Received March 7, 2018, accepted April 19, 2018, date of publication April 30, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2831244

Systematic Mapping of Process Mining Studies in Healthcare

TUGBA GURGEN ERDOGAN AND AYCA TARHAN
Computer Engineering Department, Hacettepe University, Ankara 06800, Turkey
Corresponding author: Tugba Gurgen Erdogan (tugba@cs.hacettepe.edu.tr)

ABSTRACT In the last decade, as an emerging technique for business processes management, process mining (PM) has been applied in many domains, including manufacturing, supply-chain, government, healthcare, and software engineering. Particularly in healthcare, where most processes are complex, variable, dynamic, and multi-disciplinary in nature, the application of this technique is growing yet challenging. Several literature reviews, as secondary studies, reveal the state of PM applications in healthcare from different perspectives, such as clinical pathways, oncology processes, and hospital management. In this article, we present the results of a systematic mapping (SM) study which we conducted to structure the information available in the primary studies. SM is a well-accepted method to identify and categorize research literature, in which the number of primary studies is rapidly growing. We searched for studies between 2005 and 2017 in the electronic digital libraries of scientific literature, and identified 172 studies out of the 2428 initially found on the topic of PM in healthcare. We created a concept map based on the information provided by the primary studies and classified these studies according to a number of attributes including the types of research and contribution, application context, healthcare specialty, mining activity, process modeling type and notation/language, and mining algorithm. We also reported the demographics and bibliometrics trends in this domain; namely, publication volume, top-cited papers, most contributing researchers and countries, and top venues. The results of mapping showed that, despite the healthcare data and technique related challenges, the field is rapidly growing and open for further research and practice. The researchers who are interested in the field could use the results to elicit opportunities for further research. The practitioners who are considering applications of PM, on the other hand, could observe the most common aims and specialties that PM techniques are applied.

INDEX TERMS Clinical pathway, healthcare process, process management, process mining, systematic mapping.

I. INTRODUCTION

Being heavily human-oriented and knowledge-intensive, healthcare processes and their management have a direct impact on the quality of healthcare services and related costs [1]. The complex, variable, dynamic and multi-disciplinary nature of healthcare makes process management in this domain a challenging task that should be addressed only with the right tools and techniques [2]. Process mining (PM), as a relatively young discipline for business process management, has become increasingly popular in the last decade [3] and has been applied in various domains including manufacturing, supply-chain, government, healthcare, and software engineering [4]–[6]. PM is primarily utilized in the discovery, conformance checking, and enhancement of business processes based on event logs [7]. In the healthcare domain, the application of this technique is promising due to the growing number of reported studies, yet remains challenging [8].

In cases where the number of primary studies (e.g., experience reports) in a specific area is rapidly and significantly growing, it is useful to summarize the body of knowledge by undertaking secondary study (e.g., literature review) to identify opportunities for further research [9]. A secondary study aggregates and synthesizes the content of primary studies based on a specific purpose [10]. Several literature reviews have revealed the state of PM applications in the healthcare domain in general [11]–[13] or from specific perspectives, such as clinical pathways [14] and oncology processes [15]. However, none of these secondary studies provides an extensive and aggregated map on the topic of PM in healthcare to structure the information available in the literature and highlight opportunities for further research and practice.

Based on the gap stated above, in this article, we present the results of secondary study we performed by applying systematic mapping (SM) [9] as research methodology. We searched manually for the studies published between 2005 and 2017 in the following electronic digital libraries of scientific literature (in alphabetical order): ACM, Google Scholar, Emerald, IEEE Explore, Pubmed, ScienceDirect, Scopus, SpringerLink, Web of Science, and Wiley. Of the 2428 publications found in the area, 172 were selected for a thorough review. We systematically developed a concept map of the information provided by primary studies on PM applications in healthcare and then used this map to classify the studies based on a number of attributes; namely, the types of research and contribution, application context, mining activity, process modeling type and notation, and mining algorithm. To the best of our knowledge, this is the only systematic mapping study that provides a broad and general overview of the studies on this topic. More specifically, the main contributions of this research are:

- A general classification scheme that structures the field of research on PM applications in healthcare,
- A systematic map of 172 primary studies based on the classification scheme,
- An analysis of the demographic trends and bibliometrics of the primary studies, and
- An analysis of the results, challenges, and opportunities of the field for further research and practice.

The remaining of this article is organized as follows: Section 2 summarizes the background on PM and the related works that have reviewed the applications of this technique in the healthcare domain. Section 3 overviews the research design of the current systematic mapping study by presenting research questions, details about publication selection process, and potential threats to the validity of this research. Section 4 describes the classification scheme developed iteratively to analyze and categorize the selected primary studies. Section 5 presents the results of the current research in relation to the formulated questions. Section 6 includes a summary of our findings together with related challenges. Finally, Section 7 concludes the article by providing overall results and plans for future work.

II. PROCESS MINING AND RELATED WORKS

A. OVERVIEW OF PROCESS MINING

PM is a process management technique that exploits event data on information systems. In the last decade, this technique has been increasingly used for business process management [3]. It can be considered as an analysis technique for data mining and machine learning on one side and process modeling on the other. PM is applied for three main aims [7]; i) process discovery, ii) conformance checking of process implementations according to the discovered/modeled process, and iii) enhancement of process by detecting the differences in process implementations.

Various automations of information systems store detailed information about the identity and sequence of actual activities performed during the execution of business processes [3]; e.g., business process management, enterprise resource planning, product data management and capacity requirement planning systems. This information is called event logs which is the starting point of PM [7]. In these logs, each event symbolizes one activity and each activity is a part of a process. Event logs store detailed information on events concerning the source; i.e., the person or tool that started and performed the activity, the starting and finishing time of the activity, and the data element; e.g., type, size, and comments.

Process discovery is the most important activity of the PM since it provides a base for further analyses involving the application of the remaining two types of PM, conformance checking and enhancement. In process discovery, event logs are used as input and a process model is set up without prior information [7]. When an actual process is created from event logs, many organizations can face challenges due to the differences in theory and implementation. The algorithms that are widely used for process discovery include Alpha Miner, Heuristic Miner, and Fuzzy Miner [5].

In *conformance checking*, the process model and its flow discovered from the event logs are analyzed and it is checked whether the process has been carried out as identified in the model [7]. Conformance checking measures the differences between the performed process and the process model specifications. The main aim of this process is to identify the areas that need improvement using the information gained from the actual process.

Process enhancement is the improvement of the process model based on event data. This can be undertaken by adding further data perspectives to the process model using event data, a process also known as extension. Another type of enhancement is repair, in which the quality of the process model is improved using event data and a new repaired model is defined [16].

In the PM technique, there are four perspectives; control-flow, organizational, case, and time [7]. These perspectives are not isolated and are all related to each other. The control-flow perspective focuses on the sequences of activities and the discovery of the process model and aims to find the best definitions for all possible paths. The results are represented by petri net [17] and BPMN [18]. The organizational perspective concerns the actors (human, system, and role) and the relations between these actors to classify the process model based on roles and organization. The case perspective focuses on the definition of cases and the factors which influence the real data. The time perspective is concerned with the occurrence time and frequencies of events and helps identify bottlenecks, measure service levels, track the use of resources, and predict the remaining time for ongoing events.

Various software products have PM capabilities [5], e.g., ARIS Process Performance Manager (Software AG) [19], Disco (Fluxicon) [20], ProM (TU/e) [21], PALIA-ER [22], CELONIS [23], and pMineR [24]. ProM [21] is basic PM tool, which provides a standard environment incorporating a generic open-source framework for implementing

TABLE 1. A list of secondary studies on process mining in the healthcare domain.

Type	Name of Article	# Primary Studies	Year	Ref
Survey	On process mining in health care	10	2012	[26]
Literature Review	Process mining for clinical pathway: Literature review and future directions	37	2014	[14]
Survey	Clinical processes and its data, what can we do with them	27	2015	[27]
Literature Review	Process mining in healthcare: Opportunities beyond the ordinary	40	2015	[11]
Literature Review	Process mining in healthcare: A literature review	74	2016	[12]
Systematised Lit. Review	Process mining in healthcare: A systematised literature review	168	2016	[13]
Literature Review	Process mining in oncology: A literature review	37	2016	[15]
Literature Review	Process mining for healthcare process analytics	11	2016	[28]
Systematic Mapping	Systematic mapping of process mining studies in healthcare	172	2018	This study

PM techniques. Disco [20] is a popular PM toolkit, which is powerful, easy-to-use, and fast. The revolutionary commercial PM technology in Disco helps researchers to create visual maps from process data in minutes.

B. RELATED WORKS

Although, PM has the potential to offer many advantages to healthcare professionals [8], to date, only a limited number of studies have reviewed the evidence on PM applications in analyzing and improving healthcare processes. We found nine secondary studies that reviewed primary studies in this area. Table 1 presents a list of these secondary studies (our study shown in the last row) including the type and title of the study, number of primary studies involved, and publication year.

Kaymak *et al.* [26] discussed the lack of suitability of the PM techniques for analyzing healthcare processes and examined researchers' experience on the potential use of PM methods.

They argued that the existing methods failed to identify good process models and provided a number of recommendations for applying the PM techniques. Yang and Su [14] reviewed 37 primary studies on PM for clinical pathways, specifically for the purposes of process discovery for clinical pathways design, variants analysis and control, and continuous evaluation and improvement. The authors analyzed the weaknesses of these studies, highlighted many challenges of the PM techniques in a medical environment, and suggested four areas of improvement: variants identification and analysis, customization of clinical pathways, self-learning improvement of clinical pathways, and integrated medical process management. Mans et al. [11] identified 40 primary studies that included real-life PM applications and categorized these studies based on their focus on process discovery or conformance checking. They gave examples of healthcare contexts with regard to the use of these two types of PM activities. In contrast to previous published works, Mans et al. structured a comprehensive overview of healthcare data as a healthcare reference model comprising together with a list of data quality issues. The authors also evaluated research gaps and challenges in the application of existing PM methods to the healthcare domain. Rojas et al. [12] extended their bibliographic survey [27] by conducting a literature review on the use of PM techniques in healthcare with the primary focus being on the healthcare process rather than clinical pathways. They reported the main characteristics of the reviewed studies and the emerging topics of the field. Ghasemi and Amyot [13] conducted a systematised literature survey which provides an overview of the status of PM in healthcare, and provided initial results from the search process of the systematic review. The authors identified 168 primary studies after deleting duplicates and did not explicitly specify inclusion and exclusion criteria for eliminating the primary studies. Kurniati et al. [15] reviewed 37 primary studies that applied PM techniques in oncology through a bibliographic survey. They classified limitations with regard to the use of PM techniques in oncology as: data, technique, and team.

Our study makes contributions over secondary studies listed in Table 1 in that it presents up-to-date state and maturity of research in the field based on a classification including a comprehensive set of attributes that were elicited from a large number of primary studies. We reviewed 172 studies and mapped their content based on the classification scheme we developed iteratively during the review process. This study is therefore generic and not specific to healthcare specialty or healthcare context, with an aim to enable derivation of an extensive classification scheme which would serve as reference to structure the information available in the primary studies. It is also an extension over our previous work [28] in which we had focused only on applications of PM for conformance verification studies in healthcare.

III. RESEARCH DESIGN

We applied systematic mapping as the research methodology. SM is a well-accepted method in software engineering to identify and categorize research literature, especially where the number of primary studies is increasingly growing [9]. The SM method requires an established search protocol and rigorous inclusion and exclusion criteria for the screening and selection of relevant publications. It focuses on systematically developing a classification scheme and using it to classify the content of the selected studies. Typically, the results of an SM study show the frequencies of the selected studies within the classification scheme with respect to the investigated concepts, which allows identifying trends, weaknesses, and opportunities in the research area of concern. Consequently, we found that many of its steps were also useful in the multi-disciplinary area of process mining for healthcare.

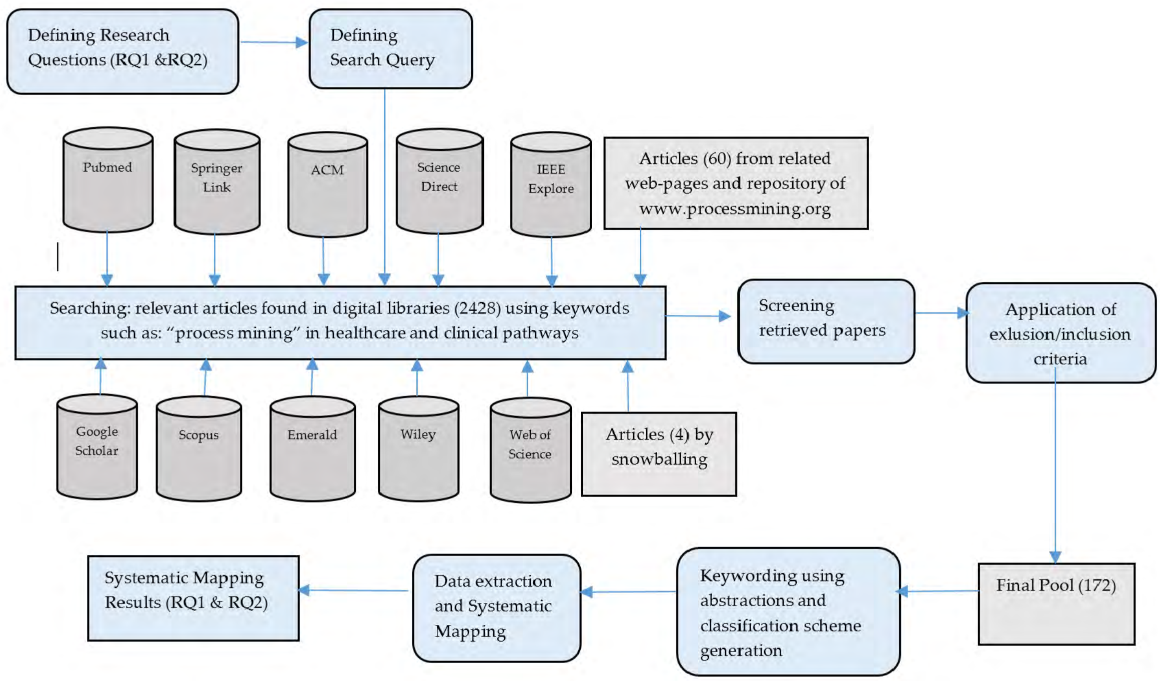


FIGURE 1. The SM process used in this study.

Fig. 1 presents the SM process used in this study as derived from the guidelines and process proposed by Petersen *et al.* [9]. Our process consists of the following steps:

- Defining research questions (RQs),
- Defining search query and searching for papers,
- Screening the retrieved papers, which results in a set of relevant papers,
- Keyword extraction from abstracts, which results in a classification scheme,
- Data extraction and SM.

The process started with defining the research questions and search query. After searching the queries in 10 digital libraries, reviewing the relevant websites, and using snowballing techniques, we gathered 2428 potentially relevant publications. Snowballing techniques are used to identify additional relevant articles through the reference lists of the articles found [29]. For screening the retrieved studies, we developed and applied inclusion/exclusion criteria and obtained a final pool of 172 studies. Then, a classification scheme was iteratively created to categorize the selected studies by keyword extraction from abstracts and full texts. After extracting the data from the studies, the results of SM were analyzed.

The remainder of this section concerns the research questions, paper selection process, and potential threats to validity.

A. RESEARCH QUESTIONS

The aim of this systematic mapping study was to identify, analyze and synthesize the quantity, type, and results of PM research in the healthcare domain. Thus, we set two main goals related to our research: 1) To systematically review the related scientific papers in the field for mapping and 2) to present bibliometrics and demographic analysis of the field. These goals led to the definition of the following research questions (RQ 1 and RQ2) which were then elaborated with a number of sub-questions:

RQ1 Systematic Mapping: What is the research space in the literature concerning PM applications in the healthcare domain? The RQ1 sub-questions were;

RQ1.1 Research Type: What was the type of research method used in the study? The empirical level of studies was determined using Petersen *et al.* [30] based on solution proposal, validation research, evaluation research, and experience paper.

RQ1.2 Contribution Type: What was the main contribution of the study to the field? How many papers have contributed a method, tool, metric, model or process?

RQ1.3 Application Context: In which context was the study carried out? The application context may be healthcare process or clinical pathway, department, or hospital.

RQ1.4 Healthcare Specialty: In which specialty was the study carried out? Which healthcare specialty was subject to more attention? The healthcare specialty may be a medical treatment process, such as skin cancer treatment or organizational healthcare process; e.g., general surgery, outpatient, and emergency department processes.

RQ1.5 Type of Process Mining Activity: What was the type of PM activity applied; e.g., process discovery, process conformance checking, or process enhancement? There are also other PM activities such as process variant analysis, performance analysis, outlier detection, and predictive monitoring, which can be used for further analysis.

RQ1.6 Process Modeling Type: How was the process model created? We expect that some studies create a process model manually, some discover it automatically using PM techniques, and others use a combination of the two.

