Process Mining in Frail Elderly Care: A Literature Review

Nik F. Farid1,2, Marc De Kamps2 and Owen A. Johnson2
1School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia
2School of Computing, University of Leeds, Leeds, U.K.

Keywords: Frail Elderly, Frailty, Healthcare Service, Process Mining.

Abstract: Process mining has proved to be a valuable technique for extracting process knowledge from data within information systems. Much work has been conducted in applying process mining to domains such as logistics, banking, transportation and many areas of the government, including healthcare. Frail elderly people who have an increased risk of adverse outcomes are amongst the main users of healthcare services and understanding healthcare processes for the frail elderly is challenging because of their diverse and complex needs combined with an often high number of co-morbidities. This paper aims to provide an overview of work applying process mining techniques to improving the care of frail elderly people. We conducted a literature search using broad criteria to identify 1,047 potential papers followed by a review of titles, abstract and content which identified eight papers where process mining techniques have been successfully applied to the care of frail elderly people. Our review shows that, to date, there has been limited application of process mining to support this important segment of the population. We summarise the results based on five themes that emerged: types of source data and process; geographical location; analysis methodology; medical domain; and challenges. Our paper concludes with a discussion on the issues and opportunities for process mining to improve the care pathways for frail elderly people.

1 INTRODUCTION

The over 60s are the main users of health and social care (Oliver, 2009) and the number of people over 60 is expected to more than double from 962 million in 2017 to 2.1 billion by 2050 (UN, 2017). While many adults remain in good health well over 60 there is an increasing risk of frailty associated with aging. Frailty is a common clinical condition among the elderly and is often associated with stress caused by a cumulative decline in organ and clinical functions over time (Clegg et al, 2013). The progression of frailty can be seen as a continuous sequence from normal ageing, to pre-frail, frailty and finally to severe frailty (Lekan et al, 2017). An inability to maintain normal body functions may result in difficulty in managing with everyday activities independently (Xue, 2011) and increases the chance of hospitalizations, institutionalization, and adverse health outcomes such as falls, delirium and even mortality (Fried et al, 2001; Crandall et al, 2016; Minitski et al, 2001; Eeles et al, 2012). Frailty progression over time is poorly understood and there is growing interest in using electronic health record data to understand and identify the factors that influence this progression. One approach can be to visualize frailty progression using multi-dimensional data including patient characteristics, diagnoses and medication history from electronic health records (Chamberlain et al, 2016). Our interest is in the extent to which process mining of these records could help improve the understanding of frailty and the pathways of care designed to support the needs of the frail elderly population.

Electronic health records (EHRs) and other health information systems store data associated with the highly complex processes involved in delivering health care services to patients and this data can be used in process mining (Mans et al, 2008). Process mining is an emerging approach that combines business process management methods with data mining technologies (Aalst, 2011). Process mining aims to explore sequence of logged events over time and to abstract generalisations of the underlying process as process models. The approach can help analysts discover how processes are followed in practice, measure the conformance of real event logs to the ideal process to identify deviations, recommend
improvements to the process and monitor the effectiveness of interventions (Aalst, 2011). When applying process mining to electronic healthcare record data we treat the pathways of care as a type of business process (Mans et al., 2008).

The phrase process mining started to appear in the literature around 2006 based on the work of van der Aalst (Aalst et al., 2006) which applied data science to businesses process improvement efforts. A wide range of tools and approaches have subsequently been developed and applied to corporate organizations (Aalst et al., 2007; Aalst, 2015) and to healthcare (Partington et al., 2015; Weber et al., 2018). Process mining is generally based on data extracted from information systems but related work has used data from sensor devices that record daily activities to develop interventions (Fatima et al., 2013) and improve personalized care (Sztyler et al., 2015).

The Process Mining Manifesto (Aalst et al., 2011) proposed the L* life-cycle methodology for process mining projects. This approach consists of five different stages (labelled 0 to 4) which are: 0) plan and justify; 1) data extraction; 2) creating a control flow model and connecting an event log; 3) creating an integrated process model; and 4) providing operational support. More recently, van Eck et al. (2015) introduced an enhanced approach called Process Mining Project Methodology (PM²).

There are a number of recent literature reviews of process mining in healthcare (Rojas, 2016; Erdogan and Tarhan, 2018; Williams et al., 2018) and other reviews which focus on specific healthcare specialties such as cancer (Kurniati et al., 2016) and cardiovascular disease (Kusuma et al., 2018). However, until now there has been no literature review specifically examining process mining for the care of the frail elderly. This paper describes the approach we adopted to identifying literature relevant to process mining for frail elderly patients. Eight papers were found and are discussed here.

