Process Mining in Cardiology: A Literature Review

Guntur P. Kusuma[1,2*], Marlous Hall[3], Chris P. Gale[3], Owen A. Johnson[1]

[1] School of Computing, Faculty Engineering, University of Leeds, United Kingdom.
[2] School of Applied Science, Telkom University, Bandung, Indonesia.
[3] Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, United Kingdom.

* Corresponding author. Tel.: +447443085087; email: scgpk@leeds.ac.uk

Abstract: Process mining is a data analytics approach which has shown promising results in healthcare including the potential to improve the management of chronic diseases such as cardiovascular disease (CVD). CVD are one of the leading causes of premature death worldwide. Helping healthcare professionals develop a better understanding of how to improve CVD care pathways may result in better outcomes for patients. This paper describes the opportunities for process mining to support such improvements and provides a summary of process mining studies undertaken in the field of cardiology. The initial search found 1,597 potentially relevant papers which, after a systematic review, identified a final set of 32 relevant studies. These studies were analyzed across five themes: process and data types; research questions; techniques, perspective and tools; methodologies; limitations and suggested future work.

Key words: Cardiology, cardiovascular disease, care pathway, process mining.

1. Introduction

Process mining is an emerging data analytics approach which takes advantage of event logs, the automatically generated records recorded in the digital information systems of organizations. By analyzing event logs, the real patterns of activity within an organization can be discovered to gain insights for process improvement. In process mining [1], software algorithms are used to automatically extract process knowledge from these event logs to build process models. Processes can be modeled diagrammatically using nodes and arcs following standard notations such as Petri Net [2], [3] and the industry standard Business Process Modeling Notation (BPMN) [4], [5]. The underlying process model can be converted between different notations [6]. Additional information related to the events (including resources, cases, and other attributes) may available within the event log.

The application of process mining in healthcare environments can be used to provide new insights that facilitate the improvement of the patient treatment processes (often referred to as *care pathways*) [7]. In hospitals, the work of physicians, nurses, and technicians involves patient related administration, the recording of medical information, procedures and decisions. Processes such as requesting laboratory tests and imaging services have to be planned and coordinated and cooperation between people from different departments through the same or linked information systems is a vital component to delivering successful care [8]. This situation characterizes medical processes as complex within a dynamic and flexible environment. Process mining has significant potential to improve our understanding of these complex clinical processes. Within cardiovascular medicine the approach may help healthcare professionals develop

a greater understanding of how to improve CVD care pathways, which may result in improved clinical outcomes.

Cardiology is the field of medicine which diagnoses and treats disorders of the heart and blood vessels such as, coronary heart disease, cerebrovascular disease, rheumatic heart disease and other cardiac conditions [9]. CVDs are the main cause of disease-related death in the world. The top two cardiovascular diseases, coronary heart disease and stroke, accounted for a combined 15 million deaths in 2015 globally [10]. From an economic perspective, CVD causes a major burden on healthcare costs, productivity loss and in the care of people living with chronic conditions [11]. Reducing cost of CVD care processes and improving quality of care are a major challenge.

The main motivation of this paper is to summarize the previously published process mining studies in cardiology and to identify future research opportunities. In order to understand the extent of the use of process mining techniques in the field of cardiology we conducted a systematic literature review. The aims of this review are to: (i) explore the value of process mining in cardiovascular care process, (ii) present an overview of previous studies, (iii) provide a review of relevant process mining algorithms, techniques, methodologies and approaches; and (iv) identify research opportunities in this emerging field of study.

2. Process Mining in Cardiology

The aim of process mining is to extract process knowledge from event logs from all kinds of information systems [1]. In the field of healthcare, such system may be termed hospital information systems or Electronic Health Record (EHR) [12]. Typically, these information systems have a large number of database tables recording event data which can be exploited to improve care processes using process mining.