RQ1.7 Modeling Notation/Language: Which modeling notation was used to create the process model?

RQ1.8 Process Mining Techniques:

a) Which techniques were used to discover the process model from its executions?
b) Which techniques were used to check the conformance between the process model and its executions?
c) Which techniques were used to enhance (repair or extend) the process model?
d) Which other PM techniques were used to analyze healthcare data?

RQ1.9 Clustering Techniques: Which clustering techniques were used to analyze healthcare data before applying PM techniques?

RQ2 Trends and Demographics of the Publications: The following set of sub-questions were formulated by reviewing the existing bibliometrics studies:

RQ2.1 Publication Count by Year: What is the yearly number of publications in the field?

RQ2.2 Top-Cited Publications: Which publications have been most cited by others?

RQ2.3 Most Contributing Researchers: Who are the most contributing researchers in the area as measured by the number of publications?

RQ2.4 Most Contributing Countries: What are the most contributing countries in the area as measured by affiliations of researchers?

RQ2.5 Top Venues: Which venues (journals or conferences) are the main targets of publications?

B. PUBLICATION SELECTION PROCESS

The literature search was performed for the studies that were published in academic journals and conference proceedings from 2005 to 2017 and available in the digital libraries of (in alphabetical order) ACM, Emerald, IEEE Explore, Pubmed, ScienceDirect, Scopus, SpringerLink, Web of Science, and Wiley. In order to select the relevant papers, we used the following final query: (''process mining'' OR ''workflow mining'') AND (''health'' OR ''care'' OR ''healthcare'' OR ''hospital'' OR ''clinic'' OR ''clinical'' OR ''pathways''). We identified these keywords by both focusing on healthcare processes and clinical pathways for process mining, and also used word embedding techniques. When we checked other phrases like sequential pattern mining, temporal data mining, and care flow mining, there was no effect on the search results.

Table 2 shows the number of studies that were initially retrieved, initially selected, and uniquely selected from the digital libraries for search keywords.

TABLE 2. Number of studies initially retrieved and selected by search query.

Digital Library	#Initially Retrieved	#Initially Selected	#Uniquely Selected
Pubmed	31	24	24
Springer Link	600	54	42
ACM	200	17	16
Science Direct	456	25	17
IEEE	40	32	23
Google Scholar	500	108	22
Scopus	271	107	19
Emerald	17	0	0
Wiley	164	2	0
Web of Science	149	78	0
	2428	**447**	**163**

To ensure the inclusion of all the relevant studies, after searching the digital libraries, we investigated the following sources:

- References of papers already included in the pool using the backward and forward snowballing methods [29],
- ResearchGate web pages of researchers,
- Three specific websites (processmining.org, processmining4healthcare.org and bpmcenter.org),
- Top venues as elicited from published papers.

We recorded the source of each paper (digital library name, website, or snowballing) in an Excel sheet to manage and improve the search process. The initial search and data extraction were undertaken by the first author and then selectively reviewed by the second author to ensure that the research process was followed correctly. Based on the review feedback from the second author, the steps and outputs of search and data extraction processes were refined.

Of the 2428 studies that were initially gathered, 447 were considered relevant to our purpose. Fifty-nine papers were extracted from the repository of the Health Analytics by Process Mining Group from the Eindhoven University of Technology and four studies were added using the snowballing method. As a result, after eliminating the duplicates, 172 studies were selected (163 from digital libraries, additional four from processmining.org, another one from processmining4health-care.org, and four from snowballing) for a thorough analysis. The exclusion criteria were: (i) publication language being other than English; (ii) full-text not being available; (iii) books and theses; and (iv) duplicate results from different search methods.

C. POTENTIAL THREATS TO VALIDITY

We systematically identified and addressed potential threats to four types of validity of our research based on the guidelines of systematic literature review and mapping studies [9], [31] and below, we describe the steps that we took to minimize or mitigate these threats as adopted from [32].

1) INTERNAL VALIDITY

Limitation of search terms and search engines can lead to an incomplete set of primary sources. In order to mitigate the risk of finding irrelevant studies, search was undertaken using defined keywords, followed by a manual search among the references in the initial pool and the ResearchGate pages of most contributing researchers in the field of study. To minimize threats that may result from search engines, we not only included comprehensive academic databases, such as Google Scholar and Pubmed but also searched special active venues and webpages related to PM. We recorded in an Excel sheet each paper found with its source. In cases where several sources returned the same paper, all sources were noted. Therefore, we believe that an adequate and inclusive basis was established for this study and if there was any missing publication, the rate would be negligible.

2) CONSTRUCT VALIDITY

In this study, threats related to this type of validity concerned the suitability of RQs and the categorization scheme used for data extraction. The research questions were specifically designed for the defined goal and different aspects of PM in healthcare. The questions were systematically answered according to a categorization scheme and finalized through several iterative improvement processes. In addition, we consulted two medical professionals when grouping healthcare specialties.

3) CONCLUSION VALIDITY

In order to ensure the reliability of our treatments, the entire pool of primary sources was analyzed and the data were reviewed, extracted, and synthesized by the first author in iterations according to a research protocol, and the whole process and all selected outputs were reviewed by the second author. In addition, following the guidelines of a systematic mapping approach and procedure ensured replicability of this study and that the results would not significantly deviate from those of other similar studies.

4) EXTERNAL VALIDITY

Defining search terms in the source selection approach resulted in obtaining only primary sources written in English language. However, the main issue concerns whether the selected works represent all types of literature in the area of study, and we consider that the relevant studies collected in the study pool contained sufficient information to represent the entire related literature.

IV. CLASSIFICATION SCHEME

To conduct systematic mapping, a stable classification scheme needs to be derived. Table 3 shows the research questions, attributes of concern for each RQ, and possible types and descriptions of each attribute as iteratively elicited from the selected studies.

As discussed by Petersen *et al.* [30], a research facet denotes the type of research approach used in each paper. This corresponds to RQ1.1 in this study. We adopted the following research facets: solution proposal, validation research, evaluation research, and experience papers. Publications that present a solution proposal and its simple example are categorized as solution proposal papers. The papers that contain validation sections with a weak empirical study are considered as validation research. If the proposed technique in a study is extensively evaluated using empirical methods (e.g., case study, comparison of existing methods) and its advantages and disadvantages are discussed, it is categorized as evaluation research. Papers that report only applications or experiences in practice are called experience papers.

RQ1.2 concerns the contribution type attribute, which denotes the type of contribution(s) proposed in each paper and can be one of the following: method/technique, tool, model, metric, process, or other [30]. For example, some studies presented new techniques, some proposed process for applying existing PM techniques, and others applied existing PM techniques to healthcare in a case study. Accordingly, we adopted the contribution facet attributes of method, tool, model, metric (similarity/distance), process, or empirical results (e.g., a case study).

RQ 1.3 is related to the application context of the PM studies considering several aspects, such as healthcare process, clinical pathway, single/multiple department(s), and single/multiple hospital(s).

RQ 1.4 is about the healthcare specialty, which denotes the name of healthcare process or clinical pathway. We extracted the name(s) of healthcare process(es), disease(s) or department(s) that were analyzed in each paper using PM techniques. These groups were reviewed by two medical professionals and revised when necessary.

RQ 1.5 concerns the type of PM activity; in particular, the PM activities of process discovery, conformance checking and enhancement in addition to some other activities including process variant analysis, performance analysis, predictive monitoring, and outlier detection.

RQ 1.6 is about the process modeling type, which can have one of the following values: defined manually with a modeling notation/language, automatically discovered using PM techniques, created manually after having been discovered by PM techniques, or not specified. This sub-RQ investigates how process model is obtained using PM techniques.

The modeling notation/language attribute corresponds to RQ1.7 that questions the modeling notation of healthcare process, which can be BPMN, flowchart, UML diagrams (state, class or activity diagrams), or formal process model such as petri net, heuristics net, fuzzy model, or declare model.

To answer RQ 1.8, we extracted the names of the process discovery, process conformance and process enhancement techniques used in each paper. We also extracted the names of the other PM techniques which were used to analyze healthcare data from several aspects, such as performance, data quality and data visualization, and by combinations of the PM techniques to obtain better results in the healthcare domain.

RQ 1.9 is related to the clustering techniques, which are used to cluster a healthcare process before applying the PM techniques. We elicited the names of the clustering techniques as well as the related ProM plugins.

TABLE 3. Classification scheme.

RQs	Attribute	Possible Types	*	Description of Types	**
RQ1.1	Research type	Solution proposal	B	description of a new solution and its applicability showed by a small example or a good line of argumentation	S
		Validation research	B	novel or practically not implemented techniques that are used for example experiments in the lab or weak empirical study	
		Evaluation research	B	evaluation of techniques are conducted in terms of implementing a solution and evaluating implementation	
		Experience paper	B	personal experiences of authors in implementing PM techniques in healthcare domain	
RQ1.2	Contribution type	Method	B	newly proposed method	M
		Tool	B	newly proposed tool that is implemented for healthcare domain	
		Metric (similarity/distance)	B	description of a new metric which can be distance or similarity	
		Process	B	definition of steps for analysis with PM techniques	
		Model	B	newly proposed model which represent healthcare data or process	
		Empirical results	B	using existing PM techniques, tools, or methodologies	
RQ1.3	Application context	Healthcare process	B	the study is applied to organizational hospital process or medical treatment process, and to healthcare data of a single department, multiple departments, a single hospital, or multiple hospitals.	M
		Clinical pathway	B		
		Single department	B		
		Multiple department	B		
		Single hospital	B		
		Multiple hospital	B		
RQ1.4	Healthcare specialty	Oncology, outpatient, surgery, emergency department, neurologic diseases, cardiovascular diseases, radiological test, respiratory diseases, diabetes, urological diseases, dental, intensive care unit process, ophthalmology, and etc.	S	name of healthcare specialty, if specified	M
RQ1.5	Type of process mining activity	Process discovery	B	discovery of healthcare process from event logs	M
		Process conformance checking	B	checking conformance of process executions to process model	
		Process enhancement	B	enhancement of process model based on event data by extending with additional perspectives or improving the quality of the model using event data	
		Process variant analysis	B	discovering process variants and identifying the most common ones	
		Performance analysis	B	process performance analysis for identifying and quantifying bottlenecks	
		Predictive monitoring	B	predictions (predicting violations) or recommendations (giving early advice) based on executions	
		Outlier detection	B	detection of abnormal values in event logs	
RQ1.6	Type of process modeling	Defined manually with a modeling notation	B	process model is manually described in a modeling notation	S
		Automatically discovered by using PM technique	B	process model is discovered automatically as a formal process model	
		Created manually after discovered	B	discovered process model is altered or corrected manually by human intervention	
		Modeling type not specified	B	there is no definition about process modeling type	
RQ1.7	Modeling notation/ language	Heuristic net, petri net, declare model, fuzzy model, BPMN, flowchart, UML diagrams, TPA, dependency graph, decision tree, social networks and etc.	S	name of notation/language used for process modeling	M
		Newly proposed notation	B	newly proposed notation to visualize healthcare process	M
		Modeling notation not specified	B	name of modeling notation is not stated in the paper	S
RQ1.8	a) Process discovery techniques	Heuristic miner, fuzzy miner, alpha/alpha++ miner, inductive visual miner, social network miner, genetic process miner, declare miner, multi-phase miner, guide tree miner, role hierarchy miner and etc.	S	name of PM technique used for process discovery	M
		Newly proposed technique	B	newly proposed PM technique used to discover process	M
		Not specified	B	name of PM technique is not stated in the paper	S
RQ1.8	b) Conformance checking techniques	LTL checker, trace alignment, conformance checker, delta analysis, replay a log on petri net, declare checker, CPN model checking, fuzzy animation	S	name of PM technique used for process conformance checking	M
		Newly proposed technique	B	newly proposed PM technique used for process conformance checking	M
		Not specified	B	name of PM technique is not stated in the paper	S

TABLE 3. *(Continued.)* **Classification scheme.**

RQ1.8	c) Process enhancement techniques	Declare repair, performance analysis with petri net, simulation techniques, extend declare map with correlation/timestamp	S	name of PM technique used for process enhancement	M
		Newly proposed technique	B	name of PM technique is not stated in the paper	
		Not specified	B	newly proposed technique is used for process enhancement	S
RQ1.8	d) Other techniques	Dotted chart, performance analysis with petri net, performance sequence diagram, pattern abstraction	S	name of other performance analysis PM techniques used for further analysis of the case-related information about the care processes	M
RQ1.9	Clustering techniques	Trace clustering, k-means clustering, agglomerative clustering, self-organizing maps, sequence clustering, hierarchical clustering, activity clustering, frequent pattern mining	S	name of clustering technique used for process clustering	M
		Newly proposed technique	B	newly proposed method is used for clustering especially by using newly proposed metric	
RQ2	Trends and Demographics	Year	N	publication count by year	S
		Number of Citations	N	top-cited papers	
		Authors	S	name of all most contributing researchers	
		Authors' country	S	country from authors	
		Venue	S	top venues where the paper is published	

* B=Boolean, N=Numberic, S=String ; ** M=Multiple Values, S=Single Value

The trends and demographics attribute regarding the types defined by RQs 2.1 to 2.5 relate to the demographic and bibliometric information of papers and include the year, number of citations, most contributing researchers and countries, and top venues (i.e., journals and conferences).

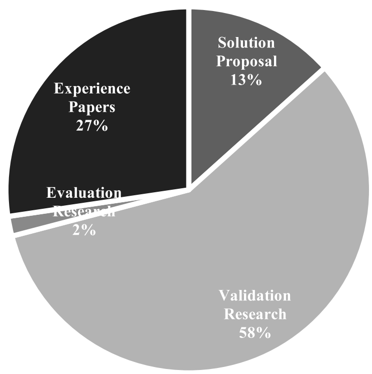
V. SYSTEMATIC MAPPING RESULTS

In this section, we present and discuss the results of the systematic mapping study in relation to RQ1 and RQ2. All papers included in systematic mapping are shared in a publicly accessible repository: https://goo.gl/fC5Ddw

A. RQ1 SYSTEMATIC MAPPING

1) TYPES OF STUDIES BY RESEARCH TYPE

Fig. 2 shows the distribution of research types of all 172 studies. The majority of the studies (58%; n = 98) were categorized as validation research, followed by experience papers (27%, n = 47), solution proposals (13%, n = 23), and evaluation research (2%, n = 4). Considering research content, validation research papers take the lead in terms of the number of studies. This shows the empirical maturity of research interest, and implies the need for stronger empirical studies for evaluation of PM techniques in healthcare.



Validation Search: [11], [16], [22], [25], [33]–[126]
Experience Papers: [26], [127]–[172]
Solution Proposal: [1], [24], [28], [173]–[191]
Evaluation Research: [192]–[195]

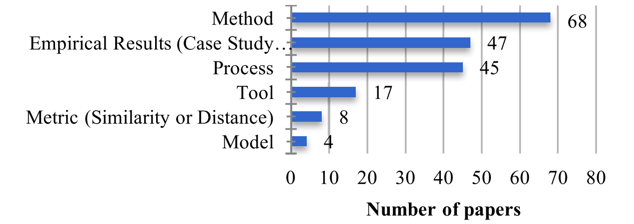
FIGURE 2. **Distribution of studies by research type.**

2) RQ1.2 TYPES OF STUDIES BY CONTRIBUTION TYPE

Fig. 3 shows the distribution of the selected papers by contribution type. Most researchers (68 studies in total) proposed a new method to deal with the complex nature of healthcare event data. This demonstrates the need for new solutions, such as algorithms, techniques, and approaches in this particular research area. The types of method proposals were data quality assessment and data visualization techniques or a combination of PM and other techniques. In addition, we observed that the number of developed tools, defined metrics and models was low while that of defined processes was high.

Relatively low number of tools can be attributed to a variety of features accessible by publicly available tools (e.g., ProM); however, implementing tools, defining metrics and models for specific purposes are among areas that are open to further research.

There were 14 articles that made two different types of contribution: [25], [66], [70], [73], [82], [83], [93], [94], [98], [108], [110], [117], [125], [174]. For example, study [70] contributed both context group learning algorithm as a new method and a context learning framework as a new process. In another study [66], the authors implemented a tool based on a metric they proposed. We also conducted a cross analysis on papers by research type versus contribution type.



Method: [35], [36], [39], [42]–[44], [48]–[50], [52]–[62], [69]–[73], [77]–[82], [85]–[88], [92], [93], [95]–[103], [105], [106], [108], [110], [112]–[115], [119], [124]–[126], [174], [181], [183], [184], [187], [190], [192], [193], [195], [196]
Empirical Results (Case Study Only): [26], [127]–[172]
Process: [1], [16], [25], [33], [34], [38], [40], [45]–[47], [63]–[65], [70], [83], [84], [89], [90], [93], [94], [98], [106]–[109], [116]–[118], [120]–[123], [125], [173]–[176], [179], [180], [182], [186], [188], [189], [194]
Tool: [22], [24], [25], [28], [41], [51], [66], [67], [74], [76], [82], [83], [106], [177], [178], [185], [191]
Metric (Similarity or Distance): [66], [68], [73], [75], [91], [94], [106], [110]
Model: [11], [37], [111], [117]

FIGURE 3. Distribution of studies by contribution type.

