explanation	Testing	Descriptive	Observational	Experimental
Maggi et al. (2014)			•			•						•			•	•		
Breuker et al. (2016)		•					•	•			•	•		•	•	•		
Senderovich et al. (2017)			•		•							•			•			
Bukhsh et al. (2019)		•				•		•		•	•	•		•	•	•		
Rehse et al. (2019)		•						•	•	•	•	•	•	•	•	•		
Verenich et al. (2019)		•			•			•		•		•		•	•	•		
Böhmer and Rinderle-Ma (2020)			•	•								•			•			
Brunk et al. (2020a)		•					•	•		•		•		•	•	•		
Galanti et al. (2020)		•						•		•		•		•	•	•	•	
Harl et al. (2020)		•							•	•				•		•		
Hanga et al. (2020)	•							•		•	•	•		•	•	•		
Jan et al. (2020)								•										
Mehdiyev and Fettke (2020a)		•						•		•		•	•	•	•	•		
Mehdiyev and Fettke (2020b)		•						•			•	•		•	•	•		
Pauwels and Calders (2020)			•				•					•			•			
Rizzi et al. (2020)	•							•		•		•		•	•	•		
Sachan et al. (2020)		•		•								•		•	•	•	•	
Sindhgatta et al. (2020b)		•						•		•	•			•		•		
Sindhgatta et al. (2020a)		•							•	•	•	•		•	•	•		
Weinzierl et al. (2020a)		•							•	•		•		•	•	•		

Table 1. Concept Matrix of Explainable Business Process Prediction Techniques.

The outcome-oriented technique from Rizzi et al. (2020) also provides *model-agnostic post-hoc* interpretability by applying LIME. It is the only work that states model debugging as the purpose of the explanation. Mehdiyev and Fettke (2020a) presented a technique for outcome prediction with post-hoc model-agnostic explanations. Similar to Rehse et al. (2019), the confidence of the identified rules is used to quantify the explanation's faithfulness for the user. Galanti et al. (2020) apply first shapley additive explanation (SHAP) (Lundberg and Lee, 2017) in a KPI-oriented (e.g., remaining time) prediction setting based on long short-term memory neural networks (LSTMs) (a type of recurrent deep neural networks). The technique is evaluated through a case study in a real-life setting.
In the area of next activity predictions, Pauwels and Calders (2020) and Brunk et al. (2020a) both apply context-sensitive dynamic Bayesian networks (DBNs), which represent *intrinsic probabilistic-based* approaches to XAI. However, Brunk et al. (2020a) go one step further and demonstrate the capabilities of *probabilistic-based* techniques for post-hoc explanation by applying an evidence sensitivity analysis on their model. Hanga et al. (2020) follow the approach of Breuker et al. (2016) to provide post-hoc explainability with an annotated process graph for model validation. Hanga et al. (2020), however, use LSTMs instead of a Bayesian network and also demonstrate local explanation with process graphs similar to Verenich et al. (2019). The only *post-hoc model-specific* prediction techniques were published by Harl et al. (2020), Sindhgatta et al. (2020a) and Weinzierl et al. (2020a). Harl et al. (2020) applied gated graph neural networks (a type of graph-based deep neural networks) in outcome predictions and provide an annotated process model as a local explanation to the user. Sindhgatta et al. (2020a) used attention-based LSTM models to extract attention weights for (local and global) post-hoc explanations. Weinzierl et al. (2020a) presented a layer-wise relevance propagation within an LSTM to enable local post-hoc explanation.

5 Discussion

Knowledge discovery is more common than model validation and debugging. Our research shows that most works specify *knowledge discovery* as the purpose of the explanation. The most common reasoning is that users need to understand the decision for adoption in practice. *Model validation and debugging* is a less common objective which might be due to our focus on IS literature. Yet, this finding makes sense as currently, PBPM techniques are directed towards process stakeholders such as process analysts. However, the trend is going from predictive to prescriptive techniques that provide recommendations (Gröger et al., 2014; Mehdiyev and Fettke, 2020b; Weinzierl et al., 2020b) or even proactively take actions in the system (Park and van der Aalst, 2020). To avoid false actions, approaches to validate and debug a model are crucial. Therefore, future research should consider providing explanations useful for data and AI experts (Mohseni et al., 2020). Sometimes the purpose of the explanation is not stated at all, which is in line with the observation that PBPM techniques are often designed rather for *all-purpose* use instead of specific use cases. Therefore, the requirements of the techniques and, subsequently, evaluation measures are generic as well. We think that more consideration should be given to the potential use cases which define, for example, the target audience or the objective of the DSS (Mehdiyev and Fettke, 2020a). Shmueli (2010) points out that explanatory and predictive modeling pursue different goals and therefore, "a priori determination of the main study goal as either explanatory or predictive is essential". Depending on the use case, explainability might not even be required.