Our research identified four literature reviews of process mining in general healthcare. The review by [13] focus on process mining for clinical pathways and described four key trends for future research directions: (1) variants identification and analysis; (2) customization of clinical pathways; (3) self-learning improvement of clinical pathways; and (4) integrated medical process management. The review by [14] found increasing publications in process mining in healthcare and addressed validity issues in the previous review. The most detailed review is that conducted by Rojas *et al* [15]. This classified process mining case studies based on medical domains including cardiology. The study is extended in [16] which provides a bibliographic survey of process mining techniques, methodologies, algorithms and tools in healthcare. The review found 10 case studies of CVD covering cardiology and stroke. The work suggests opportunities for further research in process mining in the field of cardiology. Our paper provides an updated review of the specific literature of process mining in cardiology and the method follows those developed for a similar study in oncology [17], based on the following questions:

Q1: What specific cases have been investigated in cardiology?
Q2: What research methods have previously been used in process mining implementation in cardiology?
Q3: What are the results of the previous research in process mining in cardiology?
Q4: What are the future research opportunities in this field?

3. Research Method

3.1. Search Process

Wil van der Aalst coined the term "process mining" in 1998 and introduced a range of techniques and tools, we therefore limit the literature review papers to publications since 1998. Paper searching and selection was conducted using three different sources: first, by searching in electronic databases: IEEE Xplore and ACM DL for computer science domain, Pubmed for clinical domain, Elsevier/ScienceDirect, Springer, and Google Scholar for non-specified domains. A second source was the list of publications on the

main Process Mining community website [18]; and the third source was a review of the list of process mining papers by Rojas et al [16]. The query for searching papers in each electronic database was developed by following a method used previously in our literature search of process mining in oncology [17] with modifications appropriate to the disease. The final query is presented below:

```
("process mining" OR "data mining" OR "machine learning" OR "pathway analysis") AND ("event
log" OR "clinical pathway" OR "care pathway") AND (CVD OR cardiology OR cardiovascular OR
stroke OR "atrial fibrillation" OR "heart failure" OR "unstable angina")
```

The search terms of diseases were high level and, consequently, not inclusive of all CVDs. In PubMed, we used the term 'cardiovascular disease' as a Medical Subject Heading (MeSH) term and the explode feature was selected to include all specific diagnoses under CVD.

Duplicate papers were removed from the collection of papers retrieved and then selection was based on the following inclusion and exclusion criteria:

IC1: English language articles published from January 1998 until June 2nd, 2017.

IC2: The articles should involve process mining and related to cardiology or CVDs care process.

EC1: Articles which only mention process mining.

EC2: Non peer-reviewed papers.

Criteria were applied at three stages, an initial title review, an abstract review and, for those selected, a subsequent full-text review. Those which were reviewed as full-texts were further investigated for ancestors by looking at the papers from the same authors and assessing these additional papers against the inclusion criteria. Any papers from ancestor searching were then added into the final collection from full-text based checking and ready for thematic analysis. Fig. *1* shows the review process.

3.2. Quality Assessment

A series of activities were conducted to maintain the quality of the paper search process until selection process. The first author manually conducted search processes from three different sources. Searching in electronic databases was conducted on July 2nd 2017. For Google Scholar searches, incognito mode was used to avoid potential bias resulting from automatic recommendations regarding user's browsing history [19], [20]. Supervision and verification were performed by the co-authors including defining the search strategy, discussion and reviewing papers.
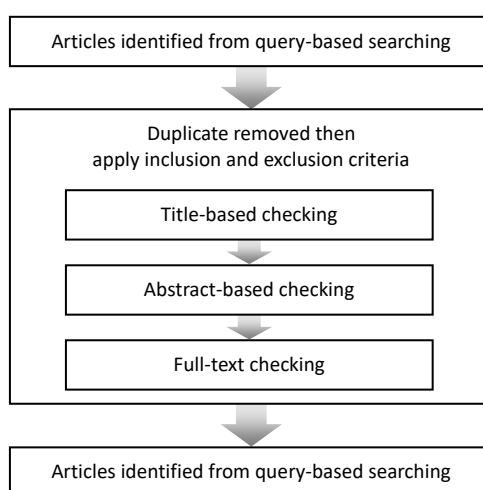

Fig. 1. The review process, following [17].