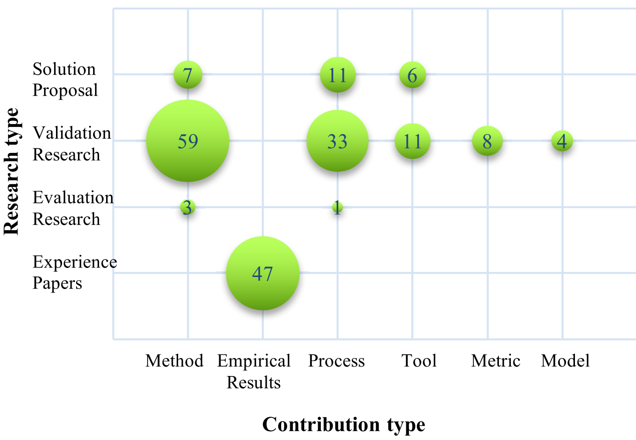


FIGURE 4. Cross analysis of studies by research versus contribution type.

Fig. 4 shows twox-y scatterplots with the number of papers in category intersections. Most papers that contributed a process or method were validated with a case study. For example, for the validation research category, there were 33 papers that validated the proposed process with a case study. In another example, six out of 17 tools were proposed only as a solution and 11 out of 17 tools were implemented, and their applicability was demonstrated using the process logs of hospital data. This analysis denotes that the majority of researchers (in 59 studies) contributed a new and inadequately validated method or reported empirical results based on their experience with existing proposals (in 47 studies).

3) RQ1.3 APPLICATION CONTEXT

To examine the application context that received more attention from researchers, we focused on two healthcare processes to which PM techniques had been applied; namely, healthcare process and clinical pathways. The majority of researchers (93 studies) used the healthcare process rather than clinical pathways (59 studies) as shown in Fig. 5. We also examined the PM application context in terms of different levels used; i.e., departmental and organizational, and found that most studies analyzed a single department or a single hospital rather than multiple departments or hospitals; i.e., most applications had a narrower scope. This can be attributed to the problems of data integration, data availability, data confidentially, or different physical locations. Based on these results, we consider that application of PM techniques to clinical pathways and multiple departments or hospitals is an important challenge that needs to be addressed.

Fig. 6 shows a cross-analysis of healthcare process and clinical pathways with respect to department or hospital level with x-y scatterplots with the number of papers shown in the category intersections. The majority of the application context was healthcare process in a single department (n=76) and a single hospital (n=85). It was followed by applications to clinical pathways in a single department (n=46) and a single hospital (n=42). Applications to healthcare process or clinical pathways in multiple departments or hospitals, on the other hand, were scarce.

4) RQ1.4 HEALTHCARE SPECIALTY

We identified and classified studies according to the healthcare specialty by holding discussions with two medical experts. Fig. 7 provides the distribution of the number of papers among 21 healthcare specialties. This distribution indicates that PM was applied to almost all healthcare specialties with the primary areas being oncology, surgery, emergency department, neurological diseases, and cardiovascular diseases. Among these, oncology was most studied probably because this process has well-defined steps and requires following rigid medical protocols. Nevertheless, it is clear from the figure that PM can be applied for various specialties with a wide range of clinical datasets.

Some of the papers were classified under more than one specialty. For example, in studies [41] and [185], the authors presented the tool they developed using data from pediatric emergency department patients with a primary diagnosis of asthma as a respiratory disease. The study [133] focused on the gynecological oncology department process and investigated the use of a specific treatment/drug. In another study [92], the authors proposed a new method and evaluated it with real-world data collected in relation to four specific diseases; bronchial lung cancer, colon cancer, gastric cancer, and cerebral infarction. Thus, we classified this study under the specialties of both oncology and neurological disease (cerebral infarction). In several publications PM was used to explore the behavior of nurses and doctors and relations between them (classified under "other" category in Fig. 7), which is an example to model role-based processes from organizational perspectives.

5) RQ1.5 TYPE OF PROCESS MINING ACTIVITIES

Taking into consideration the PM implementations reported in the selected papers, we defined other PM activities such as process variant analysis, performance analysis, outlier detection, and predictive monitoring in addition to process discovery, conformance checking, and enhancement.

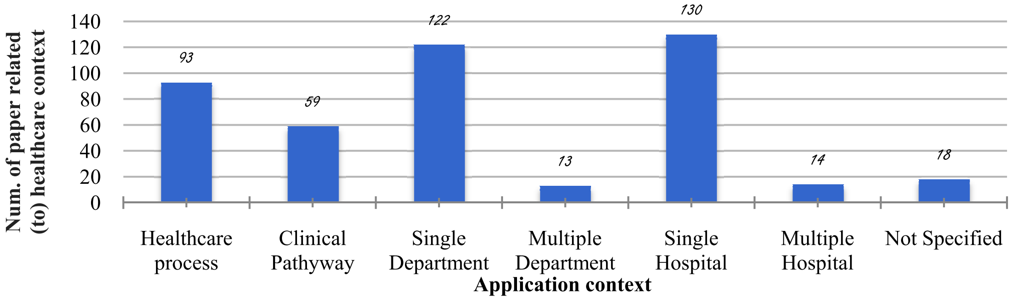


FIGURE 5. Distribution of studies by application context.

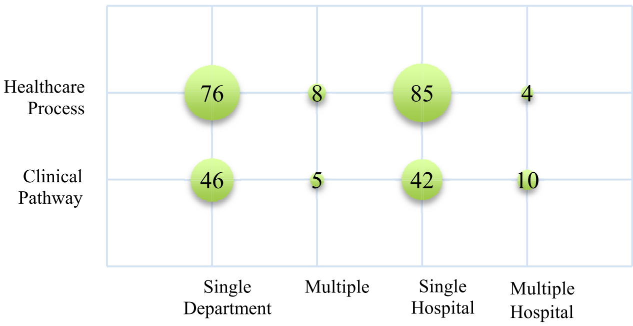


FIGURE 6. Cross analysis of studies by healthcare process/clinical pathways versus department/hospital level.

The distribution of the number of studies for all PM activities as elicited from the papers is shown in Fig. 8. Most researchers (n = 155) applied PM to discover a healthcare process and only few (n = 21) used PM for the enhancement of a healthcare process. The authors mostly tended to analyze process variants or perform performance analyses to define areas of opportunity for process enhancement. Since PM techniques are emerging and process discovery is the basis for other types of activities, there was a higher number of studies that applied PM for discovery purposes.

Fig. 8 presents the types of PM activities used in the selected studies. In this figure, the activities in the lower part need to be more frequently addressed so that PM can be used for decision-making and operational support.

Fig. 9 shows cross-analysis of PM activity types with respect to contribution type (on the left) and research type (on the right) with two x-y scatterplots with the number of papers shown in the category intersections. The majority of papers that contributed a new method (n = 58) focused on process discovery. In 88 papers, researchers validated their contribution to process discovery. In Fig. 9, the smaller circles show the research areas open to exploration; e.g., there is no tool, process or metric for predictive monitoring. The bigger circles indicate the research areas that have been frequently addressed and therefore are more likely to reach saturation.

6) RQ1.6 & RQ 1.7 PROCESS MODELING TYPE AND MODELING NOTATION

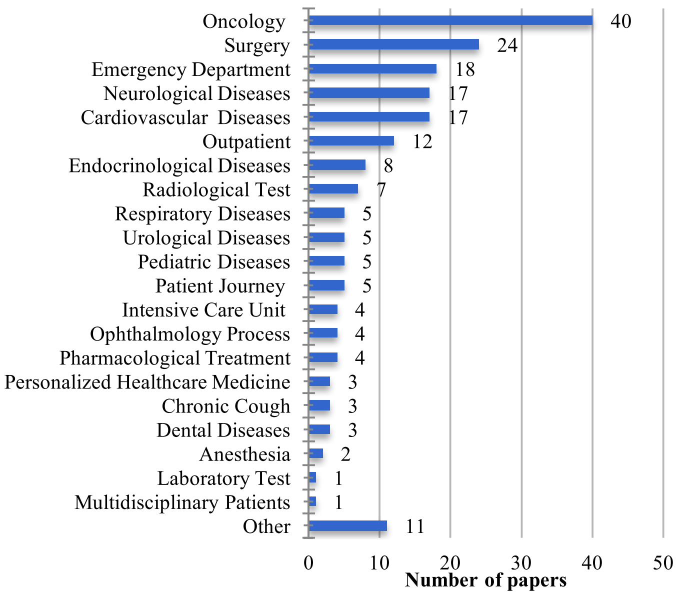
The type of process modeling notation has an influence on the discovery of the process and can change the outcomes. Through this systematic mapping analysis, we investigated modeling notations that are preferred to model healthcare processes, and grouped modeling types as manually discovered, automatically discovered, or created manually after being discovered by PM techniques. Fig. 10 shows an analysis of modeling notations versus modeling types by the number of papers. The figure indicates that there are many notations and languages that can be used according to needs. The mostly used process modeling notation was heuristic net [4] (n = 22), and all process models using this notation were discovered automatically. The other popular notations were petri net [17], fuzzy model [197], and declare model [16]. There were many newly proposed notations to visualize healthcare processes, some of which were reported to perform better than the existing ones. BPMN [18], flowchart [198], and UML diagrams [199] (class and activity diagrams) were mostly used to manually define healthcare processes.

7) RQ1.8 PROCESS MINING TECHNIQUES

a: PROCESS DISCOVERY TECHNIQUES

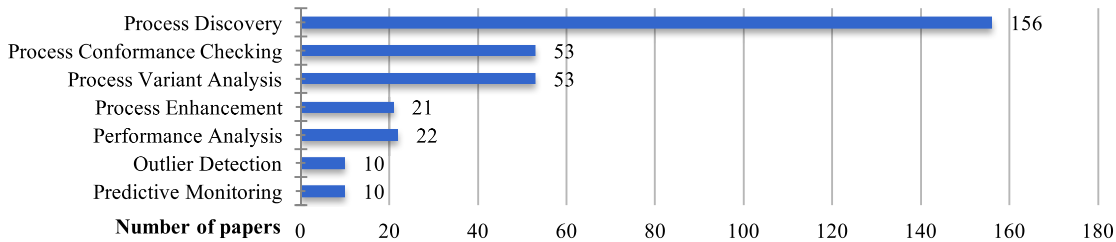
A process is considered as being successfully discovered if it accurately reflects the behaviors of the real process. In discovery, it is necessary to establish a balance between four criteria; fitness, simplicity, precision, and generality [7]. Fitness is the ability of the model to explain an observed behavior. Simplicity means that the discovered process model should be as simple as possible. Precision is avoiding under-fitting, which means preventing any behavior that is not related to the event data and generality is avoiding over-fitting, which means the discovered model should generalize the example behavior seen in the event log.

The PM discovery techniques used in the selected papers are outlined in Table 4. The most commonly used process discovery techniques were Heuristic Miner [200], Fuzzy Miner [197], and Alpha Algorithm [7]. Heuristic miner, as the second frequently used PM algorithm, follows the alpha algorithm and takes into account the frequencies and causal dependencies by considering the dependency measure [200]. It can abstract from exceptional behavior, and therefore is also suitable for many real-life logs. Fuzzy miner is one of the younger process discovery algorithms and uses significance/correlation metrics to interactively simplify the



Oncology: [11], [33], [35], [39], [45], [46], [48]–[50], [55]–[59], [62], [64], [65], [71], [77], [79], [81], [92], [93], [98], [102], [104], [107], [121], [123], [125], [130], [133], [147], [152], [157], [169], [170], [176], [192], [194]
Surgery: [11], [35], [39], [48], [50], [61], [76], [77], [83], [86], [95], [100], [101], [105], [112], [114], [121], [123], [125], [135], [152], [161], [166], [188]
Emergency Department: [22], [25], [41], [94], [103], [106], [117], [118], [131], [132], [137], [140], [142], [144], [150], [165], [178], [185]
Neurological Diseases: [1], [66], [68], [69], [75], [91], [92], [96], [106], [112]–[114], [122], [124], [137], [153], [158]
Cardiovascular Diseases: [59], [60], [72]–[74], [78], [84], [99], [109], [111], [115], [159], [177], [184], [188]
Outpatient: [37], [40], [47], [49], [51], [87]–[89], [108], [126], [136], [140], [142], [144], [148]–[150], [160], [168], [171]
Endocrinological Diseases: [34], [42], [54], [74], [116], [122], [154], [155]
Radiological Test: [25], [57], [84], [129], [139], [151], [164]
Respiratory Diseases: [41], [156], [159], [178], [185]
Urological Diseases: [16], [70], [80], [88], [119]
Pediatric Diseases: [36], [41], [106], [178], [185]
Patient Journey: [37], [131], [140], [142], [168]
Intensive Care Unit: [37], [52], [157], [163]
Ophthalmology: [95], [105], [121], [182]
Pharmacological Treatment: [72], [97], [104], [133]
Personalized Healthcare Medicine: [120], [167], [186]
Chronic Cough: [87], [89], [90]
Dental Diseases: [38], [143], [145]
Anesthesia: [26], [86]
Laboratory Test: [141]
Multidisciplinary Patients: [63]
Other: [36], [44], [52], [53], [82], [85], [103], [127], [128], [134], [180]

FIGURE 7. Distribution of studies by healthcare specialty.



Process Discovery: [1], [11], [16], [22], [24]–[26], [28], [33]–[41], [43]–[49], [51]–[60], [62]–[72], [74]–[80], [82]–[86], [88]–[90], [92]–[105], [108]–[114], [116]–[118], [120]–[136], [138]–[158], [160], [161], [163], [164], [166]–[169], [172]–[191], [193], [195], [196]
Process Conformance Checking: [11], [16], [24], [28], [36], [39]–[41], [44]–[47], [49]–[51], [61], [65], [66], [68], [71], [73], [75], [82], [83], [87], [89], [90], [93], [102], [104], [108], [116], [120], [129], [133], [136], [137], [142], [144], [145], [150], [152], [154], [164], [166], [167], [172], [173], [175], [176], [182], [184], [188]
Process Variant Analysis: [25], [28], [34], [39]–[41], [45], [47], [48], [57], [59], [65], [67], [69], [73], [74], [78], [82]–[84], [91], [94], [100], [104]–[108], [110], [111], [115], [117], [119], [122], [125], [132], [135]–[137], [140], [142], [144], [146], [150], [173]–[177], [185], [189], [190]
Process Enhancement: [16], [33], [35], [38]–[40], [49], [62], [83], [102], [106], [113], [125], [142], [145], [149], [152], [173], [175], [182], [188]
Performance Analysis: [11], [25], [40], [41], [47], [49], [83], [105], [108], [117], [130], [132], [144], [145], [150], [152], [153], [156], [164], [165], [175], [190]
Outlier Detection: [42], [57], [67], [82], [83], [91], [100], [115], [140], [175]
Predictive Monitoring: [81], [97], [106], [119], [159], [160], [190]–[192], [194]

FIGURE 8. Type of process mining activities.

process model at desired level of abstraction. Disco software, for example, is based on a fuzzy miner technique. Using this tool, researchers can create visual maps from a healthcare process in minutes in a very practical manner and replay an event log on the model. Fuzzy model has also seamless simplification, which allows researchers to replace mainstream or exceptional cases with either normal behavior or a behavior that is very rare. Alpha miner, is the first and simplest mining algorithm, creates petri nets from the discovered model [7]. Many techniques utilize these petri net models as input for other types of PM activities. Among these three techniques, heuristic miner and fuzzy miner are more preferred by researchers because they deal with noise and exceptions better and allow users to focus on the main process flow rather than details of behavior.

