

Evaluating the Process Capability Ratio of Patients' Pathways by the Application of Process Mining, SPC and RTLS

Sina Namaki Araghi, Franck Fontanili, Elyes Lamine, Nicolas Salatge, Julien Lesbegueries, Sebastien Rebiere Pouyade and Frederick Benaben

Industrial Engineering Center of IMT Mines Albi, University of Toulouse, Albi 81000, France

Keywords: Real-Time Location Systems, Process Mining, Statistical Process Control, Patient's Pathways.

Abstract: Learning how patients receive their health treatments is a critical mission for hospitals. To fulfill this task, this paper defines patients' pathways as business process models and tries to apply process mining, real-time location systems (RTLS), and statistical process control (SPC) as a set of techniques to monitor patients' pathways. This approach has been evaluated by a case study in a hospital living lab. These techniques analyze patients' pathways from two different perspectives: (1) control-flow and (2) performance perspectives. In order to do so, we gathered the location data from movements of patients and used a proof of concept framework known as R.IO-DIAG to discover the processes. To elevate the performance analyses, this paper introduces the process capability ratio of the patients' pathways by measuring the walking distance. The results lead to the evaluation of the quality of business processes.

1 INTRODUCTION

Healthcare organizations are facing the challenge of improving the quality of their services continuously (Rebuge and Ferreira, 2012). Consequently, these organizations need more knowledge about the executions of their processes. In the context of digital health, this paper emphasizes the possibility of enriching the knowledge of healthcare experts regarding to the real-time state of patients by analyzing their location data.

The movements of patients inside healthcare organizations as they are executing different tasks, could be treated as business process models and defined as *patients' pathways*. To explain this statement, we indicate that the authors in (Vanhaecht et al., 2010) defined the term of *care pathway* as a complex intervention for the mutual decision-making and organization of care processes for a well-defined group of patients during a well-defined period. It has also been signified that the aim of these care pathways is to enhance the quality of care across the continuum by improving risk-adjusted patients' outcomes, promoting patient's safety, increasing patient's satisfaction, and optimizing the use of resources. Based on these definitions and applied analyzing techniques in this paper, it could be inferred that patients' pathways can be categorized as care pathways too.

On the other hand, each patient's pathway con-

sists of the sequence of events, several steps, decision points, actors, and activities with the objective of delivering health care to the patients. Consequently, one could conceive these pathways as business process models. Similarly, Dumas et al mentioned in (Dumas et al., 2013), a business process could be seen as a mean that organizations use to deliver a service or product to clients and it is constructed from several decision points, sequences of activities, and actors' interactions.

As a service organization, a hospital needs to ensure about the quality of its services. Primarily, the quality of hospitals' services are evaluated by the duration of their processes and activities. In this paper, we evoke the idea of using the distance of patients' pathways as a variable for assessing the quality of services. There are several motives behind this choice. For instance, there are several risks regarding transferring a critically ill patient. These distances should be monitored precisely in order to be either minimized or stabilized. Needless to mention that the duration of processes could be correlated to the distance a patient should take inside the facility. Additionally, thanks to the proposed methods here, hospitals can acquire a **target value** for monitoring the distance of patients' pathways. This could lead to the detection of variations within patients' processes and consequently enhancing the quality of processes.

This paper uses DIAG approach in order to extract comprehensive knowledge from patients' pathways within hospitals. This approach which has been shown in figure 1 is an updated version of the previous one in (Araghi et al., 2018). It consists of seven main functions, that could be executed through four different states. These states are *Data state*, *Information state*, *Awareness*, and *Governance*. The seven functions in this approach would accomplish a goal which is transforming the location data of the patients into decisions. This approach is being developed within R.IO-DIAG application. Each state of the approach encompasses one or several functions. The first function is the *configuring the environment and the systems*. This function is concerned with installation of location systems and importing the primary information into the R.IO-DIAG application. These information are: identifying zones of the experiment, patient's information, and functions that could be executed in certain zones. The second function is the *location data gathering*. After obtaining the primary event logs (cf figure 2(a)) a series of interpreting rules have been devised in order to extract the corresponding activities at each zone of a hospital. A more detailed illustration of this function is included within (Namaki Araghi, 2018) article. After receiving the interpreted data, the list of un-linked activities would go within the *process modeling* function. Thanks to several process mining algorithms, a cartography of process models will be discovered and presented by declarative modeling languages or the OPC modeling language. These models and its metrics (time and distance) will be studied by the *process analyzing* function. After detecting the variations of process models, a series of cause and effect diagnoses will be performed in the *process diagnosing* to extract the cause of those variations. Finally, at *prognosing* function, thanks to discrete event simulation (DES) several scenarios will be generated for choosing the best scenario to improve the processes.