4. Results

Table 1 presents the number of papers from each source. On title review, the initial 1,597 papers were

reduced to 486 papers for abstract review. On abstract review, a further 342 papers did not meet the inclusion criteria, and the subsequent full text review excluded a further 112 papers, resulting in 32 papers that met our full search criteria.

Table 1. The Number of Retrieved Papers

#	Source		Number of papers
1	Electronic databases:		
	a.	ACM DL	10
	b.	IEEE Xplore	57
	c.	Science Direct	96
	d.	Springer	99
	e.	Pubmed	214
	f.	Google Scholar	1,100
2	processmining.org [18]		11
3	References of Rojas et al [16]		10
	Total		1,597

From the final paper collection, an in-depth ancestor investigation was conducted and no further relevant papers were identified, therefore the final number of papers for thematic analyses was 32. Figure 2 shows the number of published papers per year.
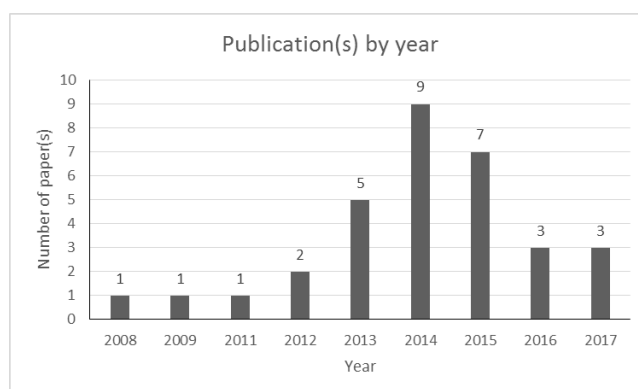

Fig. 2. Publication(s) by year.

Thematic analysis of the 32 papers was done following the five themes we had previously used in [17]: (1) process and data types; (2) research questions; (3) process mining perspectives, types and tools; (4) methodologies; and (5) limitations and future works.

All the selected papers were read by the first author for detailed analysis of the kind of processes that has been performed and data types from each paper. Research questions were reviewed to identify the problems process mining was being used to address. We classified the process mining method, the perspectives and types (following the terminology developed by [1]), together with the methodologies, limitations and discussion of future work to emphasize future research opportunities.

4.1. Thematic Analysis

4.1.1. Process and data types

The paper by [21] divided the applications of process mining to healthcare into two categories: medical treatment processes and organizational processes. From our 32 selected papers, there were seven studies which applied process mining in organizational processes [22]–[27], 24 studies of medical treatment

processes, and one study [28] which applied process mining to both organizational and medical treatment processes. There were 20 studies based on data from hospitals including these with the hospital's respective cardiology departments or stroke units. Of these 20 studies, there were four which used data from more than one hospital for comparative analyses [23], [29]–[31]. Three further studies used data from government-owned agencies which collect, host and analyze health data [22], [32], [33]. There were no papers using process mining to understand the disease trajectory of CVDs (also known as *disease progression*) although other data analytics methods have been used [34]. Fig. 3 shows stroke and unstable angina were the most common CVD diagnoses being used in studies. There was one study which involved three diagnoses [26]: unstable angina, pericarditis, and one non CVD-related diagnosis. The non-specific CVD category refers to those papers which did not explicitly report the CVD disease type.
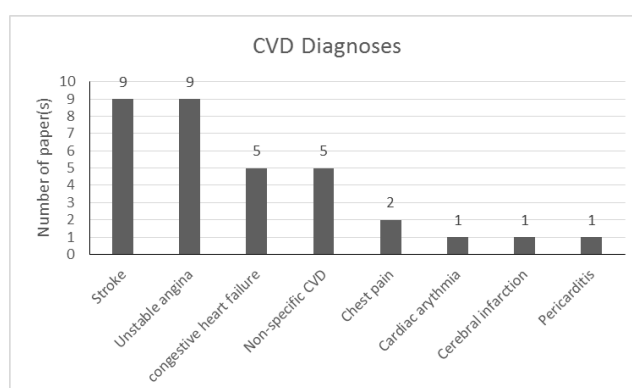

Fig. 3. Number of paper(s) based on CVD diagnoses.