Inductive miner is another powerful PM technique and has slider functionality by which to remove activities and paths that are infrequent [201]. It is possible to determine how frequently certain arcs are being taken and to animate replaying event logs on the discovered model in order to identify delays, deviations, bottlenecks, or other problems. Interestingly, this mining technique was used for healthcare process in only four of the primary studies.

b: PROCESS CONFORMANCE TECHNIQUES

In conformance checking, a log and a model are used to provide diagnostics and quantify the differences between the modeled and observed behaviors. The aim of this process is to determine the most frequent deviations, why

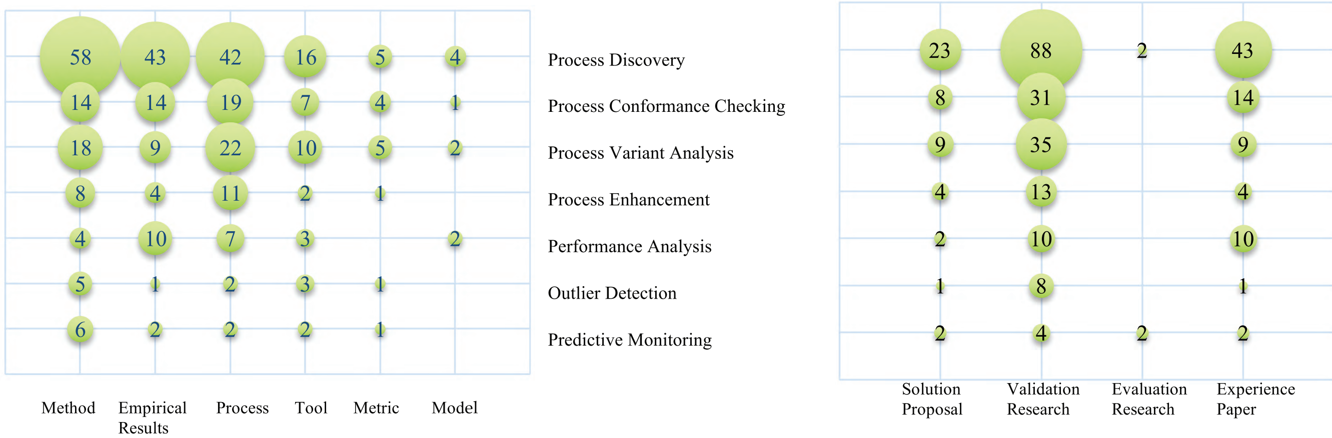
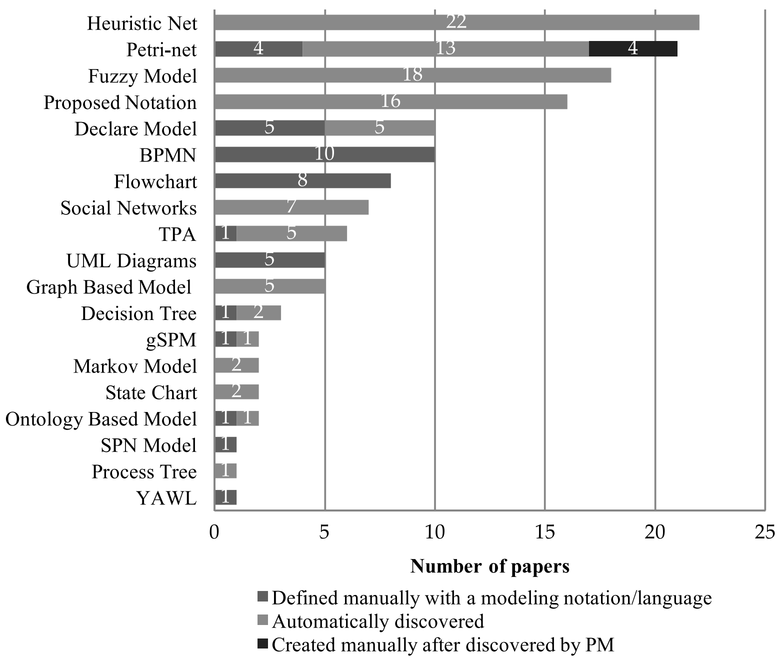


FIGURE 9. Cross analysis of studies in different process mining activity types w.r.t. contribution and research types.



Heuristic Net: [1], [26], [33], [47], [65], [67], [68], [75], [84], [98], [104], [129], [130], [133], [135], [140], [145], [148], [153], [161], [164], [168]

Petri Net: [11], [25], [36], [38], [39], [50], [55], [63], [93], [97], [141], [143], [144], [149], [151], [152], [154], [163], [168], [172], [182]

Fuzzy Model: [34], [52], [64], [88], [94], [96], [117], [118], [120], [127], [131], [134], [138], [146], [156], [162], [167], [176]

Proposed Notation: [41], [57], [60], [64], [74], [76], [78], [80], [101], [124], [125], [174], [177], [178], [183], [185]

Declare Model: [16], [35], [56], [58], [62], [79], [87], [89], [90], [102]

BPMN: [46], [49], [61], [81], [113], [139], [147], [150], [193], [195]

Flowchart: [40], [73], [132], [137], [142], [169], [179], [180]

Social Networks: [25], [103], [128], [130], [133], [145], [170]

TPA: [22], [43], [54], [82], [83], [99]

UML Diagrams: [44], [48], [51], [77], [136]

Graph Based Model: [66], [111], [112], [114], [160]

Decision Tree: [70], [121]

gSPM: [95], [105]

Markov Model: [24], [126]

State Chart: [109], [123]

Ontology Based Model: [184], [187]

SPN Model: [86]

Process Tree: [172]

YAWL: [42]

FIGURE 10. Distribution of number of studies by modeling notation and process modeling type.

they occur, whether they can be predicted, and whether the model or event log is wrong.

The PM process conformance checking techniques used in the selected papers are outlined in Table 5. These papers presented many newly proposed techniques for measuring the differences between process executions and process models. The mostly used PM techniques are LTL Checker, Replay a Log on Petri Net, Conformance Checker, and Trace Alignment. The variety in the techniques used implies that there is no standard technique adopted for conformance checking.

c: PROCESS ENHANCEMENT TECHNIQUES

In process enhancement, a log and an initial model are used to repair, enrich or extend the model with additional information. The PM process enhancement techniques used in the reviewed papers are outlined in Table 6. Most researchers either used simulation techniques for enhancing a process model or proposed a new technique. As additional information, they utilized timestamps (based on extended declare maps containing time information) or performance-related information (performance analysis with petri net) to extend the model [62].

d: OTHER PM TECHNIQUES

Other PM techniques used in the reviewed literature to analyze healthcare processes are outlined in Table 7. The mostly used techniques were dotted chart, performance analysis with petri net, and performance sequence diagrams. Dotted chart is a data visualization technique, in which every event is

TABLE 4. Process discovery techniques used in the reviewed studies.

Process Discovery Techniques	Numbers of Papers	Papers
Newly Proposed Technique	46	[22], [36], [39], [41], [43], [48], [51], [52], [54]–[56], [58]–[60], [69]–[72], [74], [76]–[80], [82], [83], [92], [93], [95], [99]–[101], [105], [110]–[114], [124], [126], [177], [181], [183], [184], [187], [196]
Heuristic Miner	39	[11], [25], [26], [33], [40], [43], [44], [47], [55], [65]–[68], [75], [80], [84], [86], [98], [99], [104], [129], [130], [132], [133], [135], [136], [140], [142], [144], [145], [147], [148], [150], [151], [153], [155], [161], [164], [168]
Fuzzy Miner	28	[11], [34], [40], [52], [55], [57], [64], [86], [88], [94], [96], [113], [117], [118], [124], [127], [131], [132], [134], [136], [138], [144], [146], [149], [150], [156], [162], [167]
Alpha Miner	12	[26], [36], [43], [84], [97], [99], [132], [139], [141], [149], [151], [168]
Social Network Miner	10	[25], [34], [103], [104], [128], [130], [133], [135], [145], [170]
Genetic Process Miner	7	[43], [84], [99], [132], [135], [151], [195]
Alpha++ Miner	6	[36], [43], [89], [90], [99], [151]
ILP Miner	4	[26], [55], [94], [154]
Declare Miner	4	[16], [35], [62], [102]
Inductive Visual Miner	4	[55], [160], [163], [172]
Enhanced Fuzzy Mining	2	[45], [176]
Guide Tree Miner	2	[45], [176]
Multi-phase Miner	2	[68], [151]
Role Hierarchy Miner	2	[34], [134]
Frequency Mining	1	[40], [108]
Decision Mining	1	[46]
DWS-Algorithm	1	[151]
Theory of Regions Based Mining	1	[151]
Handover Mining	1	[47]
Passage Miner	1	[144]
Comp Miner	1	[136]
The Evolutionary Tree Miner	1	[55]
Sequential Pattern Mining	1	[125]
Frequent Pattern Mining	1	[122]
Temporal Pattern Mining	1	[171]
Trajectory Mining	1	[202]
CPN-Tools	1	[46]

TABLE 5. Process conformance checking techniques used in the reviewed studies.

Process Conformance Checking Techniques	Numbers of Papers	Papers
Newly Proposed Technique	16	[36], [39], [41], [46], [49]–[51], [61], [66], [68], [73], [75], [82], [83], [93], [184]
LTL Checker	7	[65], [87], [89], [90], [104], [133], [154]
Replay a Log on Petri Net	5	[44], [129], [144], [164], [172]
Conformance Checker	4	[11], [116], [144], [145]
Trace Alignment	4	[45], [104], [137], [176]
Delta Analysis	3	[40], [47], [136]
Declare Checker	2	[16], [102]
CPN Model Checking	1	[87]
Fuzzy Animation	1	[150]

TABLE 6. Process enhancement techniques used in the reviewed studies.

Process Enhancement Techniques	Numbers of Papers	Papers
Using Simulation	5	[33], [38], [40], [49], [149]
Newly Proposed Technique	3	[39], [49], [83]
Declare Repair	3	[16], [35], [102]
Performance Analysis with Petri net	2	[11], [145]
Not Specified	2	[152], [173]
Extend Declare Map with Correlations	1	[62]
Extend Declare Map with Time Information	1	[62]

represented by a dot to provide a helicopter view of all the events [130]. Its usage can provide many interesting insights of the process before building process model because event data has more information, such as timing and resources, than control flow. Performance analysis with petri net involves replaying event logs on the discovered healthcare petri net process models [202]. Performance sequence diagram is used to generate process variants [21]. In order to overcome performance-related problems, healthcare data can be converted to different levels of abstraction using pattern abstraction or the PH-specific plugin in ProM. In addition, the timestamp issue detector plugin in ProM can be used to detect and analyze duplicate timestamps, differences in accuracy, or outlier [144].

TABLE 7. Other process mining techniques used by the reviewed studies.

Other PM Techniques	Numbers of Papers	Papers
Dotted Chart	8	[11], [34], [38], [40], [44], [47], [130], [152]
Performance Analysis with Petri net	4	[25], [132], [150], [153]
Performance Sequence Diagram	4	[84], [135], [140], [142]
Pattern Abstraction	2	[34], [45]
Newly Proposed Technique	2	[71], [192]
Petri Net Complexity Analysis	1	[129]
Timestamp Issue Detector	1	[144]
PH-specific Plugin	1	[146]

TABLE 8. Clustering techniques used by the reviewed studies.

Clustering Techniques	Numbers of Papers	Papers
Trace Clustering	7	[33], [34], [57], [67], [98], [130], [146]
K-Means Clustering	5	[33], [94], [98], [144], [149]
Hierarchical Clustering	4	[59], [63], [91], [103], [106]
Newly Proposed Technique	5	[42], [57], [91], [94], [103]
Agglomerative Hierarchical Clustering	3	[94], [98], [146]
Sequence Clustering	3	[25], [34], [123]
Frequent Pattern Mining	3	[67], [80], [122]
Self-Organizing Map	2	[98], [130]
Quality Threshold Clustering	1	[98]
The EM Clustering	1	[94]
Activity Clustering	1	[34]
Spectral Clustering	1	[94]
DBScan Clustering	1	[67]
Affinity Propagation	1	[106]
Density Peaks Based Clustering	1	[106]
Model Clustering	1	[194]

8) RQ1.9 CLUSTERING TECHNIQUES

Clustering algorithms for mining healthcare processes have become increasingly popular [98], [130] and used to reduce spaghetti effect on the analysis. Clustering can be considered as a data pre-processing technique among the steps of PM analysis.

Complex and big healthcare data sets are divided into similar sub-sets by filtering outliers or identifying main process. Table 8 presents the various clustering techniques used or adopted for healthcare processes in the reviewed studies. The most commonly used techniques were found to be trace clustering, k-means clustering, and hierarchical clustering. Trace clustering is splitting event logs into homogeneous subsets and creating a process model for each subset. This approach was implemented in a plugin, which contains several clustering algorithms, such as k-means, self-organizing maps, and agglomerative clustering and is used for complex and diverse event data like healthcare process data [98]. K-means clustering, hierarchical clustering and other algorithms are also available in WEKA, a data mining tool adopted by various researchers. Five of the reviewed studies [42], [57], [91], [94], [103] proposed new clustering algorithms based on newly developed similarity or distance metrics.

B. TRENDS AND DEMOGRAPHICS

1) RQ2.1 NUMBER OF STUDIES BY YEAR

The annual cumulative publication volume of the PM studies in the healthcare domain is shown in Fig. 11. Of all the reviewed papers, 156 (90%) focused on process discovery, 53 (30%) on process conformance checking, and 21 (12%) on process enhancement. It should be noted for the figure that a publication might include more than one PM activity type. The increasing number of papers shows that PM techniques are receiving increasing attention in the healthcare domain. Considering the starting year of the publications, PM techniques in the healthcare domain have a relatively short history and have significantly increased in the last few years.

2) RQ2.2 TOP-CITED STUDIES

By this RQ we identified the top-cited papers. Fig. 12 presents the visualization of the total number of citation count of each paper in an x-y plot (on the left) and average annual citation (on the right) with respect to publication year. It should be noted that citation data was extracted from Google Scholar in November 2017 and the last year shown in the x axis was taken as 2016. Furthermore, considering that the papers published in earlier years have a greater probability of having total citations, the average number of citations was used as a more reliable indicator of the citation trend. The results showed that the average number of annual citations per paper has increased in the last few years.

Several papers [25], [55], [98], [130] had a very high number of citations, placing them in the top-cited list within the pool. Among these, Mans *et al.* [130] applied PM techniques to a Dutch hospital data. Rebuge and Ferreira [25] proposed and validated a PM process and a tool that integrated all the process analysis steps of healthcare environments.

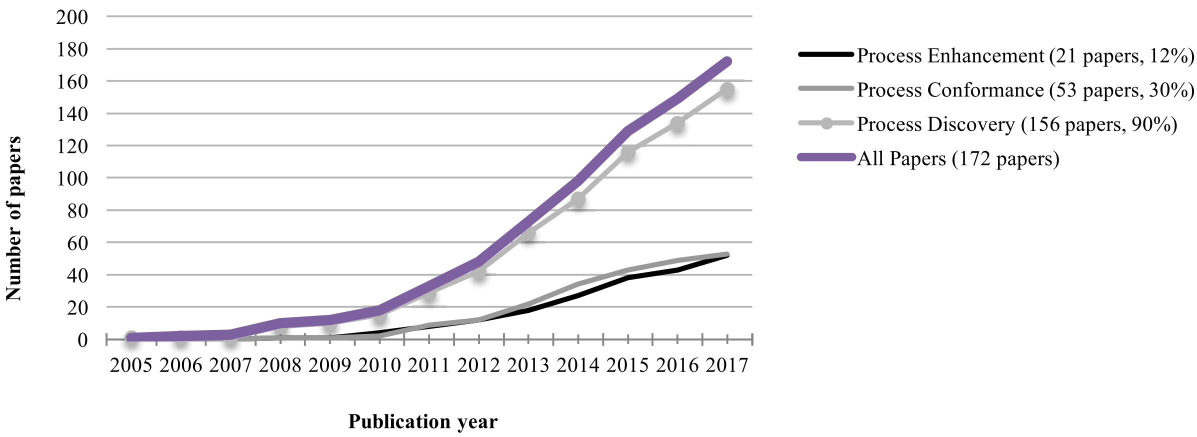


FIGURE 11. Annual trend of studies by basic process mining activity.

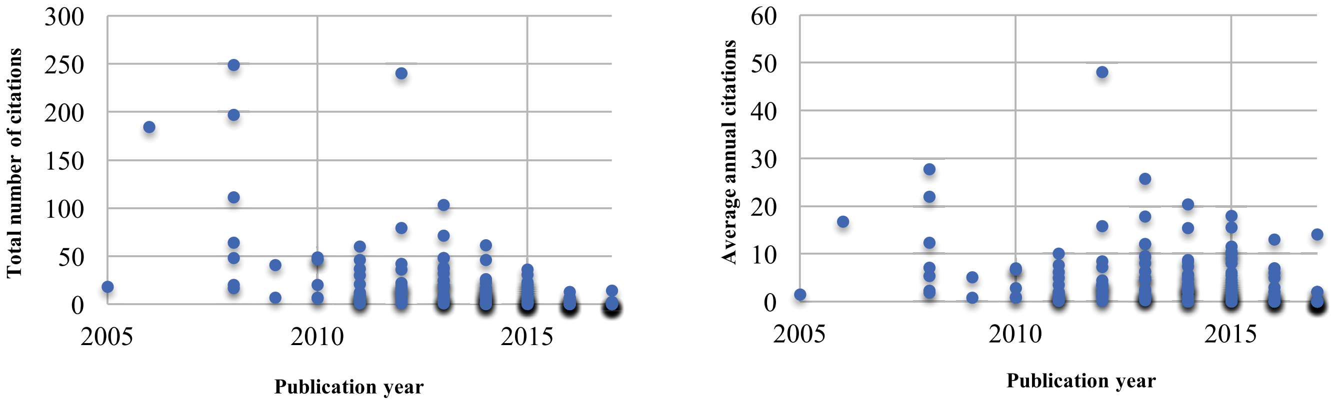


FIGURE 12. Citations by publication year.