To date there has been limited application of process mining approaches to support this important segment of the population. Our paper aims to initiate discussion on the value and potential of process mining of frail elderly care pathways and identify opportunities to work in this field of study.

2 METHODOLOGY

The literature review was conducted in October 2018 to identify papers which describe the application of process mining to care involving frail elderly people.

2.1 Search Process

A four stage approach for search and selection was used (Figure 1). The first stage covered the search of papers from medicine, technology and engineering databases; PubMed, Medline, British Medical Journal Open, ACM DL, Elsevier, ScienceDirect, database systems and logic programming (DBLP), Web of Science and Google Scholar.

![Figure 1: Summary of search process.](image)

Keywords for process mining replicated those used in previous literature reviews (Kurniati et al., 2016; Kusuma et al., 2018). To ensure the search obtained papers that cover all relevant conditions related to older people a broad selection of keyword terms was applied following the Medical Subject Headings (MeSH) terms used in the PubMed and Medline databases. MeSH terms provide a comprehensive vocabulary for journal and articles indexing in medical studies used to facilitate searching. An additional eight keywords that are synonyms of the initial six MeSH terms keywords were obtained from the thesaurus website (https://www.thesaurus.com/browse).

The following keywords were used:

("Process mining" OR "workflow mining" OR "pathway mining") AND ("Frailty" OR "Elderly" OR "Older..."
We followed the review process adopted by Kurniati et al. (2016) where paper selections was conducted based on title, abstract and content checking. We checked our search results to make sure that they included the list of papers identified on processmining.org, the main process mining research community web site. A careful filtering approach was performed at each stage to ensure all potentially relevant papers were identified. For example, papers were passed to the next stage if insufficient information was provided in either the title or the abstract. Finally, an in-depth ancestor search was conducted to identify additional relevant papers from references in the final set.

2.2 Selection Process

The initial extracted papers were selected according to the set of inclusion and exclusion criteria outlined.

2.2.1 Inclusion Criteria

The following were the inclusion criteria when selecting papers for analysis has been applied to the frail elderly domain:

IC1: Articles published from year 1998
IC2: Publication language is English
IC3: Articles published are peer-reviewed or conference proceeding articles
IC4: Articles should include case studies where process mining technique has been applied into frail elderly domain

2.2.2 Exclusion Criteria

The following were the exclusion criteria applied when choosing extracted papers:

EC1: Duplicate publication of initial searched
EC2: Books
EC3: Articles discuss case studies other than in the frail elderly domain

2.2.3 Quality Assessment Process

The following activities were undertaken to ensure the quality of the search process. The paper extraction, analysis and evaluation were done manually by the first author. Google Scholar searches were performed in incognito mode to eliminate any bias that might arise from browsing history that might influence the search results. The review and the verification of the selected publications in the final stage was supervised by all co-authors.

3 RESULT

Eight papers were identified after a comprehensive search. This section describes the search process and an analysis of the selected papers. Table 1 provides an overview of the number of papers initially extracted and the final selection of published articles from the different sources. The initial search retrieved a total number of 1,091 papers from ten different search engines. We note that zero results were returned from BMJ Open, Elsevier/Springer, DBLP and processmining.org sources and most final selected papers were from the Google Scholar search.

<table>
<thead>
<tr>
<th>Sources</th>
<th>Initial Extraction</th>
<th>Final Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar</td>
<td>991</td>
<td>5</td>
</tr>
<tr>
<td>PubMed</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>BMJ Open</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>Elsevier/Springer</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ACM</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Web of Science</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Medline</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>DBLP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>processmining.org</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>1,091</td>
<td>8</td>
</tr>
</tbody>
</table>

The duplication step removed 44 papers and left 1,047 papers for the next stage. The inclusion and exclusion criteria were applied to the remaining papers to carefully select those papers that met with the aim of the work. Figure 1 details the number of papers excluded at each stage of the process based on the inclusion and exclusion criteria.

3.1 Characterisation of Element Analysis

The review identified eight papers and these are presented here. A complete list of reviewed papers are summarised in Appendix 1. A thematic analysis of the papers identified five themes: (1) data and process type; (2) geographic analysis; (3) methodology; (4) medical domain; (5) challenges that arose when conducting the study.