4.1.2. Research questions

Each of the selected studies were designed to address problems within healthcare settings using process mining techniques or proposing new approaches in the field of process mining. Process mining studies in cardiology were dominated by proposals for new approaches and case studies undertaken to implement these approaches. Seven studies proposed new approaches to extract a process model from the event log: including the Parallel Activity-based Log Inference Algorithm (PALIA) [35], the log-tree algorithm and trace retrieval algorithm [36], probabilistic topic models [37], semi-automated implementation [38], Log-based clinical pathway summarization [39], Integer Linear Programming (ILP) [33], Genetic Process Mining (GPM) and Particle Swarm Optimization (PSO) [40]. In term of deviance detection of clinical pathway, there were three studies which proposed new approaches: (1) using Heuristics Miner and Taxonomic Distance [24]; (2) A* algorithm [41]; (3) by combining several methods Decision Tree, k-Nearest Neighbor (k-NN), and Iterative Pattern-based classification algorithm [31]. There were four studies which focus on prediction and/or recommendation: including predicting local anomalies in the care pathway [42], outcome prediction [43], process prediction [26], and a clinical pathway recommender for personalized treatment [32]. Furthermore, there was one study which proposed a new process mining methodology called Interactive Pattern Recognition (IPR) [28] which utilizes patient's lifestyle data and the behavior related to the treatments prescribed by physicians.

4.1.3. Process mining perspectives, types and tools

Process mining types are described into three categories: including discovery, conformance, and enhancement [1]. In addition, there are three different perspectives: control-flow, performance, and organizational. Most of the retrieved studies (31 out of 32) discussed at least one process mining perspectives. Only one study [28] discussed all three types, and they recommend process mining

methodology using process mining technology over an IPR framework. The author suggested the methodology be applied to monitor the continuous follow-up of patients based on the idea a committee of medical experts could use process mining based technologies to closely monitor the real-world picture of the clinical pathways in action.

In term of process mining perspectives, sixteen studies only discussed the control-flow perspective [22], [25], [29], [33], [36], [37], [39], [41], [42], [44]–[50], fourteen studies discussed control-flow and performance perspectives [23], [24], [26], [30]–[32], [35], [38], [40], [51]–[55], and one study discussed all three perspectives [28]. There is one further study which discusses methods to ensure the provision high quality data for process mining [27] and this study proposed solutions to reducing errors during data-entry.

Tools for performing process mining are varied, one open source option is the ProM toolkit [56] which is the most frequently used by researchers in the process mining field and the de-facto standard in the research community. For process mining studies in cardiology, the ProM toolkit has been utilized by seven studies [23]–[25], [29]–[31], [46] and two of them were combined with another tools such as WEKA, a data mining toolkit [57], or RapidMiner, a proprietary data science platform [58]. Another proprietary tool for process mining is Disco [59] of which the earlier version Nitro was used by one study [23].

4.1.4. Methodologies

Several methodologies to perform process mining projects have been proposed. van der Aalst *et al.* [1] proposed the L* life-cycle model which consists of five stages, a study from Caron *et al.* [60] proposed the Clinical Pathway Analysis Method (CPAM) a process mining methodology for clinical pathway analysis in medical setting. For quick process diagnostics, and in order to get a broad overview of the process within a short period of time, Bozkaya *et al.* [61] proposed the Process Diagnostic Method (PDM). Furthermore, Rebuge and Ferreira [62] extended the PDM methodology by adding particular characteristics of healthcare processes and named the model the Business Process Analysis in Healthcare (BPA-H) model. Subsequently, van Eck *et al.* [63] introduced PM2: Process Mining Project Methodology which covers a wide range of process mining and other analysis techniques, and is suitable for the analysis of both structured and unstructured processes.

Awareness of the range of methodologies to conduct process mining projects is low. In this review, there is only one paper [23] that reported following the L* life-cycle methodology. The study analyzed care pathway behaviors of four Australian public hospitals using four comparison points.