The next section provides an overview of process mining, RTLS and the related works. However, the third section presents the main focus of this paper which is on the analyzing function of DIAG approach, where we aim at evaluating the capability of processes in order to ensure the quality of services. The fourth section evaluates and illustrates the research work by an experiment. Finally, the last section concludes the paper.

2 BACKGROUND

There are two problematics that motivate experts to consider patients' pathways as business processes. These problematics are social and technical. Concerning the social aspect, Guardian issued a study in UK (Pinchin, 2015) based on data from 2013 that approximately 7 million hospital appointments have been missed due to patients being lost in the facilities and each costs on average £108. The impact of facility design on patients' safety is a serious issue and is related to the distance that patients walk in the facility. Therefore, monitoring the distance traveled by patients helps hospitals to reduce these errors. Although, measuring the exact value is only attainable by using proper techniques and technologies such as RTLS.

In the context of smart healthcare, the application of RTLS is growing rapidly as a prerequisite. These systems consist of three main components that communicate with each other thanks to radio-frequency signals. The first component is a tag which could be attached to each object that need to be located. Antennas are the second component. They find the position of the tags over the location area. The third component is a location engine. This software uses different types of algorithms and localization techniques to calculate the positions of tags. Some of these techniques are: triangulation, trilateration, Angle of Arrival (AOA), Time Difference of Arrival (TDOA), and Received Signal Strength Indicator (RSSI) (Cotera et al., 2016).

2.1 Process Mining

Recently, the application of process mining has been developed in the context of healthcare (Rojas et al., 2016) as an *evidence-based* methods to provide a fast, automatic and efficient way to map the processes. The concentration of process mining is on discovering knowledge from event logs which are registered in an information system. Nowadays, the knowledge generated by process mining is based on three activities: process discovery, conformance checking and enhancement of business processes (Aalst, 2016). Several researchers have defined process mining as a practice which is being derived from the field of data mining (Tiwari et al., 2008). However, Van der Aalst defines process mining as a bridge between process science and data science (Aalst, 2016). He identifies that process mining and data mining both start from data, however, data mining techniques are not typically "process-centric". Concerning process mining in healthcare, it mainly received attentions to-