(1) Data and Process Type: the classification of process and data type followed the approach in Rojas
et al, (2016) that divided sources of data extraction by clinical or administrative healthcare dataset and the process type categorized as clinical treatment process or organizational process. However in this literature review, the most commonly extracted data were from sensors collected either from elderly behaviour living in smart environment (Vitali and Pernici, 2015; Tax et al, 2018); mined process data collected from an MIT smart home dataset (Tapia et al, 2004) and from nursing homes as in Llatas et al, (2011), Wolf et al (2013) and Munstermann et al, (2012) for patients who require ambulant services. Meanwhile Triki et al, (2015) analyse data from scenario generators for elderly people’s daily activities. There are only two papers that directly study electronic health record (EHR) data and one related to acute care and simple one day surgery (Najjar et al, 2018); while Conca et al, (2018) used administrative data, which identified different healthcare discipline roles. The nature of the data will determines the type of analysis possible. Najjar et al, (2018) investigated the clinical treatment while Conca et al, (2018) discussed the organizational process of collaboration between physicians, nurses and dietician. The other six papers (Munstermann et al, 2012; Llatas et al, 2011; Tax et al, 2018; Triki et al, 2015; Wolf et al, 2013; Vitali and Pernici, 2015) analysed processes which looked into daily activity of elderly people.


(3) Methodology: none of the papers described followed the process mining methodologies of the L* life cycle or PM2. All papers reported that they had developed their own methodology. It was evident that each had carried out process mining using clustering techniques from event logs generated from either EHRs or from sensor devices.

(4) Medical Domain: three different care processes within the medical domain have been investigated. Two papers (Llatas et al, 2011; Wolf et al, 2013) analysed processes to detect or reduce the progression of dementia. Najjar et al, (2018) obtained data from patients who suffered from heart diseases, whereas Conca et al, (2018) collected data from patients who had Type 2 diabetes mellitus. The other papers (Triki et al, 2015; Tax et al, 2018; Vitali and Pernici, 2015; Munstermann et al, 2012) did not describe the medical domain associated with their work.

(5) Challenges: the challenges could be categorized as technique, data and team limitations from sensor devices as in Kurniati et al, (2016). The papers working with sensor data experienced data quality issues related to granularity (Triki et al, 2015; Llatas et al, 2011; Wolf et al, 2013; Vitali and Pernici, 2015; Tax et al, 2018; Munstermann et al, 2012). The other limitation was data that was incomplete or inconsistent. This was the main issue in Conca et al, (2018). Najjar et al, (2018) suggested pre-processing of the extracted pathway data through multiple iterations to narrow the model to specific elements of interest. Conca et al, (2018) used a medical expert to help address their process mining challenges.

3.2 Evaluation of Experimental Result

Most of the case studied in the papers concerned traces collected from sensor devices (Wolf et al, 2013; Tax et al, 2016; Vitali and Pernici, 2015; Munstermann et al, 2012; Llatas et al, 2011). Two papers conducted experiments using EHR data (Najjar et al, 2018; Conca et al, 2018) and one used a scenario generator (Triki et al, 2015). There are two papers that applied a clustering algorithm to cluster set of events such as Hidden Markov Model (Najjar et al, 2018) and a combination of flow disintegration functionality and measuring dissimilarity based on heuristic topological editing distance (Conca et al, 2018). Llatas et al, (2011) used a workflow mining technique based on the Workflow Instance Acceptor Algorithm. Wolf et al, (2013) employed classification techniques on sensor data for activity recognition of traces before labelling the recorded activities to be modelled as processes. Triki et al, (2015) generated a variety of simulations to describe different models. Vitali and Pernici, (2015) constructed a methodology to understand the connection between process and events.

There are two papers that performed evaluation on the proposed algorithms. Wolf et al, (2013) showed that their subjective logic condition evaluator was outperformed by both a fuzzy event assignment method and a method that combines the FlowCon and FlexCon algorithms. Tax et al, (2016) describes the effect of proposed algorithms on the computation time. Munstermann et al, (2012) used a dynamic threshold for an F2-measure that combines both precision and recall to differentiate between the abnormal and normal days of activities. A set of abnormal traces were artificially created from normal sensor data by simulating four different types of

4 DISCUSSION

4.1 A Lack of Focus on the Frail Elderly

Most of the finally selected papers are very recent and this suggests that work using process mining for the care of older adults has started to gain attention only recently. We noted that, during the literature search, a number of papers on assisting the living of elderly people were also identified. Although many of these were excluded due to our focus on process mining techniques, it is evident that this is an important theme where process mining could help. Several of the process mining papers use sensor data to detect changes in behaviour concerning the daily activities (Munstermann et al, 2012; Triki et al, 2015; Vitali and Pernici, 2015). Smart living environments for elderly people can be fitted with home sensors, and monitoring using wearable devices can include not only the elderly but also their carers (Wolf et al, 2013) and this allows very rich data to be gathered for process analysis. For example, the data gathered and analysed in nursing homes conducted by Llatas et al, (2011) helped to detect abnormal behaviour by elderly people who suffer from dementia and cognitive impairment.