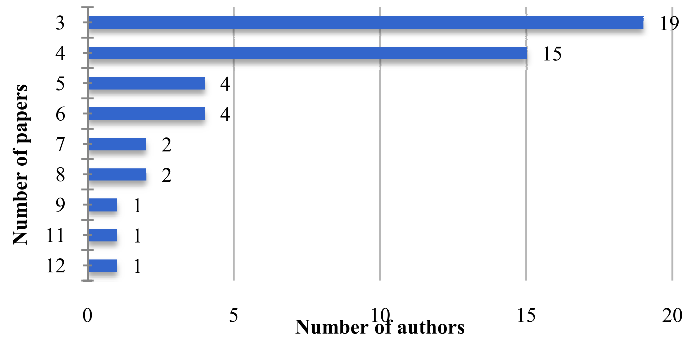
Song *et al.* [98] implemented a trace clustering ProM plugin with a generic methodology that contained several clustering algorithms, such as K-means, self- organizing maps, and agglomerative clustering. The authors also demonstrated the applicability of their proposals with real process logs of hospital data. This methodology and plugin have been used in many following papers. Leemans *et al.* [55] presented a technique to cope with infrequent behavior and large event logs while ensuring soundness and implemented this technique, which is called Inductive Miner – infrequent (IMi), in ProM.

3) RQ2.3 MOST CONTRIBUTING RESEARCHERS

By including up to all authors of each article, we calculated the number of articles published by each author to identify the most contributing researchers. We considered most contributing researchers to be those who had published a minimum of three article in the area (Fig. 13). Thus, 39 authors with three published articles, 68 authors with two published articles, and 262 with only one published article in this area were not included in this researchers list. The most contributing authors were found to be Maggi with 12, Van der Aalst with 11, and Fernandez-LIatas with 9 papers. Maggi, published a paper with Bose and Van der Aalst, in which they proposed declare maps to represent complex healthcare processes and experimented with this modeling language [35]. Van der Aalst and Mans demonstrated that PM can be applied to healthcare processes by presenting the first case study that reported on the qualitative benefits of PM in the healthcare domain [130]. Together with Vanwersch, these authors also published a springer brief [8] that summarized the most important issues of PM applications in this area. The third most contributing author, Fernandez-Llatas *et al.* [43] proposed a new algorithm which is called PALIA to solve the problems related to the existing PM techniques and presented a tool and framework by validating the new algorithm [51]. Among the most contributing researchers, Van der Aalst, Fernandez-LIatas, Traver, and Rojas are the members of Process Mining for Healthcare Group [203].

4) RQ2.4 MOST CONTRIBUTING COUNTRIES

We ranked the numbers of studies according to the countries of origin of the author affiliations. The authors who applied PM techniques in healthcare were from 33 different countries. As shown in Fig. 14, the Netherlands, USA, China, Italy, Spain, and Germany were the most contributing countries.



3: Appelrath, HJ; Basole, RC; Braunstein, M; Burd, RS; Capurro, D; Cavallini, A; Chau, DH; Dagliati, A; Ganesha, K; Marsic,I; Micieli, I; Montani, S; Munoz-Gama, J; Park, H; Sacchi, L; Santos, EP; Sepulveda, M; Vogelgesang, T; Wei, Z; Yang, S
4: Bose, RPJC; Caron, F; Di Francescomarino, C; Hwang, H; Ji, L; Jin, T; Kim, E; Leonardi, G; Lu, X; Quaglini, S; Rojas, E; Song, M; Vanthienen, J, Wang, J; Xu, X
5: Benedi, J; Cho, M; Kim, S;, Premchaiswadi, W
6: Baesens, B; Dng, W; Dumas, M; Yoo, S
7: Mans, R; Traver, V;
8: Duan, H; Huang, Z
9: Fernandez-Llatas, C
11: Van der Aalst, W. M. P.
12: Maggi, F.

FIGURE 13. Most contributing researchers with at least three studies.

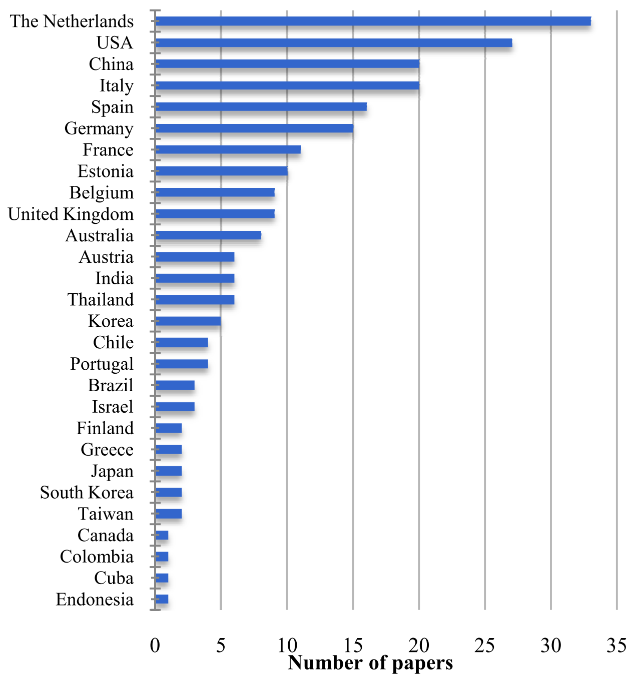


FIGURE 14. Distributions of countries that contributed to PM studies in healthcare.

126 of the papers were written by authors from one country; for 37, the authors were from two countries, and the remaining papers were written by authors from more than two countries.

5) RQ2.5 TOP VENUES

We calculated the number of published papers in each venue to identify the top venues by type. Of the 105 venues on this list, three were a website, four were books, four were symposiums, eight were workshops, 46 were conferences, and 40 were journals. Table 9 presents the venues in which at least two papers were published. The appendix contains the complete list of venues.

VI. SUMMARY OF FINDINGS AND CHALLENGES

A. SUMMARY OF FINDINGS

This systematic mapping study sought an answer to research questions concerning various aspects of PM application in the healthcare domain. The findings related to each RQ are given below:

RQ 1.1 (Research Type): Approximately 85% of the studies were of one of the two research types; validation (58%) and experience (27%), followed by 13% that proposed a new solution. Only 2% of all papers were of evaluation type. This result confirms the empirical maturity of research interest but indicates that empirical techniques are not sufficiently used for evaluation. In other words, research needs to move towards more rigorous validation approaches.

RQ 1.2 (Contribution Type): The majority of the papers contributed a new but poorly validated method (40%), reported empirical results based on experience with methods that had previously been proposed (27%), or proposed a new analysis process (26%). This can be attributed to PM being a relatively young discipline for healthcare processes. The ratio of studies that reported only solution proposals without validation were low (14%), which is a positive observation for the quality of the reported studies. The studies that contributed a new tool, model or metric had a low rate (17%), which denotes the need for more attention.

RQ 1.3 (Application Context): The most frequently analyzed application context was healthcare processes (61%) as an organizational process in a single department of a single hospital. There was a gap in applying PM at larger scales; e.g., in clinical pathways, multiple departments, or multiple hospitals.

RQ 1.4 (Healthcare Specialty): Research on PM implementation was undertaken in 21 healthcare specialties, indicating that PM can be applied for a wide range of clinical datasets. Among these specialties, the most analyzed was oncology process as also reported by other secondary studies [12], [15]. This was followed by surgery process, emergency department processes, and neurological diseases. We encountered several papers in which PM was reported to provide valuable insights from different perspectives such as organizational perspective; e.g., the behavior of doctors/nurses resource perspective [36], [52], [53], [82], [85], [103], [128], [134]; e.g., for personalized healthcare process [120], [167], [186]; e.g., use of prescribed medication [97], [133].

TABLE 9. Top venues with least at two studies (ranked by number of studies).

Venue Type	Venue	Number of papers
Conference	International Conference on Business Process Management	21
Journal	Journal of Biomedical Informatics	8
Conference	International Conference on ICT and Knowledge Engineering	5
Journal	Artificial Intelligence in Medicine	4
Technical Report	bpmcenter.org	4
Workshop	CEUR Workshop Proceedings	4
Conference	Engineering in Medicine and Biology Society (EMBC)	4
Journal	Information Systems	3
Conference	International Conference on Advanced Information Systems Engineering	3
Conference	Asia Pacific Conference	2
Workshop	Australasian Workshop on Health Informatics and Knowledge Management (HIKM)	2
Journal	Computers in Biology and Medicine	2
Journal	Decision Support Systems	2
Journal	Enterprise Information Systems	2
Journal	Expert Systems with Applications	2
Journal	Healthcare Informatics Research	2
Conference	International Conference Automation Science and Engineering (CASE	2
Conference	International Conference on Health Informatics	2
Conference	International Conference on Healthcare Informatics	2
Conference	International Conference on Inventive Communication and Computational Technologies (ICICCT)	2
Conference	International Joint Conference on Biomedical Engineering Systems and Technologies	2
Journal	International Journal of Computer Assisted Radiology and Surgery	2
Journal	International Journal of Medical Informatics	2
Journal	Journal of Healthcare Engineering	2
Journal	Sensors	2
Journal	Studies in Health Technology and Informatics	2
Journal	Transactions on Services Computing	2

RQ 1.5 (Type of Process Mining Activity): The cross-analysis of studies in different PM activity types with respect to contribution and research types showed that process discovery (90%) was the most frequently addressed PM activity type and that the studies including process conformance (30%), process variant analysis (30%), or process enhancement (12%) were less frequent by decreasing degrees of maturity in contribution and research types. This is understandable since process discovery is the basic activity type for other type of activities, yet the observation indicates that the field needs further studies on conformance analysis or process enhancement to improve healthcare processes.

RQ 1.6 & RQ 1.7 (Process Modeling Type & Modeling Notation): PM offers automatic modeling of real processes. Modeling notation may differ according to the healthcare specialty. Among the reviewed papers, the most frequently used notations were heuristic net (13%), petri net (12%), and fuzzy model (10%). In terms of manual methods for healthcare process modeling, BPMN, Flowchart, and UML diagrams were addressed by 13% of the reviewed studies. Process models can also be refined manually to obtain better representations after an automatic discovery process (e.g., [11], [144], [152]). To represent a healthcare process in an interactive and user-friendly manner, researchers (in 9% of the studies) also introduced new notations by proposing tools or techniques.

RQ 1.8 (Type of PM Techniques): Below are the mostly used techniques for basic PM activities undertaken for healthcare processes:

- Heuristic miner, fuzzy miner, and alpha miner for discovery;
- LTL checker, trace alignment, and conformance checker for conformance checking;
- Simulation, declare repair, and performance analysis plugin with petri net for process enhancement; and
- Dotted chart, performance analysis with petri net, and performance sequence diagram for performance analysis.

In addition to these, many new techniques have also been proposed for process discovery and conformance checking.

RQ 1.9 (clustering techniques): The most common of these techniques were found to be the trace clustering plugin, k-means algorithm, and hierarchical clustering algorithm.

RQ 2 (Trends and Demographics): The annual cumulative publication volume of PM studies showed that PM applications for healthcare processes have attracted increasing attention from researchers, particularly in the last few years. Several papers, e.g., [130], [25], [55] and [98], had top citations with their methodologies having been followed by many later studies. Maggi, Van der Aalst, and Fernandez-LIatas were the most contributing researchers in the field and the most contributing countries were the Netherlands, USA, and China. The top visited conference was the International Conference on Business Process Management (BPM) and the top preferred journal was the Journal of Biomedical Informatics.

B. CHALLENGES OBSERVED

Based on the observations gathered from the studies in our pool, we identified a number of challenges related to healthcare data and PM techniques. We summarize these below.

1) HEALTHCARE DATA CHALLENGES

Of the 172 studies included in our pool, 47 studies were concerned with exploring ways of evaluating healthcare data by generating event logs, repairing event logs, eliminating noisy data, data extraction, data modeling, data integration, and data preprocessing.

Healthcare data are gathered at different levels of granularity from different data sources including hospital information systems, clinical data warehouses, real time data sources (RFID tags, RTLS), medical devices (X-ray machines), video reviews, and datasets created for financial purposes (e.g., billing data). Identifying and merging healthcare data with different characteristics from these sources are challenging tasks. In addition, event data are often incomplete and object-centric rather than process-centric, and contain outliers and events at different levels of granularity [5]. For the effective use of PM techniques in the healthcare domain, problems related to data modeling, extraction, integration, preprocessing, abstraction, visualization, and inspection need to be addressed and resolved.

Dealing with big and complex healthcare data also brings performance efficiency challenges. These can be solved with abstraction levels of time and event data, splitting event logs using clustering algorithms and sophisticated PM techniques and tools. In some studies, researchers recommended using alternative plugins [34], [144], [146] to solve performance problems related to healthcare data or proposed new PM techniques or tools [24], [55], [59], [177], [185] for the summarization of healthcare data to allow focusing on the main features of the diverse and big data.

In addition to the above issues, representation of complex and variable healthcare processes is expected to improve usability and understandability not only for medical experts but also for non-experts. The challenge is to hide the complex PM techniques behind user-friendly and interactive interfaces and notations, which can automatically set parameters and filters, and suggest suitable types of analysis [5]. During our mapping we encountered studies that proposed data visualization techniques including a performance summary [64], [80], allowed the interactive specification of a patient group through filtering [74], [177], [178], [185], and enabled comparisons between different patient populations or process models [41], [174]. Although there were four studies [41], [74], [177], [178] that developed a visual analytics tool, three involving a new modeling notation, they only reported preliminary results from the case studies they performed.

In future studies, researchers may consider using public repositories, such as gynecology data [204] and sepsis data [205], which have already been processed and prepared for PM.

2) TECHNIQUE-RELATED CHALLENGES

There are many PM techniques available and various software vendors offer different software products for general purposes. This mapping study clearly revealed that there were many newly proposed techniques ($n = 68$) and tools ($n = 17$) specifically in the healthcare domain. This diversity is good on one side but it may create difficulties especially for potential users on the other side. There is a place for independent evidence or a comparison on the quality of these new techniques or tools. There is also a place for studies that report process discovery techniques by measuring the quality of discovery metrics (fitness, simplicity, precision, and generalization). Creating representative benchmarks for evaluating the PM techniques and tools for strengths and weaknesses remains as a challenge and will be useful, especially to guide practitioners in the pool of the various assets.

VII. CONCLUSION

PM is an emerging set of techniques applied for business process management in the healthcare domain and this study provided a descriptive analysis of the related literature by applying systematic mapping on PM studies targeted for healthcare. In the mapping study, we included 172 relevant papers and analyzed them with respect to various aspects including research and contribution type, application context and healthcare specialty, process modeling type and notation, PM techniques, and demographic and bibliometric analysis.

The mapping of included studies showed that the field is rapidly growing, and open for further research and practice. A large number of studies that reported validations of their proposals or experiences based on previous proposals indicates increasing empirical maturity of the field. Designing and conducting further studies with even stronger validation approaches will contribute to maturity of research and practice.

Many of the studies proposed methods and processes, and few introduced tool, metric and models. There was also a variety in the techniques proposed for different purposes of use such as process discovery, conformance analysis, process enhancement, and predictive monitoring. These findings highlight dynamic and emerging nature of research in the field.

There is an opportunity to propose or use PM techniques in healthcare contexts such as clinical pathways, multiple departments and multiple hospitals, in which study frequencies were observed fewer in comparison to healthcare processes in a single department or single hospital. Also, there is a need for further studies on conformance analysis or process enhancement in various contexts to better demonstrate the usefulness of PM techniques for healthcare business process management.

In addition to the body of evidence on the applicability and usefulness of PM techniques in the healthcare domain, there are also significant challenges to be addressed. The basic

TABLE 10. Complete list of venues of the primary studies included in this systematic mapping.