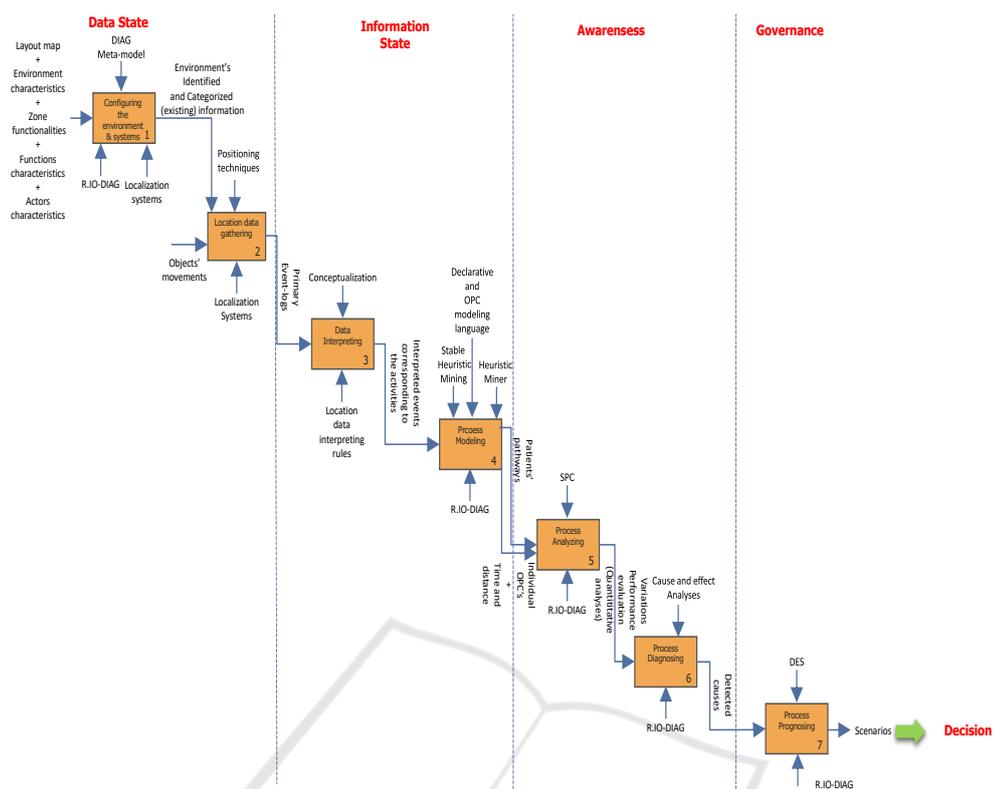


Figure 1: DIAG approach for extracting knowledge from location event logs.

wards designing and applying discovering algorithms. The enhancement and performance analysis are part of process mining techniques that have not evolved significantly mainly in the area of patients' pathways (Partington et al., 2015). The focus of this paper is on this gap. Therefore, the question for this research work is: how hospitals can measure the quality of patients' pathways?

Most of the applied methodologies in this area contain three main activities: (i)cleaning and preparing the data, (ii) using discovery algorithms and (iii) analyzing strategies (Rojas et al., 2016). As examples of these works, the authors in (Partington et al., 2015) performed a research in Australia to evaluate the performance of a care pathway in different hospitals. They analyzed the processes based on the number of patients who have been admitted, the throughput time of activities and the frequency of used healthcare procedures. They also provided an evaluation of process mining literature in healthcare. Authors highlighted a large gap among process mining activities. Based on their evaluation, only 3.5 % of research works are concerned with performance evaluation and enhancement of processes. They indicated that 82 % of works are oriented towards process discovery and 22% are related to conformance checking. The asso-

ciation of process mining and RTLS did not receive a lot of attention. However, it showed promising results in previous works. In this regard, the current approaches are highly dependent on the type of the configured location system. In (Fernandez-Llatas et al., 2015), Fernandez-Llatas et al. defined seven steps to transform location data in order to reach an enhancement phase of processes. These steps have been defined in the context of a tool which transforms the location data into the model-based analysis. They illustrated their research by the PALIA-ILS SUITE application (<http://pmuc.ing.puc.cl/>). In (Yang and Su, 2014), Yang and Su, reviewed the same mentioned works in (Fernandez-Llatas et al., 2015) and emphasized the need to improve the performance evaluation aspect of their work.

Most of these research works focused on the control-flow perspective. The main interest was to map the patients' pathways through process discovery. However, this paper expands this area by adding quantitative analysis. This would help experts to not only have a vision of the processes' executions, but also to acquire a real-time performance perspective of the patients' pathways by controlling the variations within processes.

3 METHOD

To fill the mentioned gaps, this work tends to extend the DIAG approach presented in (Araghi et al., 2018) by adding control charts and process capability analysis to a process mining application in order to analyze the real-time location data and quality of patients' pathways. This approach supports the performance and control-flow perspectives of a process mining application in healthcare.