Some of the papers presented additional visualizations of processes. Tax et al, (2018) used a plot of the events from the log file as a dotted chart with coloured dots representing events over time. Conca et al, (2018) created process models from the collaboration patterns they found. Najjar et al, (2018) used the percentage of participation, referral and self-referral to describe significant patterns.

Our search has some limitations. The review was limited to papers on the recent status of process mining related to frail elderly patients. The inclusion and exclusion decisions were made by a single reviewer. Google Scholar matching for search criteria can have variability in its results. However the search was comprehensive and followed well established methods.

To date there has been no work that directly addresses the care and management of the condition of frailty in elderly patients. This despite the growing number of older adults globally and the growing recognition in the medical world that frailty demands specific attention as a complex set of diseases and needs. Kim and Jang (2018) have argued that, historically, medical attention has often focused on single diseases and the study and management of frailty introduces a more holistic approach centred on the patients, their experience of a range of often interconnected diseases and the specific needs that frail individuals have to maintain the best possible quality of life.

5 CONCLUSION

Process mining is an emerging field within data science and presents a fresh set of methods for process improvement in healthcare. Our group are developing methods for meaningful care pathway analysis using clinical reference groups and multidisciplinary domain experts to iteratively improve understanding of current pathways and identify potential improvements.

This paper has reviewed the small number of papers that use process mining of both sensor data and electronic health records for older adults and those with frailty. These demonstrate the potential for process mining to play an important role in improving our understanding of how best to manage the care of the frail elderly. However the opportunities for researchers to apply process mining to improve frailty care has not yet been explored in any detail. Our review of the literature shows that novel approaches are just starting to emerge. Our next step is to use primary care data from the UK NHS to examine how care pathways vary between different categories of frailty over time and the relationship this has to prescribing patterns.

ACKNOWLEDGEMENTS

The first author would like to thank the Ministry of Higher Education of Malaysia for the sponsorship and all the support given throughout the research. All authors gratefully acknowledge support from the Connected Health Cities (CHC) programme commissioned by the Northern Health Science Alliance (NHSA) and funded by the UK Department of Health.
REFERENCES


Aalst, W.M. van der., 2015. Extracting event data from databases to unleash process mining in BPM-Driving innovation in a digital world (pp. 105-128). Springer, Cham.


## APPENDIX

### Table 2.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Setting</th>
<th>Domain</th>
<th>Data Sources</th>
<th>Case Persp ective</th>
<th>Challenges</th>
<th>No of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Najjar <em>et al.</em></td>
<td>2018</td>
<td>Quebec, Canada</td>
<td>Heart failure</td>
<td>EHR of acute care and one day surgery</td>
<td>Yes</td>
<td>Data (pre-processing) and team limitation</td>
<td>180,027</td>
</tr>
<tr>
<td>Conca <em>et al.</em></td>
<td>2018</td>
<td>Chile, Santiago</td>
<td>Type 2 Diabetes</td>
<td>EHR of administrative data</td>
<td>No</td>
<td>Data (inconsistent and/or incomplete) and team limitation</td>
<td>2,843</td>
</tr>
<tr>
<td>Llatas <em>et al.</em></td>
<td>2018</td>
<td>Spain</td>
<td>Dementia</td>
<td>Nursing home</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>One</td>
</tr>
<tr>
<td>Tax <em>et al.</em></td>
<td>2018</td>
<td>Eindhoven, Netherlands</td>
<td>Elderly behaviour</td>
<td>Secondary data (Tapia &amp; <em>et al.</em>, 2004) of elderly living in smart environment</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>Four different datasets with number of events ranging from 220 to 1,962</td>
</tr>
<tr>
<td>Vitali and Pernici</td>
<td>2016</td>
<td>Milano, Italy</td>
<td>Elderly behaviour</td>
<td>Sensor collected from smart living environment of home care</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>One person with nine classification of events</td>
</tr>
<tr>
<td>Triki <em>et al.</em></td>
<td>2015</td>
<td>Toulouse, France</td>
<td>Elderly behaviour (outdoors)</td>
<td>Scenario generator of elderly e.g. outdoor activities</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>One person Several scenarios consist of three actors</td>
</tr>
<tr>
<td>Wolf <em>et al.</em></td>
<td>2013</td>
<td>Mainkofen, Germany</td>
<td>Dementia</td>
<td>Geriatric ward of nursing home</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>135 randomly selected person</td>
</tr>
<tr>
<td>Munstermann <em>et al.</em></td>
<td>2012</td>
<td>Duisburg, Germany</td>
<td>Elderly behaviour</td>
<td>Patient who uses ambulant service</td>
<td>Yes</td>
<td>Data limitation - (high granularity)</td>
<td>Five</td>
</tr>
</tbody>
</table>