Venue Type	Venue	Number of papers
Journal	Artificial Intelligence in Medicine	4
Journal	Computers in Biology and Medicine	2
Journal	Data Mining and Knowledge Discovery	1
Journal	Decision Support Systems	2
Journal	Enterprise Information Systems	2
Journal	Expert Systems with Applications	2
Journal	Future Generation Computer Systems	1
Journal	Health Environments Research &Design Journal	1
Journal	Health Informatics Journal	1
Journal	Health Information Management Journal	1
Journal	Healthcare Informatics Research	2
Journal	IEEE Software	1
Journal	Information Systems	3
Journal	International Journal of Computer Assisted Radiology and Surgery	2
Journal	International Journal of Electronics and Telecommunications	1
Journal	International Journal of Medical Informatics	2
Journal	International Journal of Parallel Programming	1
Journal	Journal of Biomedical and Health Informatics	1
Journal	Journal of Biomedical Informatics	8
Journal	Journal of Computers	1
Journal	Journal of Digital Imaging	1
Journal	Journal of Healthcare Engineering	2
Journal	Journal of Medical Systems	1
Journal	Journal of Software Maintenance and Evolution: Research and Practice	1
Journal	Journal of Systems Science and Complexity	1
Journal	Journal of the American College of Surgeons	1
Journal	Journal of the American Medical Informatics Association	1
Journal	Journal of Theoretical and Applied Information Technology	1
Journal	Knowledge-Based Systems	1
Journal	Methods of Information in Medicine	1
Journal	Procedia Technology	1
Journal	Sensors	2
Journal	Software & Systems Modeling	1
Journal	SSRN	1
Journal	Studies in Health Technology and Informatics	2
Journal	Transactions on Information Technology in Biomedicine	1
Journal	Transactions on Intelligent Systems and Technology (TIST)	1
Journal	Transactions on Management Information Systems (TMIS)	1
Journal	Transactions on Services Computing	2
Book	Advances in Transdisciplinary Engineering	1
Book	Data Mining in Clinical Medicine	1
Book	Encyclopedia of Business Analytics and Optimization	1
Book	Transactions on Petri Nets and Other Models of Concurrency XI	1
Conference	Americas Conference on Information Systems (AMCIS)	1
Conference	Asia Pacific Conference	2
Conference	Benelux Conference on Artificial Intelligence (BNAIC)	1
Conference	Biomedical and Health Informatics (BHI)	1
Conference	Engineering in Medicine and Biology Society (EMBC)	4
Conference	HIMSS conference proceedings	1
Conference	Human Factors in Computing Systems	1
Conference	International Conference Automation Science and Engineering (CASE	2
Conference	International Conference of Modeling and Simulation (MOSIM)	1
Conference	International Conference on Advanced Information Systems Engineering	3
Conference	International Conference on Advances in Information Communication Technology & Computing	1
Conference	International Conference on Business Informatics Research	1
Conference	International Conference on Business Process Management	21
Conference	International Conference on Case-Based Reasoning	1
Conference	International Conference on Conference on Information and Knowledge Management	1
Conference	International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)	1
Conference	International Conference on Data Mining Workshops (ICDMW)	1
Conference	International Conference on e-Health Networking, Applications and Services (Healthcom)	1
Conference	International Conference on Grid and Pervasive Computing	1
Conference	International Conference on Health Informatics	2
Conference	International Conference on Healthcare Informatics	2
Conference	International Conference on ICT and Knowledge Engineering	5

TABLE 10. *(Continued.)* Complete list of venues of the primary studies included in this systematic mapping.

Conference	International Conference on Industrial conference on data mining (ICDM)	1
Conference	International Conference on Industrial Engineering and Engineering Management (IEEM)	1
Conference	International Conference on Information & Communication Technology and Systems (ICTS)	1
Conference	International Conference on Information Society and Techology (ICIST)	1
Conference	International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI)	1
Conference	International Conference on Inventive Communication and Computational Technologies (ICICCT)	2
Conference	International Conference on Knowledge Discovery and Data Mining (SIGKDD)	1
Conference	International Conference on Networking, Sensing and Control (ICNSC)	1
Conference	International Conference on Services Computing Context-Aware	1
Conference	International Congress of the European Federation for Medical Informatics	1
Conference	International Joint Conference on Biomedical Engineering Systems and Technologies	2
Conference	International Joint Conference on Computer Science and Software Engineering (JCSSE)	1
Conference	On the Move to Meaningful Internet Systems: OTM Conferences	1
Conference	Pacific-Asia Conference on Knowledge Discovery and Data Mining	1
Conference	Pervasive Computing Technologies for Healthcare (PervasiveHealth)	1
Conference	Proc Healthcare Systems Process Improvement Conference	1
Conference	SIGKDD conference on knowledge discovery and data mining workshop on data science for social good	1
Conference	Software Measurement and the International Conference on Software Process and Product Measurement	1
Conference	Systems, Man, and Cybernetics (SMC)	1
Conference	Trust, Security and Privacy in Computing and Communications (TrustCom)	1
Conference	Winter Simulation Conference	1
Conference	Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society (USAB)	1
Conference	Working Conference on Virtual Enterprises	1
Conference	World Congress on Health and Biomedical Informatics	2
Workshop	Australasian Workshop on Health Informatics and Knowledge Management (HIKM)	2
Workshop	CEUR Workshop Proceedings	4
Workshop	EDBT/ICDT 2013 Workshops	1
Workshop	International Workshop on Algorithms & Theories for the Analysis of Event Data	1
Workshop	Learning from Medical Data Streams Workshop (LEMEDS)	1
Workshop	Workshop on Knowledge Representation for Health-Care Data, Processes and Guidelines	1
Workshop	Workshop on Medical Cyber-Physical Systems	1
Workshop	Workshop on Visual Analytics in Healthcare	1
Symposium	International Symposium on Applied Sciences in Biomedical and Communication Technologies	1
Symposium	International Symposium on Computer-Based Medical Systems	1
Symposium	International Symposium on Data-Driven Process Discovery and Analysis	1
Symposium	SIGHIT International Health Informatics Symposium	1
Website	bpmcenter.org	4
Website	BMC Health Services Research	1
Website	Hasso Plattner Institute	1

challenges observed during mapping are related to healthcare data, PM techniques, performance, and data visualization. We should note that many of our observations and challenges identified are in parallel with the findings reported in previous secondary research [5], [12], [14], [151], [206]. We expect that this study will be useful, as the previous ones, in making the community more aware of the need and importance for further PM studies in healthcare.

As the future work, we plan to deepen the descriptive analysis that we performed in this mapping study by conducting systematic literature reviews (SLRs), as also suggested by [9]. As an example, we are currently working on an SLR that will highlight the most beneficial and problematic PM techniques as experienced in the healthcare settings. We expect the results from such review will better reflect the advantages and the difficulties of using the PM techniques, and provide deeper insights for directing future efforts. In addition, based on the findings and challenges we have reported, PM researchers and practitioners might target future efforts on, e.g.: 1) defining comprehensive methodologies for in-depth analysis of healthcare processes for management purposes, 2) extracting and evaluating quality of healthcare process data based on the goals of PM projects, 3) developing domain-specific tools to support on-site analysis of healthcare process data by medical professionals and hospital managers, 4) developing summary dashboards which show performance indicators as measured by applying the PM techniques on healthcare processes, and 5) conducting evaluation studies of PM projects within multiple departments of multiple hospitals via healthcare process comparison or benchmarking.

Appendix

See Table 10.

ACKNOWLEDGEMENTS

The authors thank Prof. Dr. N. Çağla Tarhan, a full professor of radiology, and Şebnem Yücel, M.D., an assistant professor of medical oncology, for their suggestions on the grouping of healthcare specialties in this research.

REFERENCES

[1] S. Quaglini, "Information and communication technology for process management in healthcare: A contribution to change the culture of blame," *J. Softw., Evol. Process*, vol. 22, nos. 6–7, pp. 435–448, 2010.

[2] M. Kirchmer, S. Laengle, and V. Masias, "Transparency-driven business process management in healthcare settings," *IEEE Technol. Soc. Mag.*, vol. 32, no. 4, pp. 14–16, Dec. 2013.

[3] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*. Heidelberg, Germany: Springer, 2013.

[4] W. M. P. van der Aalst, *Process Mining: Data Science in Action*. Springer, 2016.

[5] W. M. P. van der Aalst, "Process mining manifesto," in *Proc. Bus. Process Manage. Workshops*, 2011, pp. 169–194.

[6] W. M. P. van der Aalst and A. J. M. M. Weijters, "Process mining: A research agenda," *Comput. Ind.*, vol. 53, no. 3, pp. 231–244, 2004.

[7] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

[8] R. S. Mans, W. M. P. van der Aalst, and R. J. B. Vanwersch, *Process Mining in Healthcare Evaluating and Exploiting Operational Healthcare Processes*. Heidelberg, Germany: Springer, 2015.

[9] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, Aug. 2015.

[10] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering version 2.3," *Engineering*, vol. 45, no. 4ve, p. 1051, 2007.

[11] R. Mans, W. M. P. van der Aalst, and R. Vanwersch, "Process mining in healthcare: Opportunities beyond the ordinary," Tech. Rep., 2013.

[12] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *J. Biomed. Informat.*, vol. 61, pp. 224–236, Jun. 2016.

[13] M. Ghasemi and D. Amyot, "Process mining in healthcare: A systematised literature review," *Int. J. Electron. Healthc.*, vol. 9, no. 1, pp. 60–88, 2016.

[14] W. Yang and Q. Su, "Process mining for clinical pathway: Literature review and future directions," in *Proc. ICSSSM*, 2014, pp. 1–5.

[15] A. Kurniati, O. Johnson, D. Hogg, and G. Hall, "Process mining in oncology: A literature review," in *Proc. ICICM*, Oct. 2016, pp. 291–297.

[16] M. Rovani, F. Maggi, M. De Leoni, and W. M. P. van der Aalst, "Declarative process mining in healthcare," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9236–9251, 2015.

[17] J. Peterson, *Petri Net Theory and the Modeling of Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1981.

[18] R. M. Dijkman, M. Dumas, and C. Ouyang, "Semantics and analysis of business process models in BPMN," *Inf. Softw. Technol.*, vol. 50, no. 12, pp. 1281–1294, Nov. 2008.

[19] I. D. Scheer. (2002). *ARIS Process Performance Manager (ARIS PPM)*. [Online]. Available: http://www.ids-scheer.com

[20] *Process Mining and Automated Process Discovery Software for Professionals—Fluxicon Disco*. Accessed: Feb. 26, 2018. [Online]. Available: https://fluxicon.com/disco/

[21] *ProM—The Leading Process Mining Toolkit*, Process Mining Group, BPM (Demos) , 2014.

[22] E. Rojas *et al.*, "PALIA-ER: Bringing question-driven process mining closer to the emergency room," in *Proc. CEUR Workshop*, vol. 1920. 2017, pp. 1–5.

[23] *About Celonis|The Leader & Innovator in Process Mining*. Accessed: Feb. 26, 2018. [Online]. Available: https://www.celonis.com/company/

[24] R. Gatta *et al.*, "pMineR: An innovative R library for performing process mining in medicine," in *Artificial Intelligence in Medicine*, vol. 10259. Cham, Switzerland: Springer, 2017, pp. 351–355.

[25] Á. Rebuge and D. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Inf. Syst.*, vol. 37, no. 2, pp. 99–116, 2012.

[26] U. Kaymak, R. Mans, T. van de Steeg, and M. Dierks, "On process mining in health care," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 1859–1864.

[27] E. Rojas, M. Arias, and M. Sepúlveda, "Clinical processes and its data, what can we do with them?" in *Proc. 8th Int. Joint Conf. Biomed. Eng. Syst. Technol. (BIOSTEC)*, 2015, pp. 642–647.

[28] T. Erdogan and A. Tarhan, "Process mining for healthcare process analytics," in *Proc. 26th Int. Workshop Softw. Meas. (IWSM) 11th Int. Conf. Softw. Process Product Meas.*, 2017, pp. 125–130.

[29] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proc. 18th Int. Conf. Eval. Assess. Softw. Eng. (EASE)*, 2014, pp. 1–10.

[30] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proc. 12th Int. Conf. Eval. Assess. Softw. Eng. (EASE)*, 2008, pp. 68–77.

[31] B. A. Kitchenham, D. Budgen, and O. P. Brereton, "Using mapping studies as the basis for further research—A participant-observer case study," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 638–651, 2011.

[32] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, "Experimentation in software engineering," in *Experimentation in Software Engineering*. Springer, 2012, pp. 1–236.

[33] T. Goila and A. Agrahari,"A framework for evidence based business process redesign," SSRN, Tech. Rep., 2014.

[34] J. Lismont, A. Janssens, I. Odnoletkova, S. van den Broucke, F. Caron, and J. Vanthienen, "A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways," *Comput. Biol. Med.*, vol. 77, pp. 125–134, Oct. 2016.

[35] F. Maggi, R. Bose, and W. M. P. van der Aalst, "A knowledge-based integrated approach for discovering and repairing declare maps," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.*, vol. 7908. 2013, pp. 433–448.

[36] M. Kuo and Y. Chen, "A method to identify the difference between two process models," *J. Comput.*, vol. 7, no. 4, pp. 998–1005, 2012.

[37] C. McGregor, C. Catley, and A. James, "A process mining driven framework for clinical guideline improvement in critical care," in *Proc. Learn. Med. Data Streams Workshop*, vol. 765. 2011, pp. 35–46.

[38] R. Mans, H. Reijers, D. Wismeijer, and M. Van Genuchten, "A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare," *Inf. Syst.*, vol. 38, no. 8, pp. 1097–1115, 2013.

[39] Y. Sun, Y. Du, and M. Li, "A repair of workflow models based on mirroring matrices," *Int. J. Parallel Program.*, vol. 45, no. 4, pp. 1001–1020, 2016.

[40] M. Cho, M. Song, and S. Yoo, "A systematic methodology for outpatient process analysis based on process mining," in *Proc. Asia–Pacific Conf. Bus. Process Manage.*, vol. 181. 2014, pp. 31–42.

[41] R. Basole, H. Park, M. Gupta, M. Braunstein, D. Chau, and M. Thompson, "A visual analytics approach to understanding care process variation and conformance," in *Proc. Workshop Vis. Anal. Healthcare.*, 2015, pp. 6-1–6-8.

[42] B. Han, L. Jiang, and H. Cai, "Abnormal process instances identification method in healthcare environment," in *Proc. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, 2011, pp. 1387–1392.

[43] C. Fernández-Llatas, T. Meneu, J. Benedí, and V. Traver, "Activity-based process mining for clinical pathways computer aided design," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2010, pp. 6178–6181.

[44] A. García, D. Pérez-Alfonso, and O. U. L. Armenteros, "Analysis of hospital processes with process mining techniques," in *Proc. World Congr. Heal. Biomed. Informat.*, vol. 216. 2015, pp. 310–314.

[45] R. Bose and W. M. P. van der Aalst, "Analysis of patient treatment procedures: The BPI challenge case study," 2011.

[46] R. Dunkl, K. Fröschl, W. Grossmann, and S. Rinderle-Ma, "Assessing medical treatment compliance based on formal process modeling," in *Proc. Usability Eng. Austrian Comput. Soc. (USAB)*, 2011, pp. 533–546.

[47] S. Yoo *et al.*, "Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital," *Int. J. Med. Informat.*, vol. 88, pp. 34–43, Apr. 2016.

[48] J. Poelmans, G. Dedene, G. Verheyden, H. Van Der Mussele, S. Viaene, and E. Peters, "Combining business process and data discovery techniques for analyzing and improving integrated care pathways," in *Proc. Adv. Data Mining Appl. Theor. Aspects 10th Ind. Conf. Data Mining*, vol. 6171. 2010, pp. 505–517.

[49] A. Senderovich, M. Weidlich, L. Yedidsion, A. Gal, A. Mandelbaum, and S. Kadish, "Conformance checking and performance improvement in scheduled processes: A queueing-network perspective," in *Proc. CEUR Workshop*, vol. 1701. pp. 72–74, 2016.

[50] X. Lu, R. Mans, D. Fahland, and W. M. P. van der Aalst, "Conformance checking in healthcare based on partially ordered event data," in *Proc. Int. Conf. Bus. Process Manage.*, vol. 202. 2015, pp. 75–88.

[51] C. Fernández-Llatas, T. Meneu, J. Benedí, and V. Traver, "Continuous clinical pathways evaluation by using automatic learning algorithms," in *Proc. Int. Conf. Health Informat. HEALTHINF*, 2011, pp. 228–234.

[52] H. Wolf, K. Herrmann, and K. Rothermel, "Dealing with uncertainty: Robust workflow navigation in the healthcare domain," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–23, 2013.

[53] Y. Zhang, O. Martikainen, P. Pulli, and V. Naumov, "Developing a real-time process data acquisition system for automatic process measurement," in *Proc. Int. Conf. Grid Pervasive Comput.*, 2011, pp. 115–124.

[54] C. Fernández-Llatas, A. Martinez-Millana, A. Martinez-Romero, J. M. Benedí, and V. Traver, "Diabetes care related process modelling using process mining techniques. lessons learned in the application of interactive pattern recognition: Coping with the spaghetti effect," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Nov. 2015, pp. 2127–2130.

[55] S. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," in *Proc. Int. Conf. Bus. Process Manage.*, vol. 171. 2013, pp. 66–78.

[56] S. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering data-aware declarative process models from event logs," in *Business Process Management*, vol. 8094. Berlin, Germany: Springer, 2013, pp. 81–96.

[57] B. Hompes, J. Buijs, W. M. P. van der Aalst, P. Dixit, and J. Buurman, "Discovering deviating cases and process variants using trace clustering," in *Proc. 27th Benelux Conf. Artif. Intell. (BNAIC)*, 2015, pp. 5–6.