3.1 Statistical Process Control(SPC)

SPC is a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variations in the process. SPC has already made its way into the healthcare sector (Thor et al., 2007). However, it has been used mainly for analyzing biological experiments but not in a sense of analyzing patients' pathways. The Shewhart control charts are one of the most sophisticated techniques of SPC. A typical control chart has three indicators which are known as center-line (CL), upper control limit (UCL), and lower control limit (LCL). These lines are being represented in the charts as horizontal lines. The UCL and LCL are indicating that if a process is in control, then nearly all of the sample points would fall between them. As long as all of the points of the samples are between LCL and UCL, no action is necessary. But, if a point falls beyond those limits, it could be inferred that the process is out of control due to the high level of variations. Therefore, some inspections on different aspects of the process are required. There are several types of control charts, such as \bar{x} chart, R-chart (range chart), P-chart, and C-chart. The application of each of these charts depends on the types of data and analysis that one could require. In this research work, \bar{x} and R charts will be used since the outcomes are based on two types of numeric data: Time and Distance of patients pathways. In followings, the mathematical principles for constructing the control limits will be presented.

Let $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$, be the average values of the samples, then the process average is the grand average ($\bar{\bar{x}}$):

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (1)$$

If the range of each sample equals to R then:

$$R = x_{max} - x_{min} \quad (2)$$

Now let R_1, R_2, \dots, R_m be the ranges of samples then the average range of process is:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (3)$$

The control limits to construct the \bar{x} and R-charts are as follows:

$$\begin{aligned} UCL_{\bar{x}} &= \bar{\bar{x}} + A_2 \bar{R} & UCL_R &= \bar{R} D_4 \\ CL_{\bar{x}} &= \bar{\bar{x}} & CL_R &= \bar{R} \\ LCL_{\bar{x}} &= \bar{\bar{x}} - A_2 \bar{R} & LCL_R &= \bar{R} D_3 \end{aligned} \quad (4)$$

The constant A_2, D_3 , and D_4 in (4) changes based on the size of the samples. These constants exist in most of the mathematical statistics references (Joglekar, 2005).

3.1.1 Process Capability Ratio

Another method to analyze the performance of processes is in terms of process capability ratio (PCR) or C_p . PCR is a statistical metric for making a comparison between the output of a process and with the specifications limits of the process. A process which all of its outcomes fall between the specification limits is considered as a capable process. For example, if hospitals define certain specifications as the expected length of stay for patients' pathways; C_p ratio helps them to evaluate their performance. PCR analysis could be defined by adjusting two new limits as Upper Specification Limit (USL) and Lower Specification Limit (LSL). These limits are specifications relevant to the quality characteristics that one desires to analyze (such as reliability of a process). In this research work, the USL and LSL could be defined manually by the healthcare experts, or be calculated by analyzing the distribution of the gathered data. Equation (5) shows the mathematical expressions to calculate C_p . Where σ is the standard deviation of samples. Equation (6) shows how to calculate the **specification limit** regarding to the distribution of the data.

$$C_p = \frac{USL - LSL}{6\sigma} \quad (5)$$

$$USL = \bar{\bar{x}} + 3\sigma$$

$$LSL = \bar{\bar{x}} - 3\sigma \quad (6)$$

C_p could have three states, which help experts to analyze the capability of the As-Is processes: If $C_p < 1$; it means that process is using up more than 100 % of the tolerance band which means the process is not capable to provide the desired outcome continuously. If $C_p = 1$; it means that process is using 100% of its tolerance band. This implies that process may provide some undesirable outcomes, but statistically is predictable and capable of satisfying the current specification defined by the organization. If $C_p > 1$; the process is using much less than 100 % of its tolerance band. As a result, relatively few undesirable outcomes could be produced by the process. These analyses can be seen concretely by the description of the study case in the next section.