Recent works prefer DL-based approaches over intrinsic models. The majority of techniques before 2020 applied intrinsically interpretable techniques combined with post-hoc explanation. However, in most recent publications, DL-based approaches seem to be the new normal. Most likely, this is because these techniques often provide better predictive quality when benchmarked against traditional approaches. However, if model explainability is a requirement, it is controversial whether small improvements in predictive quality are worthwhile when intrinsic interpretability is sacrificed as a consequence. Designers of PBPM techniques should state the purpose of a technique and, subsequently, evaluate the need for explainability. Future research should investigate methods to evaluate the trade-off between predictive quality and explainability to guide the design of techniques (Stierle et al., 2020).

Model-specific post-hoc explanations are underrepresented. Only four — very recent publications — apply *model-specific post-hoc* explanations to make their techniques interpretable. Given that more advanced DL-based techniques outperform simpler (*intrinsic*) techniques, the gained predictive quality can also be leveraged for better post-hoc interpretability features. Additionally, since model-specific techniques build upon model-internal characteristics, they can give more valuable insights for model validation and debugging purposes than model-agnostic approaches. Therefore, we call for more research that applies *model-specific* techniques in models with high predictive quality. In the ML literature, there exist several popular and effective *model-specific* methods (Barredo Arrieta et al., 2020), which are rather unseen in XPBPM, visualizing activations (Shrikumar et al., 2016) or modifying a deep neural network architecture (Hendricks et al., 2016), which could be adapted for usage in XPBPM.

Predictive quality is evaluated, explainability is only demonstrated. Most works evaluate their techniques in regards to *predictive quality*. Many authors argue that if their model can compete with state-of-the-art techniques regarding predictive quality, explainability is a welcome addition. However, with explainability being a design criterion, the evaluation of the techniques needs to evolve. The usefulness of the explanations for users needs to be assessed (Miller, 2019). Only two publications discuss the faithfulness of the explanations, as suggested by Du et al. (2019). Further aspects such as the robustness of the explanation are not discussed (Alvarez-Melis and Jaakkola, 2018). Most works at least demonstrate or describe how their techniques can be applied for explainability, e.g., by using the models created during the evaluation of the predictive quality. While this is a first step towards evaluating the applicability of XPBPM techniques, future work should assess the techniques in real-life settings with end-users. Both the design goals and the evaluation measures need to be chosen according to the target users of the explanation (Mohseni et al., 2020). Two works presented a case study applying a real-life data set, yet, without deploying the technique in the companies. So far, no work exists with an experimental evaluation method. In this regard Yang et al. (2019) proposed a framework that addresses the explanation evaluation of interpretable ML-based systems. They differentiate between the objective constructs of *generalizability* and *fidelity* and the subjective *persuasibility* of an explanation. Future endeavors should more rigorously distinguish between evaluating the predictive quality and the explanations of a technique.

6 Conclusion and Further Steps

Our research aims to provide researchers with an overview and classification of existing XPBPM techniques and raise awareness of essential concepts. The research field of XAI is a promising avenue for the broader adoption of ML-based DSSs in practice. Therefore, it is great to see the field of XPBPM develop so rapidly. However, a systematic approach is necessary to design and develop XPBPM techniques that advance the field. We are confident that our work supports this cause and will be useful for other researchers.

The subjective decisions that have to be made when conducting systematic literature reviews, such as selecting keywords, databases, and concepts, represent a limitation of our work. However, we minimized this bias by involving multiple authors in the search and coding process over multiple iterations. While our review focused on works mentioning explainability, we suppose that further intrinsically explainable techniques exist that do not explicitly state this design property or were designed without a reference to XAI.

Our next step will be to translate the findings of our review into a conceptual framework that will support the design of XPBPM techniques. We plan to emphasize the explanation's purpose and its receiver to provide adequate measures and methods to evaluate the techniques.

References

Adadi, A. and M. Berrada (2018). "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE Access* 6, 52138–52160. DOI: 10.1109/ACCESS.2018.2870052.