[58] F. Maggi, "Discovering metric temporal business constraints from event logs," in *Proc. Int. Conf. Bus. Informat. Res.*, vol. 194. Sep. 2014, pp. 261–275.

[59] Z. Huang, W. Dong, L. Ji, C. Gan, X. Lu, and H. Duan, "Discovery of clinical pathway patterns from event logs using probabilistic topic models," *J. Biomed. Informat.*, vol. 47, pp. 39–57, Feb. 2014.

[60] M. Prodel, V. Augusto, and X. Xie, "Discovery of patient pathways from a national hospital database using process mining and integer linear programming," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2015, pp. 1409–1414.

[61] K. Kirchner, N. Herzberg, A. Rogge-Solti, and M. Weske, "Embedding conformance checking in a process intelligence system in hospital environments," in *Proc. Process Support Knowl. Represent. Health Care*, vol. 16. 2013, pp. 35–42.

[62] R. Bose, F. Maggi, and W. M. P. van der Aalst, "Enhancing declare maps based on event correlations," in *Business Process Management*, vol. 8094. Berlin, Germany: Springer, 2013, pp. 97–112.

[63] L. Maruster and R. J. Jorna, "From data to knowledge: A method for modeling hospital logistic processes," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 2, pp. 248–255, Jun. 2005.

[64] J. De Weerdt, F. Caron, J. Vanthienen, and B. Baesens, "Getting a grasp on clinical pathway data: An approach based on process mining," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2012, pp. 22–35.

[65] F. Caron, J. Vanthienen, and B. Baesens, "Healthcare analytics: Examining the diagnosis-treatment cycle," in *Procedia Technol.*, vol. 9, pp. 996–1004, Oct. 2016.

[66] S. Montani, G. Leonardi, S. Quaglini, A. Cavallini, and G. Micieli, "Improving structural medical process comparison by exploiting domain knowledge and mined information," *Artif. Intell. Med.*, vol. 62, no. 1, pp. 33–45, 2014.

[67] G. Lakshmanan, S. Rozsnyai, and F. Wang, "Investigating clinical care pathways correlated with outcomes," in *Business Process Management*, vol. 8094. Berlin, Germany: Springer, Aug. 2013, pp. 323–338.

[68] S. Montani, G. Leonardi, S. Quaglini, A. Cavallini, and G. Micieli, "Knowledge-intensive medical process similarity," in *Proc. Workshop Knowl. Represent. Health Care Data, Process. Guid.*, 2014, pp. 1–13.

[69] Z. Huang, X. Lu, and H. Duan, "Latent treatment pattern discovery for clinical processes," *J. Med. Syst.*, vol. 37, no. 2, p. 9915, 2013.

[70] J. Ghattas, M. Peleg, P. Soffer, and Y. Denekamp, "Learning the context of a clinical process," in *Proc. Int. Conf. Bus. Process Manage.*, vol. 43. 2009, pp. 545–556.

[71] C. Ciccio, F. Maggi, M. Mecella, and J. Mendling, "Log-based understanding of business processes through temporal logic query checking log-based understanding of business processes," in *Proc. Move Meaningful Internet Syst., OTM Conf.*, vol. 8841. Oct. 2014, pp. 75–92.

[72] R. Williams, B. Brown, N. Peek, and I. Buchan, "Making medication data meaningful: Illustrated with hypertension," *Stud. Health Technol. Informat.*, vol. 228, pp. 247–251, Jan. 2016.

[73] J. Van de Klundert, P. Gorissen, and S. Zeemering, "Measuring clinical pathway adherence," *J. Biomed. Informat.*, vol. 43, no. 6, pp. 861–872, 2010.

[74] A. Perer, F. Wang, and J. Hu, "Mining and exploring care pathways from electronic medical records with visual analytics," *J. Biomed. Informat.*, vol. 56, pp. 369–378, Aug. 2015.

[75] S. Montani, G. Leonardi, S. Quaglini, A. Cavallini, and G. Micieli, "Mining and retrieving medical processes to assess the quality of care," in *Proc. Int. Conf. Case-Based Reason.*, vol. 7969. 2013, pp. 233–240.

[76] T. Vogelgesang and H. Appelrath, "Multidimensional process mining with PMCube explorer," in *Proc. BPM (Demos)*, vol. 1418. 2015, pp. 90–94.

[77] Z. Huang, X. Lu, and H. Duan, "On mining clinical pathway patterns from medical behaviors," *Artif. Intell. Med.*, vol. 56, no. 1, pp. 35–50, 2012.

[78] Z. Huang, W. Dong, P. Bath, L. Ji, and H. Duan, "On mining latent treatment patterns from electronic medical records," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 914–949, 2015.

[79] A. Burattin, M. Cimitile, F. Maggi, and A. Sperduti, "Online discovery of declarative process models from event streams," *IEEE Trans. Serv. Comput.*, vol. 8, no. 6, pp. 833–846, Nov./Dec. 2015.

[80] Y. Zhang, R. Padman, and N. Patel, "Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data," *J. Biomed. Informat.*, vol. 58, pp. 186–197, Dec. 2015.

[81] F. Maggi, C. Di Francescomarino, M. Dumas, and C. Ghidini, "Predictive monitoring of business processes," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.*, 2014, pp. 457–472.

[82] C. Fernández-Llatas, J.-M. Benedi, J. M. García-Gómez, and V. Traver, "Process mining for individualized behavior modeling using wireless tracking in nursing homes," *Sensors*, vol. 13, no. 11, pp. 15434–15451, 2013.

[83] C. Fernández-Llatas, A. Lizondo, E. Monton, J. Benedi, and V. Traver, "Process mining methodology for health process tracking using real-time indoor location systems," *Sensors (Switzerland)*, vol. 15, no. 12, pp. 29821–29840, 2015.

[84] P. Helmering, P. Harrison, V. Iyer, A. Kabra, and J. Van Slette, "Process mining of clinical workflows for quality and process improvement," in *Proc. HIMSS Conf.*, 2012, pp. 1–7.

[85] Y. Zhang, O. Martikainen, P. Pulli, and V. Naumov, "Real-time process data acquisition with Bluetooth," in *Proc. 4th Int. Symp. Appl. Sci. Biomed. Commun. Technol.*, 2011, pp. 1–5.

[86] A. Rogge-Solti, R. Mans, W. M. P. van der Aalst, and M. Weske, *Repairing Event Logs Using Stochastic Process Models*. Potsdam, Germany: Univ. Potsdam, 2013.

[87] M. Grando, W. M. P. van der Aalst, and R. Mans, "Reusing a declarative specification to check the conformance of different CIGs," in *Proc. Int. Conf. Bus. Process Manage.*, 2012, pp. 188–199.

[88] R. Micio, F. Fontanili, G. Marquès, P. Bomert, and M. Lauras, "RTLS-based process mining: Towards an automatic process diagnosis in healthcare," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Oct. 2015, pp. 1397–1402.

[89] M. Grando, M. Schonenberg, and W. M. P. van der Aalst, "Semantic process mining for the verification of medical recommendations," in *Proc. Int. Conf. Health Informat. (HEALTHINF)*, Jan. 2011, pp. 5–16.

[90] M. A. Grando, M. H. Schonenberg, and W. M. P. van der Aalst, "Semantic-based conformance checking of computer interpretable medical guidelines," *Commun. Comput. Inf. Sci.*, vol. 273, pp. 219–232, Jan. 2011.

[91] Z. Huang, W. Dong, H. Duan, and H. Li, "Similarity measure between patient traces for clinical pathway analysis: Problem, method, and applications," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 1, pp. 4–14, Jan. 2014.

[92] Z. Huang, X. Lu, H. Duan, and W. Fan, "Summarizing clinical pathways from event logs," *J. Biomed. Informat.*, vol. 46, no. 1, pp. 111–127, 2013.

[93] E. Ramezani, D. Fahland, and W. M. P. van der Aalst, "Supporting domain experts to select and configure precise compliance rules," in *Proc. Lect. Notes Bus. Inf. Process.*, vol. 171. 2014, pp. 498–512.

[94] P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas, and N. Matsatsinis, "Supporting healthcare management decisions via robust clustering of event logs," *Knowl.-Based Syst.*, vol. 84, pp. 203–213, Aug. 2015.

[95] T. Neumuth, P. Liebmann, P. Wiedemann, and J. Meixensberger, "Surgical workflow management schemata for cataract procedures process model-based design and validation of workflow schemata," *Methods Inf. Med.*, vol. 51, no. 5, pp. 371–382, 2012.

[96] X. Xu, T. Jin, Z. Wei, C. Lv, and J. Wang, "TCPM: Topic-based clinical pathway mining," in *Proc. IEEE 1st Int. Conf. Connect. Health Appl. Syst. Eng. Technol. (CHASE)*, Jun. 2016, pp. 292–301.

[97] B. Xu, L. Xu, H. Cai, L. Jiang, and Y. Luo, "The design of an m-Health monitoring system based on a cloud computing platform," *Enterp. Inf. Syst.*, vol. 11, no. 1, pp. 17–36, Jun. 2017.

[98] M. Song, C. Günther, and W. M. P. van der Aalst, "Trace clustering in process mining," in *Proc. Int. Conf. Bus. Process Manage.*, 2008, pp. 109–120.

[99] C. Fernández-Llatas, B. Valdivieso, V. Traver, and J. Benedi, "Using process mining for automatic support of clinical pathways design.," in *Data Mining in Clinical Medicine*, vol. 1246, J. M. García-Gómez, Ed. New York, NY, USA: Springer, 2015, pp. 79–88.

[100] L. Bouarfa and J. Dankelman, "Workflow mining and outlier detection from clinical activity logs," *J. Biomed. Informat.*, vol. 45, no. 6, pp. 1185–1190, 2012.

[101] T. Blum, N. Padoy, H. Feußner, and N. Navab, "Workflow mining for visualization and analysis of surgeries," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 3, no. 5, pp. 379–386, 2008.

[102] F. M. Maggi, "Declarative process mining with the declare component of ProM," in *Proc. BPM*, vol. 32. 2013, pp. 1–5.

[103] D. Ferreira and C. Alves, "Discovering user communities in large event logs," in *Proc. Int. Conf. Bus. Process Manage.*, 2011, pp. 123–134.

[104] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. De Weerdt, and B. Baesens, "Monitoring care processes in the gynecologic oncology department," *Comput. Biol. Med.*, vol. 44, no. 1, pp. 88–96, 2014.

[105] T. Neumuth, P. Jannin, J. Schlomberg, J. Meixensberger, P. Wiedemann, and O. Burgert, "Analysis of surgical intervention populations using generic surgical process models," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, no. 1, pp. 59–71, 2011.

[106] S. Yang *et al.*, "A data-driven process recommender framework," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 2111–2120.

[107] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Syst. Appl.*, vol. 31, no. 1, pp. 56–58, 2006.

[108] M. Cho, M. Song, M. Comuzzi, and S. Yoo, "Evaluating the effect of best practices for business process redesign: An evidence-based approach based on process mining techniques," *Decision Support Syst.*, vol. 104, pp. 92–103, 2017.

[109] V. Augusto, X. Xie, M. Prodel, B. Jouaneton, and L. Lamarsalle, "Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation," in *Proc. Winter Simul. Conf.*, 2017, pp. 2135–2146.

[110] M. Zhou *et al.*, "Evaluation of trace alignment quality and its application in medical process mining," in *Proc. IEEE Int. Conf. Healthcare Inform.*, Aug. 2017, pp. 258–267.

[111] Z. Huang, W. Dong, L. Ji, C. He, and H. Duan, "Incorporating comorbidities into latent treatment pattern mining for clinical pathways," *J. Biomed. Inform.*, vol. 59, pp. 227–239, Feb. 2016.

[112] X. Xu, T. Jin, Z. Wei, and J. Wang, "Incorporating domain knowledge into clinical goal discovering for clinical pathway mining," *J. Healthcare Eng.*, vol. 2017, pp. 261–264, 2017.

[113] S. Yang *et al.*, "Medical workflow modeling using alignment-guided state-splitting HMM," in *Proc. IEEE Int. Conf. Healthcare Inform.*, Aug. 2017, pp. 144–153.

[114] X. Xu, T. Jin, Z. Wei, and J. Wang, "Incorporating topic assignment constraint and topic correlation limitation into clinical goal discovering for clinical pathway mining," *J. Healthcare Eng.*, vol. 2017, May 2017, Art. no. 5208072.

[115] Z. Huang, W. Dong, L. Ji, L. Yin, and H. Duan, "On local anomaly detection and analysis for clinical pathways," *Artif. Intell. Med.*, vol. 65, no. 3, pp. 167–177, 2015.

[116] R. D. Dewandono, R. Fauzan, R. Sarno, and M. Sidiq, "Ontology and process mining for diabetic medical treatment sequencing," in *Proc. Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, 2013, pp. 171–178.

[117] E. Rojas, M. Sepúlveda, J. Munoz-Gama, D. Capurro, V. Traver, and C. Fernandez-Llatas, "Question-driven methodology for analyzing emergency room processes using process mining," *Appl. Sci.*, vol. 7, no. 3, p. 302, 2017.

[118] F. Rismanchian and Y. H. Lee, "Process mining-based method of designing and optimizing the layouts of emergency departments in hospitals," *Heal. Environ. Res. Des. J.*, vol. 10, no. 4, pp. 105–120, 2017.

[119] L. Thomas, M. V. M. Kumar, and B. Annappa, "Recommending an alternative path of execution using an online decision support system," in *Proc. Int. Conf. Intell. Syst. Metaheuristics Swarm Intell.*, 2017, pp. 108–112.

[120] T. Sztyler, J. Carmona, J. Völker, and H. Stuckenschmidt, "Self-tracking reloaded: Applying process mining to personalized health care from labeled sensor data," in *Transactions on Petri Nets and Other Models of Concurrency XI*, vol. 9930. Berlin, Germany: Springer, 2015, pp. 160–180.

[121] S. Tsumoto, H. Iwata, S. Hirano, and Y. Tsumoto, "Similarity-based behavior and process mining of medical practices," *Future Generat. Comput. Syst.*, vol. 33, pp. 21–31, Apr. 2014.

[122] T. Wang, X. Tian, M. Yu, X. Qi, and L. Yang, "Stage division and pattern discovery of complex patient care processes," *J. Syst. Sci. Complex.*, vol. 30, no. 5, pp. 1136–1159, 2017.

[123] M. D. P. Villamil *et al.*, "Strategies for the quality assessment of the health care service providers in the treatment of Gastric Cancer in Colombia," *BMC Health Services Res.*, vol. 17, no. 1, p. 654, 2017.

[124] X. Xu, T. Jin, and J. Wang, "Summarizing patient daily activities for clinical pathway mining," in *Proc. Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, 2016, pp. 1–6.

[125] A. Dagliati *et al.*, "Temporal electronic phenotyping by mining careflows of breast cancer patients," *J. Biomed. Inform.*, vol. 66, pp. 136–147, Feb. 2017.

[126] H. Liu, Y. Liu, X. Li, G. Xie, and G. T. Lakshmanan, "Towards pathway variation identification: Aligning patient records with a care pathway," in *Proc. ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2014, pp. 1359–1368.

[127] Á. Rebuge, L. V. Lapão, A. Freitas, and R. Cruz-Correia, "A process mining analysis on a virtual electronic patient record system," in *Proc. IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 554–555.

[128] P. Rattanavayakorn and W. Premchaiswadi, "Analysis of the social network miner (working together) of physicians," in *Proc. Int. Conf. ICT Knowl. Eng.*, 2015, pp. 121–124.

[129] D. Forsberg, B. Rosipko, and J. L. Sunshine, "Analyzing PACS usage patterns by means of process mining: steps toward a more detailed workflow analysis in radiology," *J. Digit. Imag.*, vol. 29, no. 1, pp. 47–58, 2016.

[130] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker, "Application of process mining in healthcare—A case study in a Dutch hospital," in *Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol.*, vol. 25. 2008, pp. 425–438.

[131] L. Perimal-Lewis, D. Teubner, P. Hakendorf, and C. Horwood, "Application of process mining to assess the data quality of routinely collected time-based performance data sourced from electronic health records by validating process conformance," *Health Inform. J.*, vol. 22, no. 3, pp. 1017–1029, 2016.

[132] P. Delias, P. Manolitzas, E. Grigoroudis, and N. Matsatsinis, "Applying process mining to the emergency department," in *Encyclopedia of Business Analytics and Optimization*. IGI Global, Nov. 2016, pp. 168–178.

[133] F. Caron, J. Vanthienen, J. De Weerdt, B. Baesens, J. De Weerdt, and B. Baesens, "Beyond X-raying a care-flow: Adopting different focuses on care-flow mining gaining," in *Proc. 1st Int. Bus. Process Intell. Challenge (BPIC)*, 2011, pp. 1–11.