4.1.5. Limitations and future work

In the limitations and future work sections of the selected papers we found limitations related to data, technique, and team. Data limitations were identified in thirteen papers [28], [23], [22], [51], [46], [53], [35], [41], [54], [29], [30], [26], [27]. The issues included the noise of the data, missing timestamp information, and data quality control. Technical limitations were identified in 13 papers [28], [23], [51], [46], [41], [27], [44], [45], [32], [33], [37], [48], [55] including limitations in computer processing power, memory usage, bias associated with the mining algorithm, and simplifying the control flow due to computational complexity. There were no studies which reported limitation of team composition, however, 14 studies reported that they had involved clinician(s) to help the team with analysis or verification of results [28], [23], [46], [41], [54], [30], [27], [32], [33], [48], [55], [49], [50], [42].

A number of opportunities for further research have been proposed. Improvements in data were suggested by [22], [41], [29], [26], [32], [55], [50], [42], [43], [24], [25], [39] regarding additional health parameters of patients (including physiologic measures, laboratory test results, patient-reported health status), analysis of different CVD diagnoses, as well as the opportunity to include potentially related data from other departments. Technical aspects of future research suggestions were proposed by 19 studies, primarily suggestions to use different techniques or algorithms, remove limitations, improve classification

of patients, extending the methods, compliance checking with clinical guidelines, and building prototypes for further validation. Moreover, future research ideas surrounding the team were mainly focused on suggestions to work more closely with domain experts, agreement with another institution to get access for interconnected data, and working with system developers for integration of the proposed approaches into the healthcare system.

4.2. Process Mining Contributions in CVD

The literature review found less than half of the papers demonstrated clinician involvement in the team. Partington *et al.* [23] did identify significant variety in comparative pathways and presented to a clinical stakeholder group with positive feedback. Huang *et al.* [37] discovered care pathway patterns which were evaluated by hospital managers and clinical experts. They concluded that the discovered pathway patterns from clinical event logs supported care pathway redesign and improvement.

Recognition of the potential benefits from process mining has encouraged at least one initiative to provide higher quality data for better process mining analysis. The paper by [27] identified data quality issues which led to re-engineering the architecture of a registry system to ensure event sequences were recorded correctly in order to use process mining to improve stroke care pathways.

5. Conclusion

The implementation of process mining in cardiology has shown promising opportunities and the potential to support medical experts develop better understanding of care. Through multi-disciplinary collaboration with experts, more detailed understanding of the actual care pathways can be developed, interventions devolved and clinical outcomes improved.

This paper provides a summary of process mining studies in cardiology with respect to domain problems, processes and data types, perspectives, algorithms, and methodologies. The limitations and future work analysis highlights data quality as a major issue that still needs to be addressed. The low number of studies using formal process mining methodologies indicates limited awareness of these methods within the healthcare field. Further comparative studies between healthcare providers from different countries would provide better insights on international best practices, which has potential to improve the quality of care and outcomes for CVD patients on an international scale. There are many opportunities for process mining to support further research on CVD care pathways.

Acknowledgment

References

[1]   Aalst, W. M. P. (2011). Process mining: Discovery, conformance and enhancement of business process. *Process Mining* (Vol. 5). Springer Verlag Berlin-Heidelberg.

[2]   Peterson, J. L. (1977). Petri nets. *ACM Computing Surveys (CSUR)*, 9(3), 223–252.

[3]   Aalst, W. M. P. (1998). The application of petri nets to workflow management. *Journal of Circuits, Systems and Computers*, *8(1)*, 21–66.

[4]   OMG. (2011). About the Business Process Model and Notation Specification Version 2.0. Retrieved from the website: http://www.omg.org/spec/BPMN/2.0/

[5]   Aalst, W. M. P. (2009). Business process modeling notation. *Encyclopedia of Database Systems*, 293–294. Springer US.

[6]   Dumas, M., & García-Bañuelos, L. (2015). *Process Mining Reloaded: Event Structures as a Unified*

Representation of Process Models and Event Logs, 33–48. Springer, Cham.

[7] Mans, R. S., Schonenberg, M. H., Song, M., Aalst, W. M. P., & Bakker, P. J. M. (2008). Process mining in healthcare — A case study. *Proceedings of HEALTHINF 2008, International Conference in Health Informatics* (pp. 118–125). Portugal: INSTICC.