Alvarez-Melis, D. and T. S. Jaakkola (2018). *On the Robustness of Interpretability Methods*. arXiv: 1806.08049 [cs.LG].

Barredo Arrieta, A., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion* 58, 82–115. ISSN: 15662535. DOI: 10.1016/j.inffus.2019.12.012.

Böhmer, K. and S. Rinderle-Ma (2020). "LoGo: Combining Local and Global Techniques for Predictive Business Process Monitoring." In: *Advanced Information Systems Engineering*. Ed. by S. Dustdar, E. Yu, C. Salinesi, D. Rieu, and V. Pant. Cham: Springer International Publishing, pp. 283–298. ISBN: 978-3-030-49435-3.

Breiman, L. (2001). "Random Forests." *Machine Learning* 45 (1), 5–32.

Breuker, D., M. Matzner, P. Delfmann, and J. Becker (Dec. 2016). "Comprehensible Predictive Models for Business Processes." *MIS Quarterly* 40 (4), 1009–A9. ISSN: 02767783.

Brunk, J., M. Stierle, L. Papke, K. Revoredo, M. Matzner, and J. Becker (2020a). "Cause vs. effect in context-sensitive prediction of business process instances." *Information Systems* 95. ISSN: 0306-4379. DOI: 10.1016/j.is.2020.101635.

Brunk, J., J. Stottmeister, S. Weinzierl, M. Matzner, and J. Becker (2020b). "Exploring the effect of context information on deep learning business process predictions." *Journal of Decision Systems*, 1–16. DOI: 10.1080/12460125.2020.1790183.

Bukhsh, Z. A., A. Saeed, I. Stipanovic, and A. G. Doree (2019). "Predictive maintenance using tree-based classification techniques: A case of railway switches." *Transportation Research Part C: Emerging Technologies* 101, 35–54. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2019.02.001.

Chen, T. and C. Guestrin (2016). "Xgboost: A scalable tree boosting system." In: *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 785–794.

Cooper, H. M. (1988). "Organizing knowledge syntheses: A taxonomy of literature reviews." *Knowledge in society* 1 (1), 104.

Craven, M. W. (1996). "EXTRACTING COMPREHENSIBLE MODELS FROM TRAINED NEURAL NETWORKS." PhD thesis. Madison: University of Wisconsin - Madison.

Du, M., N. Liu, and X. Hu (2019). "Techniques for interpretable machine learning." *Communications of the ACM* 63 (1), 68–77.

Galanti, R., B. Coma-Puig, M. d. Leoni, J. Carmona, and N. Navarin (2020). "Explainable Predictive Process Monitoring." In: *Proceedings of the 2nd International Conference on Process Mining (ICPM)*, pp. 1–8. DOI: 10.1109/ICPM49681.2020.00012.

Gleicher, M. (2016). "A Framework for Considering Comprehensibility in Modeling." *Big data* 4 (2), 75–88. DOI: 10.1089/big.2016.0007.

Gröger, C., H. Schwarz, and B. Mitschang (2014). "Prescriptive analytics for recommendation-based business process optimization." In: *International Conference on Business Information Systems*. Springer, pp. 25–37.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi (2019). "A Survey of Methods for Explaining Black Box Models." *ACM Computing Surveys* 51 (5), 1–42. ISSN: 0360-0300. DOI: 10.1145/3236009.

Gunning, D. (2017). "Explainable artificial intelligence (XAI)." *Defense Advanced Research Projects Agency* 2.

Gunning, D., M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang (2019). "XAI-Explainable artificial intelligence." *Science robotics* 4 (37). DOI: 10.1126/scirobotics.aay7120.

Hanga, K. M., Y. Kovalchuk, and M. M. Gaber (2020). “A Graph-Based Approach to Interpreting Recurrent Neural Networks in Process Mining.” *IEEE Access* 8, 172923–172938.

Harl, M., S. Weinzierl, M. Stierle, and M. Matzner (2020). “Explainable predictive business process monitoring using gated graph neural networks.” *Journal of Decision Systems*, 1–16. DOI: 10.1080/12460125.2020.1780780.

Hendricks, L. A., Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell (2016). “Generating visual explanations.” In: *Proceedings of the 14th European Conference on Computer Vision*. Springer, pp. 3–19.