[134] C. Krutanard, P. Porouhan, and W. Premchaiswadi, "Discovering organizational process models of resources in a hospital using Role Hierarchy Miner," in *Proc. Int. Conf. ICT Knowl. Eng.*, Nov. 2015, pp. 125–130.

[135] H. Fei and N. Meskens, "Discovering patient care process models from event logs," in *Proc. 8th Int. Conf. Modeling Simulation (MOSIM)*, 2010, pp. 10–12.

[136] E. Kim *et al.*, "Discovery of outpatient care process of a tertiary university hospital using process mining," *Healthcare Inform. Res.*, vol. 19, no. 1, pp. 42–49, 2013.

[137] D. C. Kelleher, R. P. J. C. Bose, L. J. Waterhouse, E. A. Carter, and R. S. Burd, "Effect of a checklist on advanced trauma life support workflow deviations during trauma resuscitations without pre-arrival notification," *J. Amer. College Surg.*, vol. 218, no. 3, pp. 459–466, 2014.

[138] C. Webster, "EHR business process management: From process mining to process improvement to process usability," in *Proc. Healthcare Syst. Process Improvement Conf.*, 2012, pp. 1–6.

[139] E. Helm and F. Paster, "First steps towards process mining in distributed health information systems," *Int. J. Electron. Telecommun.*, vol. 61, no. 2, pp. 137–142, 2015.

[140] L. Perimal-Lewis, S. Qin, C. Thompson, and P. Hakendorf, "Gaining insight from patient journey data using a process-oriented analysis approach," in *Proc. 5th Australasian Workshop Health Inform. Knowl. Manage. (HIKM)*, vol. 129. 2012, pp. 59–66.

[141] K. Ganesha, R. Nagaraj, and M. D. Nayana, "Health care analysis for process deviation using alpha-fitness algorithm in process mining," *J. Theor. Appl. Inf. Technol.*, vol. 88, no. 3, pp. 511–516, 2016.
[142] L. Perimal-Lewis, D. De Vries, and C. H. Thompson, "Health intelligence: Discovering the process model using process mining by constructing start-to-end patient journeys," in *Proc. 7th Australasian Workshop Health Inform. Knowl. Manag. (Hikm)*, 2014, pp. 59–67.
[143] M. van Genuchten, R. Mans, H. Reijers, and D. Wismeijer, "Is your upgrade worth it? Process mining can tell," *IEEE Softw.*, vol. 31, no. 5, pp. 94–100, Sep./Oct. 2014.
[144] S. Suriadi, R. S. Mans, M. T. Wynn, A. Partington, and J. Karnon, "Measuring patient flow variations: A cross-organisational process mining approach," in *Proc. Asia–Pacific Conf. Bus. Process Manage.*, 2014, pp. 43–58.
[145] R. Mans, H. Reijers, M. van Genuchten, and D. Wismeijer, "Mining processes in dentistry," in *Proc. 2nd ACM SIGHIT Int. Health Inform. Symp.*, 2012, pp. 379–388.
[146] C. W. Günther, A. Rozinat, W. M. P. van der Aalst, and K. van Uden, "Monitoring deployed application usage with process mining," Tech. Rep. BPM Center Report BPM-08-11, 2008.
[147] M. Binder *et al.*, "On analyzing process compliance in skin cancer treatment: An experience report from the evidence-based medical compliance cluster (EBMC^2)," in *Advanced Information Systems Engineering*. 2012, pp. 398–413.
[148] X. Zhang and S. Chen, "Pathway identification via process mining for patients with multiple conditions," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2012, pp. 1754–1758.
[149] Z. Zhou, Y. Wang, and L. Li, "Process mining based modeling and analysis of workflows in clinical care—A case study in a chicago outpatient clinic," in *Proc. IEEE 11th Int. Conf. Netw., Sens. Control (ICNSC)*, Apr. 2014, pp. 590–595.
[150] A. Partington, M. T. Wynn, S. Suriadi, C. Ouyang, and J. Karnon, "Process mining for clinical processes: A comparative analysis of four Australian hospitals," *ACM Trans. Manag. Inf. Syst.*, vol. 5, no. 4, pp. 1–18, 2015.
[151] M. Lang, T. Bürkle, S. Laumann, and H. U. Prokosch, "Process mining for clinical workflows: Challenges and current limitations," in *Proc. MIE 21th Int. Congr. Eur. Fed. Med. Inform.*, 2008, p. 229.
[152] R. S. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, and A. J. Moleman, "Process mining in healthcare: Data challenges when answering frequently posed questions," in *Process Support and Knowledge Representation in Health Care*. 2013, pp. 140–153.
[153] R. Mans *et al.*, "Process mining techniques: An application to stroke care," *Stud. Health Technol. Inform.*, vol. 136, pp. 573–578, Dec. 2015.
[154] V. Sfyrla, J. Carmona, and P. Henck, "Process-oriented analysis for medical devices," in *Proc. 5th Workshop Med. Cyber-Phys. Syst. (MCPS)*, vol. 36. 2014, pp. 143–146.
[155] A. Dagliati *et al.*, "Temporal data mining and process mining techniques to identify cardiovascular risk-associated clinical pathways in Type 2 diabetes patients," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Jun. 2014, pp. 240–243.
[156] P. Jaisook and W. Premchaiswadi, "Time performance analysis of medical treatment processes by using disco," in *Proc. ICT Knowl. Eng. (ICT Knowl. Eng.)*, 2016, pp. 110–115.
[157] R. P. J. C. Bose, R. S. Mans, and W. M. P. van der Aalst, "Wanna improve process mining results? It's high time we consider data quality issues seriously," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, 2013, pp. 127–134.
[158] S. Quaglini, "Process mining in healthcare: A contribution to change the culture of blame," in *Proc. Int. Conf. Bus. Process Manage.*, 2008, pp. 308–311.
[159] S. van der Spoel, M. van Keulen, and C. Amrit, "Process prediction in noisy data sets: A case study in a Dutch Hospital," in *Proc. Int. Symp. Data-Driven Process Discovery Anal.*, 2012, pp. 60–83.
[160] T. Yampaka and P. Chongstitvatana, "An application of process mining for queueing system in health service," in *Proc. Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, 2016, pp. 1–6.
[161] W. Neamsirorat and W. Premchaiswadi, "Analysis of surgical event logs in a hospital by using heuristics miner technique," in *Proc. Int. Conf. ICT Knowl. Eng.*, 2015, pp. 105–109.
[162] E. Jaroenphol, P. Porouhan, and W. Premchaiswadi, "Analysis of the patients' treatment process in a hospital in Thailand using fuzzy mining algorithms," *Proc. Int. Conf. ICT Knowl. Eng.*, 2015, pp. 131–136.
[163] F. Mannhardt and D. Blinde, "Analyzing the trajectories of patients with sepsis using process mining," in *Proc. CEUR Workshop*, vol. 1859. 2017, pp. 72–80.
[164] K. Ganesha, K. V. Supriya, and M. Soundarya, "Analyzing the waiting time of patients in hospital by applying heuristics process miner," in *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, 2017, pp. 500–505.
[165] E. Rojas, M. Sepúlveda, J. Munoz-Gama, D. Capurro, and A. Seret, "Calculating emergency room metrics through process mining," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2016.
[166] S. Yoo *et al.*, "Conformance analysis of clinical pathway using electronic health record data," *Healthc. Inform. Res.*, vol. 21, no. 3, pp. 161–166, 2015.
[167] T. Sztyler, J. Völker, J. V. Carmona, O. Meier, and H. Stuckenschmidt, "Discovery of personal processes from labeled sensor data: An application of process mining to personalized health care," in *Proc. Int. Workshop Algorithms Theor. Anal. Event Data*, 2015, pp. 31–46.
[168] N. Garg and S. Agarwal, "Process Mining for Clinical Workflows," in *Proc. Int. Conf. Adv. Inf. Commun. Technol. Comput.*, 2016, pp. 1–5.
[169] K. Baker *et al.*, "Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy," *Int. J. Med. Inform.*, vol. 103, pp. 32–41, Jul. 2017.
[170] G. Riz, E. A. P. Santos, and E. de Freitas Rocha Loures, "Process mining to knowledge discovery in healthcare processes," *Adv. Transdisciplinary Eng.*, vol. 4, pp. 1019–1028, 2016.
[171] S. Hirano, "Temporal data mining on the stay time of outpatients and treatment processes," in *Proc. Data Mining Workshops*, 2016, pp. 527–530.
[172] K. Ganesha, M. Soundarya, and K. V. Supriya, "The best fit process model for the utilization of the physical resources in hospitals by applying inductive visual miner," *Proc. Int. Conf. Inven. Commun. Comput. Technol. (ICICCT)*, 2017, pp. 318–322.
[173] J. Cardoso, R. Mans, P. R. da Cunha, W. Van der Aalst, H. Berthold, and S. Quaglini, "A framework for next generation e-health systems and services," in *Proc. Amer. Conf. Inf. Syst. (AMCIS)*, 2015, pp. 1–11.
[174] C. Cordes, T. Vogelgesang, and H.-J. Appelrath, "A generic approach for calculating and visualizing differences between process models in multidimensional process mining," in *Proc. Int. Conf. Bus. Process Manage.*, vol. 202. 2015, pp. 383–394.
[175] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. Deweerdt, and B. Baesens, "A process mining-based investigation of adverse events in care processes," *Heal. Inf. Manage. J.*, vol. 43, no. 1, pp. 16–25, 2014.
[176] R. P. J. C. Bose and W. M. P. van der Aalst, "Analysis of patient treatment procedures," in *Bus. Process Manage. Workshops*, vol. 99. 2012, pp. 165–166.
[177] A. Perer and D. Gotz, "Data-driven exploration of care plans for patients," in *Proc. Extended Abstracts Hum. Factors Comput. Syst. (CHI)*, 2013, pp. 439–444.
[178] V. Kumar *et al.*, "Exploring clinical care processes using visual and data analytics: Challenges and opportunities," in *Proc. 20th ACM SIGKDD Conf. Knowl. Discovery Data Mining Workshop Data Sci. Soc. Good*, 2014, pp. 1–5.
[179] W. Zhou and S. Piramuthu, "Framework, strategy and evaluation of health care processes with RFID," *Decision Support Syst.*, vol. 50, no. 1, pp. 222–233, 2010.
[180] W. Zhou and S. Piramuthu, "Healthcare process mining with RFID," in *Proc. Int. Conf. Bus. Process Manage.*, vol. 43. 2010, pp. 405–411.
[181] T. Meneu, V. Traver, S. Guillén, B. Valdivieso, J. Benedí, and C. Fernández-Llatas, "Heart Cycle: Facilitating the deployment of advanced care processes," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 6996–6999.
[182] M. Peleg, P. Soffer, and J. Ghattas, "Mining process execution and outcomes—Position paper," in *Proc. Bus. Process Manage. Workshops*, vol. 4928. 2008, pp. 395–400.
[183] T. Vogelgesang and H.-J. Appelrath, "Multidimensional process mining: A flexible analysis approach for health services research," in *Proc. Joint EDBT/ICDT Workshops*, 2013, pp. 17–22.
[184] Z. Huang, Y. Bao, W. Dong, X. Lu, and H. Duan, "Online treatment compliance checking for clinical pathways," *J. Med. Syst.*, vol. 38, no. 10, p. 123, 2014.
[185] R. C. Basole *et al.*, "Understanding variations in pediatric asthma care processes in the emergency department using visual analytics," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 2, pp. 318–323, 2015.

[186] J. Moon and D. Kim, "Context-aware business process management for personalized healthcare services," in *Proc. Int. Conf. Services Comput. Context-Aware*, 2013, pp. 757–758.
[187] S. P. Detro, D. Morozov, M. Lezoche, H. Panetto, E. P. Santos, and M. Zdravkovic, "Enhancing semantic interoperability in healthcare using semantic process mining," in *Proc. Int. Conf. Inf. Soc. Techology (ICIST)*, 2016, pp. 80–85.
[188] C. Fernandez-Llatas, J. L. Bayo, A. Martinez-Romero, J. M. Benedí, and V. Traver, "Interactive pattern recognition in cardiovascular disease management. A process mining approach," in *Proc. IEEE-EMBS Int. Conf. Biomed. Heal. Inform.*, Feb. 2016, pp. 348–351.
[189] S. P. Detro, E. A. P. Santos, M. Lezoche, and E. de Freitas Rocha Loures, "Managing business process variability through process mining and semantic reasoning: An application in healthcare," in *Proc. Work. Conf. Virtual Enterprises*, 2017, pp. 333–340.
[190] H. Nguyen, M. La Rosa, M. Dumas, and A. H. M. ter Hofstede, "Stage-based business process mining," in *Proc. CEUR Workshop*, vol. 1848. 2017, pp. 161–169.
[191] S. Mertens, F. Gailly, and G. Poels, "Supporting and assisting the execution of flexible healthcare processes," in *Proc. Pervas. Comput. Technol. Healthcare (PervasiveHealth)*, 2015, pp. 334–337.
[192] A. Leontjeva, R. Conforti, C. Di Francescomarino, M. Dumas, and F. M. Maggi, "Complex symbolic sequence encodings for predictive monitoring of business processes," in *Business Process Management* (Lecture Notes in Computer Science), vol. 9253. Cham, Switzerland: Springer, 2015, pp. 297–313.
[193] R. Pérez-Castillo, B. Weber, I. G.-R. de Guzmán, M. Piattini, and J. Pinggera, "Assessing event correlation in non-process-aware information systems," *Softw. Syst. Model.*, vol. 13, no. 3, pp. 1117–1139, 2014.
[194] C. Di Francescomarino, M. Dumas, F. M. Maggi, and I. Teinemaa, "Clustering-based predictive process monitoring," *IEEE Trans. Serv. Comput.*, to be published.
[195] R. Pérez-Castillo, B. Weber, J. Pinggera, S. Zugal, I. G.-R. de Guzmán, and M. Piattini, "Generating event logs from non-process-aware systems enabling business process mining," *Enterprise Inf. Syst.*, vol. 5, no. 3, pp. 301–335, 2011.
[196] L. Canensi, G. Leonardi, S. Montani, and P. Terenziani, "A context-aware miner for medical processes," in *Proc. ICCBR Workshops*, 2016, pp. 192–201.
[197] C. W. Günther and W. M. P. van der Aalst, "Fuzzy mining—Adaptive process simplification based on multi-perspective metrics," in *Business Process Management* (Lecture Notes in Computer Science), vol. 4714. Berlin, Germany: Springer, 2007, pp. 328–343.
[198] I. Nassi and B. Shneiderman, "Flowchart techniques for structured programming," *ACM SIGPLAN Notices*, vol. 8, no. 8, pp. 12–26, 1973.
[199] *Unified Modeling LanguageTM (OMG UML), Infrastructure, Version 2.3*, OMG, OMG document formal/2010-05-03, 2010.
[200] A. J. M. M. Weijters, W. M. P. van der Aalst, and A. K. A. de Medeiros, "Process mining with the heuristicsminer algorithm," Tech. Univ. Eindhoven, Eindhoven, The Netherlands, Tech. Rep. WP 166, 2006, pp. 1–34.
[201] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Process and deviation exploration with inductive visual miner," in *Proc. CEUR Workshop*, vol. 1295. 2014, pp. 46–50.
[202] P. T. Hornix, "Performance analysis of business processes through process mining," M.S. thesis, Eindhoven Univ. Technol., Eindhoven, The Netherlands, Tech. Rep., Jan. 2007.
[203] *Home—Process Mining for Health Care*. Accessed: Feb. 26, 2018. [Online]. Available: http://www.processmining4healthcare.org/
[204] B. F. van Dongen. (2011). *Real-Life Event Logs—Hospital Log*. [Online]. Available: http://data.4tu.nl/repository/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54
[205] F. (Felix) Mannhardt. (2016). *Sepsis Cases—Event Log*. [Online]. Available: http://data.4tu.nl/repository/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460
[206] W. M. P. Van der Aalst, "Challenges in business process mining," in *Applied Stochastic Models in Business and Industry*, Citeseer, 2010.

TUGBA GURGEN ERDOGAN received the B.S. and M.S. degrees in computer engineering from Hacettepe University, in 2012, where she is currently pursuing the Ph.D. degree with a focus on process improvement by using process mining for healthcare processes. She is also a Research Assistant with the Department of Computer Engineering, Hacettepe University.

AYCA TARHAN has been involved in the area of software engineering and information systems for 15 years. She was a Lecturer of the Software Management Program, Informatics Institute, Middle East Technical University, from 2002 to 2006. She was a Visiting Researcher with the Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, from 2013 to 2015. She is currently an Assistant Professor with the Computer Engineering Department, Hacettepe University, Ankara, Turkey, where she is also the Head of the Health Informatics Sub-Division, Informatics Institute. She has led or been involved in projects originating from industry-academia collaborations on software quality evaluation, software process improvement, business process modeling, and system requirements elicitation.