[8] Lenz, R., & Reichert, M. (2007). IT support for healthcare processes - premises, challenges, perspectives. *Data and Knowledge Engineering*, *61*(1), 39–58.

[9] *Cardiovascular Diseases (CVDs)*. (2017). *World Health Organization*. Retrieved from the website: http://www.who.int/cardiovascular_diseases/en/

[10] *Fact Sheet — The Top 10 Causes of Death*. (2017). *World Health Organization*. Retrieved from the website: http://www.who.int/mediacentre/factsheets/fs317/en/

[11] Wilkins, E., *et al*. (2017). European cardiovascular disease statistics 2017 edition. *European Heart Network, Brussels*, 192.

[12] Mans, R. S., Aalst, W. M. P., & Vanwersch, R. J. B. (2015). *Process Mining in Healthcare Evaluating and Exploiting Operational Healthcare Processes*.

[13] Yang, W., & Su, Q. (2014). Process mining for clinical pathway literature review and future directions. *Proceedings of 2014 11th International Conference on Service Systems and Service Management (ICSSSM),* (2010).

[14] Ghasemi, M., & Amyot, D. (2016). Process mining in healthcare: A systematised literature review. *Int. J. Electronic Healthcare*, *9(1)*, 60–88.

[15] Rojas, E., Arias, M., & Sepúlveda, M. (2015). Clinical processes and its data, what can we do with them? *Proceedings of the International Conference on Health Informatics* (pp. 642–647). SCITEPRESS - Science and and Technology Publications.

[16] Rojas, E., Munoz-Gama, J., Sepulveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, *61*, 224–236.

[17] Kurniati, A. P., Johnson, O., Hogg, D., & Hall, G. (2016). Process mining in oncology: A literature review. *Information Communication and Management (ICICM)*. Hatfield, UK: IEEE.

[18] Process Mining Group. (2017). Process Mining. Retrieved from the website: www.processmining.org

[19] Said, H., Al Mutawa, N., Al Awadhi, I., & Guimaraes, M. (2011). Forensic analysis of private browsing artifacts. *Proceedings of 2011 International Conference on Innovations in Information Technology* (pp. 197–202).

[20] Aggarwal, G., Bursztein, E., Jackson, C., & Boneh, D. (2010). An analysis of private browsing modes in modern browsers. *Proceedings of 19th USENIX conference on Security* (pp. 6–20).

[21] Kaymak, U., Mans, R., Steeg, De, T., & Dierks, M. (2012). On process mining in health care. *Proceedings of 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1859–1864).

[22] Augusto, V., Xie, X., Prodel, M., Jouaneton, B., & Lamarsalle, L. (2017). Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation. *Proceedings - Winter Simulation Conference* (pp. 2135–2146).

[23] Partington, A., Karnon, J., Wynn, M., Suriadi, S., & Ouyang, C. (2015). Process mining for clinical processes: A comparative analysis of four Australian hospitals process mining for clinical processes: A comparative analysis of four australian hospitals. *ACM Trans. Manag. Inform. Syst. Article*, *5(19)*.

[24] Montani, S., Leonardi, G., Quaglini, S., Cavallini, A., & Micieli, G. (2013). Mining and retrieving medical processes to assess the quality of care. *Lecture Notes in Computer Science,* 233–240. Springer, Berlin, Heidelberg.

[25] Quaglini, S. (2009). Process mining in healthcare: A contribution to change the culture of blame. *Lecture Notes in Business Information Processing*, *17*, 308–311.

[26] Spoel, S., Keulen, M., & Amrit, C. (2013). Process prediction in noisy data sets: A case study in a Dutch hospital. *Data-Driven Process Discovery and Analysis,*. 60–83. Springer Berlin Heidelberg.

[27] Lanzola, G., Parimbelli, E., Micieli, G., Cavallini, A., & Quaglini, S. (2014). Data quality and completeness in a web stroke registry as the basis for data and process mining. *Journal of Healthcare Engineering*, *5(2)*, 163–184.