Hevner, A. R., S. T. March, J. Park, and S. Ram (2004). “Design science in information systems research.” *MIS quarterly* 28 (1), 75–105.

Jan, S. T. K., V. Ishakian, and V. Muthusamy (2020). “AI Trust in business processes: The need for process-aware explanations.” In: *Proceedings of the 32nd Innovative Applications of Artificial Intelligence Conference (IAAI-20)*. Vol. 34. 8, pp. 13403–13404. DOI: 10.1609/aaai.v34i08.7056.

Kvamme, H., N. Sellereite, K. Aas, and S. Sjursen (2018). “Predicting mortgage default using convolutional neural networks.” *Expert Systems with Applications* 102, 207–217. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2018.02.029.

Lipton, Z. C. (2018). “The mythos of model interpretability.” *Communications of the ACM* 61 (10), 36–43. ISSN: 0001-0782. DOI: 10.1145/3233231.

Lundberg, S. M. and S.-I. Lee (2017). “A Unified Approach to Interpreting Model Predictions.” In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., pp. 4765–4774.

Maggi, F. M., C. Di Francescomarino, M. Dumas, and C. Ghidini (2014). “Predictive Monitoring of Business Processes.” In: *Advanced Information Systems Engineering*. Ed. by M. Jarke, J. Mylopoulos, C. Quix, C. Rolland, Y. Manolopoulos, H. Mouratidis, and J. Horkoff. Cham: Springer International Publishing, pp. 457–472.

Marquez-Chamorro, A. E., M. Resinas, and A. Ruiz-Cortes (2018). “Predictive monitoring of business processes: A survey.” *IEEE Transactions on Services Computing* 11 (6), 962–977. ISSN: 19391374. DOI: 10.1109/TSC.2017.2772256.

Mehdiyev, N. and P. Fettke (2020a). “Explainable Artificial Intelligence for Process Mining: A General Overview and Application of a Novel Local Explanation Approach for Predictive Process Monitoring.” In: *Interpretable Machine Learning from Granular Computing Perspective*. Ed. by R. Navarro-Almanza, J. R. Castro, and M. A. Sanchez. Cham: Springer International Publishing. ISBN: 978-3-030-17985-4. DOI: 10.1007/978-3-030-17985-4_8.

— (2020b). “Prescriptive process analytics with deep learning and explainable artificial intelligence.” In: *Proceedings of the European Conference on Information Systems*. AISeL.

Miller, T. (2019). “Explanation in artificial intelligence: Insights from the social sciences.” *Artificial Intelligence* 267, 1–38. ISSN: 0004-3702. DOI: https://doi.org/10.1016/j.artint.2018.07.007. URL: https://www.sciencedirect.com/science/article/pii/S0004370218305988.

Mohseni, S., N. Zarei, and E. D. Ragan (2020). *A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems*. arXiv: 1811.11839 [cs.HC].

Molnar, C. (2020). *Interpretable Machine Learning* -. Raleigh, North Carolina: Lulu.com. ISBN: 978-0-244-76852-2.

Murdoch, W. J., C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu (2019). “Definitions, methods, and applications in interpretable machine learning.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (44), 22071–22080. DOI: 10.1073/pnas.1900654116.

Okoli, C. and K. Schabram (May 2010). “A Guide to Conducting a Systematic Literature Review of Information Systems Research.” *SSRN Electronic Journal* 10. DOI: 10.2139/ssrn.1954824.

Paré, G., M.-C. Trudel, M. Jaana, and S. Kitsiou (2015). “Synthesizing information systems knowledge: A typology of literature reviews.” *Information & Management* 52 (2), 183–199.

Park, G. and W. M. van der Aalst (2020). “A General Framework for Action-Oriented Process Mining.” In: *International Conference on Business Process Management*. Springer, pp. 206–218.

Pauwels, S. and T. Calders (2020). “Bayesian Network Based Predictions of Business Processes.” In: *Business Process Management Forum*. Ed. by D. Fahland, C. Ghidini, J. Becker, and M. Dumas. Cham: Springer International Publishing, pp. 159–175.

Poll, R., A. Polyvyanyy, M. Rosemann, M. Röglinger, and L. Rupprecht (2018). “Process forecasting: Towards proactive business process management.” In: *Proceeding of the 16th International Conference on Business Process Management (BPM)*. Springer, pp. 496–512.

Rehse, J.-R., N. Mehdiyev, and P. Fettke (2019). “Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory.” *Künstliche Intelligenz* 33 (2), 181–187.

Ribeiro, M. T., S. Singh, and C. Guestrin (2016). ““Why should I trust you?” Explaining the predictions of any classifier.” In: *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144.

Rizzi, W., C. Di Francescomarino, and F. M. Maggi (2020). “Explainability in Predictive Process Monitoring: When Understanding Helps Improving.” In: *Business Process Management Forum*. Ed. by D. Fahland, C. Ghidini, J. Becker, and M. Dumas. Cham: Springer International Publishing, pp. 141–158.

Rogge-Solti, A. and M. Weske (2013). “Prediction of remaining service execution time using stochastic petri nets with arbitrary firing delays.” In: *International conference on service-oriented computing*. Springer, pp. 389–403.

Sachan, S., J.-B. Yang, D.-L. Xu, D. E. Benavides, and Y. Li (2020). “An explainable AI decision-support-system to automate loan underwriting.” *Expert Systems with Applications* 144, 113100. ISSN: 0957-4174. DOI: `https://doi.org/10.1016/j.eswa.2019.113100`.

Senderovich, A., C. Di Francescomarino, C. Ghidini, K. Jorbina, and F. M. Maggi (2017). “Intra and inter-case features in predictive process monitoring: A tale of two dimensions.” In: *Proceedings of the 15th International Conference on Business Process Management*. Springer, pp. 306–323.

Shmueli, G. (2010). “To Explain or to Predict?” *Statistical Science* 25 (3), 289–310. DOI: `10.1214/10-STS330`. URL: `https://doi.org/10.1214/10-STS330`.

Shrikumar, A., P. Greenside, A. Shcherbina, and A. Kundaje (2016). “Not just a black box: Learning important features through propagating activation differences.” *arXiv preprint arXiv:1605.01713*.

Sindhgatta, R., C. Moreira, C. Ouyang, and A. Barros (2020a). “Exploring Interpretable Predictive Models for Business Processes.” In: *Proceedings of the 18th International Conference on Business Process Management (BPM)*. Ed. by D. Fahland, C. Ghidini, J. Becker, and M. Dumas. Cham: Springer International Publishing, pp. 257–272.

Sindhgatta, R., C. Ouyang, and C. Moreira (2020b). “Exploring interpretability for predictive process analytics.” In: *Proceedings of the 18th International Conference on Service-Oriented Computing (ICSOC)*. Springer International Publishing.

Stierle, M., S. Zilker, S. Dunzer, J. Tenschert, and G. Karagegova (2020). “Design Principles for Comprehensible Process Discovery in Process Mining.” In: *Proceedings of the European Conference on Information Systems*.

van der Aalst, W., M. H. Schonenberg, and M. Song (2011). “Time prediction based on process mining.” *Information systems* 36 (2), 450–475.

Verenich, I., M. Dumas, M. La Rosa, and H. Nguyen (2019). “Predicting process performance: A white-box approach based on process models.” *Journal of Software: Evolution and Process* 31 (6). e2170 smr.2170, e2170. DOI: `10.1002/smr.2170`.

Wanner, J., K. Heinrich, C. Janiesch, and P. Zschech (2020). “How Much AI Do You Require? Decision Factors for Adopting AI Technology.” In: *International Conference on Information Systems (ICIS)*, pp. 1–17.

Webster, J. and R. T. Watson (2002). “Analyzing the past to prepare for the future: Writing a literature review.” *MIS Quarterly*, xiii–xxiii.

Weinzierl, S., S. Zilker, J. Brunk, K. Revoredo, M. Matzner, and J. Becker (2020a). "XNAP: Making LSTM-based Next Activity Predictions Explainable by Using LRP." In: *4th International Workshop on Artificial Intelligence for Business Process Management 2020, Sevilla, Spain*.
Weinzierl, S., S. Zilker, M. Stierle, G. Park, and M. Matzner (Mar. 8–11, 2020b). "From predictive to prescriptive process monitoring: Recommending the next best actions instead of calculating the next most likely events." In: *Proceedings of the 15th International Conference on Wirtschaftsinformatik* (Potsdam, Germany), pp. 1–5. DOI: 10.30844/wi_2020_c12-weinzierl.
Yang, F., M. Du, and X. Hu (2019). *Evaluating Explanation Without Ground Truth in Interpretable Machine Learning*. arXiv: 1907.06831.