[28] Fernandez-Llatas, C., *et al.* (2016). Interactive pattern recognition in cardiovascular disease management. A process mining approach. *Proceedings of 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 348–351).

[29] Suriadi, S., Partington, A., Wynn, M. T., & Karnon, J. (2014). Measuring patient flow variations: A cross-organisational process mining approach. *Lecture Notes in Business Information Processing.*

[30] Mans, R., & Schonenberg, H. (2008). Process mining techniques: an application to stroke care. *Studies in health technology and informatics*, *136*, 573–578.

[31] Rosa, L., *et al.* (2014). Mining business process deviance: A quest for accuracy. *Proceedings Lecture Notes in Computer Science*, *8841*, 436–445.

[32] Liu, R., Srinivasan, R. V., Zolfaghar, K., Chin, S. C., Roy, S. B., Hasan, A., & Hazel, D. (2015). Pathway-finder: An interactive recommender system for supporting personalized care pathways. *IEEE International Conference on Data Mining Workshops* (pp. 1219–1222).

[33] Prodel, M., Augusto, V., Xie, X., Jouaneton, B., & Lamarsalle, L. (2015). Discovery of patient pathways from a national hospital database using process mining and integer linear programming. 1409–1414.

[34] Jensen, A. B., *et al.* (2014). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nature Communications*, *5*, 1–10.

[35] Fernandez-Llatas, C., Valdivieso, B., Traver, V., & Benedi, J. M. (2015). Using process mining for automatic support of clinical pathways design. *Data Mining in Clinical Medicine*, *1246*, 79–88. New York: Humana Press.

[36] Luca, C., Giorgio, L., Stefania, M., & Paolo, T. (2016). Mining the log-tree of process traces: Current approach and future perspectives. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI* (Vol. 2016, pp. 310–316).

[37] Huang, Z., Dong, W., Ji, L., Gan, C., Lu, X., & Duan, H. (2014). Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics*, *47*, 39–57.

[38] Helmering, P. (2012). Process mining of clinical workflows for quality and process improvement. *Proceedings of Healthcare Information and Management Systems Society 2012 (HIMSS 2012)* (pp. 1–7). Citeseer.

[39] Huang, Z., Lu, X., Duan, H., & Fan, W. (2013). Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics*, *46(1)*, 111–127.

[40] Cheng, H. J., Ou-Yang, C., & Juan, Y. C. (2012). A hybrid evolutionary-based process mining technology to discover parallelism structures. *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management* (pp. 1573–1577).

[41] Yan, H., *et al.* (2017). Aligning event logs to task: Time matrix clinical pathways in BPMN for variance analysis. *Proceedings of International Conference on Biomedical and Health Informatics (BHI)* (pp. 1–7).

[42] Huang, Z., Juarez, J. M., Dong, W., Ji, L., & Duan, H. (2015). Predictive monitoring of local anomalies in clinical treatment process. *Lecture Notes in Artificial Intelligence*. Pavia, Italy: Springer, Cham.

[43] Huang, Z., Dong, W., Ji, L., & Duan, H. (2016). Outcome prediction in clinical treatment processes. *Journal of Medical Systems*, *40(1)*, 8.

[44] Yan, H., *et al.* (2013). Analyzing conformance to clinical protocols involving advanced synchronizations. *Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013*.

[45] Fernandez-Llatas, C., Meneu, T., Miguel Benedi, J., & Traver, V. (2011). Continuous clinical pathways evaluation by using automatic learning algorithms. *Healthinf 2011: Proceedings of the International Conference on Health Informatics* (pp. 228–234).

[46] Perimal-Lewis, L., Vries, D., & Thompson, C. H. (2014). Health intelligence: Discovering the process model using process mining by constructing start-to-end patient journeys. *Proceedings of the Seventh Australasian Workshop on Health Informatics and Knowledge Management* (vol. *153*, pp. 59–67).

[47] Huang, Z., Dong, W., Wang, F., & Duan, H. (2015). Medical inpatient journey modeling and clustering: a bayesian hidden markov model based approach. *AMIA Annual Symposium Proceeding* (pp. 649–658).

[48] Montani, S., Leonardi, G., Quaglini, S., Cavallini, A., & Micieli, G. (2014). Improving structural medical process comparison by exploiting domain knowledge and mined information. *Artificial Intelligence in Medicine*, *62(1)*, 33–45.

[49] Detro, S. P., Santos, E. P., & Loures, E. R. (2017). Managing business process variability through process mining and semantic reasoning : An application in healthcare. *IFIP Advances in Information and Communication Technology,* 333–340. Vicenza, Italy: Springer, Cham.

[50] Hamet, P., & Tremblay, J. (2017). Multi-level interactive medical process mining. *Metabolism: Clinical and Experimental*, *69(iii)*, S36–S40.

[51] Huang, Z., *et al.* (2014). Online treatment compliance checking for clinical pathways. *J Med Syst*, *38*.

[52] Huang, Z., Dong, W., Ji, L., He, C., & Duan, H. (2016). Incorporating comorbidities into latent treatment pattern mining for clinical pathways.

[53] Liu, H., 0007, Y. L., Li, X., Xie, G. T., & Lakshmanan, G. T. (2014). Towards pathway variation identification: Aligning Patient records with a care pathway. *Cikm*, 1359–1368.

[54] Lakshmanan, G. T., Rozsnyai, S., & Wang, F. (2013). Investigating clinical care pathways correlated with outcomes. *Business Process Management - Lecture Notes in Computer Science* (vol. 8094, pp. 323–338). Berlin: Springer, Berlin, Heidelberg.

[55] Huang, Z., Dong, W., Ji, L., Yin, L., & Duan, H. (2015). On local anomaly detection and analysis for clinical pathways. *Artificial Intelligence in Medicine*, *65*, 167–177.

[56] *ProM - Process Mining Toolkit*. (2010). *Process Mining Group*.

[57] Weka 3: Data mining software in java. (2017). Machine Learning Group at University of Waikato.

[58] *RapidMiner*. (2017). RapidMiner: Data Science Platform.

[59] Günther, C. W., & Rozinat, A. (2012). Disco: Discover your processes. *BPM 2012 Demonstration Track*, 46–50. Tall.

[60] Caron, F., Vanthienen, J., Vanhaecht, K., Limbergen, E., Deweerdt, J., & Baesens, B. (2014). A process mining-based investigation of adverse events in care processes. *Health Information Management Journal*, *43(1)*, 16–25.

[61] Bozkaya, M., Gabriels, J., & Werf, J. M. (2009). Process diagnostics: A method based on process mining. *Proceedings of 2009 International Conference on Information, Process, and Knowledge Management* (pp. 22–27). IEEE.

[62] Rebuge, Á., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, *37(2)*, 99–116.

[63] Eck, M. L., Lu, X., Leemans, S. J. J., & Aalst, W. M. P. (2015). PM 2 : A process mining project methodology. *Proceedings of International Conference on Advanced Information Systems Engineering (CAiSE) 2015. Lecture Notes in Computer Science* (vol. 909, pp. 297–313). Springer, Cham.

Guntur P. Kusuma received a BSc in informatics engineering from Telkom University (formerly STT Telkom), Bandung, Indonesia (2004) and an MSc in information systems from Institut Teknologi Bandung, Indonesia (2010). He is currently a PhD candidate in computer science at University of Leeds, UK. His research interests include health data analytics, process analytics including process and data mining, and the application of machine learning to data analytics.

Marlous Hall qualified in mathematics and statistics at the University of Sheffield and is a senior epidemiologist and Sir Henry Wellcome Fellow in Cardiovascular Epidemiology at the University of Leeds, UK. Dr Hall has expertise in the applications of advanced statistical techniques to large scale electronic healthcare data.

Chris P. Gale qualified in medicine at the London Hospital Medical College and is professor of Cardiovascular Medicine and Honorary Consultant Cardiologist at the University of Leeds, UK.

Owen A. Johnson received a BSc in management science from Loughborough University, UK (1985) and an MSc in information systems from University of Leeds, UK (2001). He is a senior fellow at the University of Leeds. His research interests are in e-health record process analytics including data and process mining, modeling, conformance, simulation and visualization.