4 CASE STUDY AND THE EXTRACTED RESULTS

To validate this approach, an experiment in a hospital living lab has been conducted. In this regard, R.IO-DIAG application has been developed within R.IO SUITE platform (<https://research-gi.mines-albi.fr/display/RIOSUITE/R-IOSuite+Home>) as a proof of concept, and for the illustration of the results. This tool receives the location data, refines them and generates business process models with the addition of robust statistical analyses and diagnoses. The results of this study case are the outcomes of two core functions of DIAG approach: process modeling and analyzing. This experiment had been carried out during ten days, and 150 patients have been monitored. They have been divided into 10 samples with the unique size of 15. These patients had similar profiles regarding which health care procedure they required. Event logs from the RTLS have been generated by the location engine with the JSON format. Each event contains several information such as event ID, time-stamp of entering a zone or exiting one, location data and other complementary data such as room temperature, humidity and tag's battery level. Figure 2 shows an example of patients tracking view and the primary event logs. To perform process discovery, a set of complex event processing rules have been defined based on DIAG reference model (Namaki Araghi, 2018). This reference model helps to identify which types of activities could happen in a certain zone. The details of these rules is within location data interpreting function of DIAG, which are evoked in (Namaki Araghi, 2018). After gathering the location data of pathways, the modeling function is the next step to gain a view on the way processes are being executed.

4.1 A Control-flow Perspective

The first step to obtain a model-based analysis of patients' pathways is to perform automatic process discovery. There are several existing process discovery algorithms (Augusto et al., 2017). DIAG uses stable heuristic mining algorithm since it can be applied for a collective modeling of pathways (describing the algorithm in detail is beyond the limits of this article). As it has been shown in figure 3 a qualitative analysis can be extracted by illustrating patients' pathways as process models. After analyzing the model, it has been indicated that most of the cases in the model have ignored one important step in their processes. This step was the "UROLOGY_checkout". The primary diagnosis indicated this could be due

to the long waiting period at the "waiting area for checkout_UROLOGY". Such qualitative analysis can be more useful for the organizations if it is being supported by quantitative analyses too. Existing mathematical analyses within process mining tools do not evaluate the quality of processes. The performance perspective section of this paper targets this problematic.

4.2 The Performance Perspective

It is obvious that the pathways could differ for each patient. However, it is a non-trivial task for hospitals to offer a stable service quality to patients. As Montgomery emphasizes in (Montgomery, 2007), the quality of services which patients are conceiving is directly influenced by the stability of processes. During this experiment, we encountered that patients did not have problems with the waiting times or length of pathways in particular but they expressed their dissatisfaction when they had to spend longer periods of time than expected in order to perform certain tasks. Additionally, they had difficulties for finding the rooms in the hospital. Therefore, we proposed to use process control charts to enhance the performance by discovering and diagnosing the variations.

Since the control limits on the \bar{x} -chart (Figure 4b) depend on the process variability, it is ideal to begin with the R -chart (Figure 4a). Unless the process variability is in control, these limits (in \bar{x} -chart) will not have much meaning (noa,) and the process is already out-of-control.

The R -chart helps to ensure the stability of the extracted data. The CL is the average of all the subgroups' ranges. The other control limits are set by a distance of 3σ (standard deviation) above and below the center line. These thresholds define the limits for expected variations in the subgroups ranges. The figure 4a shows R -chart for analyzing the range of distances for patients' pathways. Based on the stability in ranges, we may now construct the \bar{x} -chart presented in figure 4b. This figure presents the instability of the average traveled distance by patients. The red points in the \bar{x} -chart indicate that within three days of the experiment, there were some assignable causes that affect the distance of pathways for the patients. These variations in the process have been investigated by the help of healthcare professionals. Because of an increase in the number of admitted patients on those days, the department had the lack of resources to perform medical examination for all the patients. Therefore, nurses asked some patients to report their problems to other exam rooms in the hospital. This caused the traveled distance of some pathways exceed

```

{
  "event_id": 195284,
  "topic": "measurement.update.position",
  "attributes": {
    "z": "Salle d'attente ACCUEIL",
    "lon": 2.17933787647385,
    "lat": 43.9218877209485,
    "distance": 281719.68305639544,
    "x": 10.089952235424914,
    "y": 2.4143585199153503
  },
  "date_time": "2017/09/14,10:50:14",
  "eui": "0xd0d01e1a01000029"
},
{
  "event_id": 195282,
  "topic": "measurement.update.zone",
  "attributes": {
    "zone": "Salle d'attente ACCUEIL",
    "in_zone": false
  },
}

```

(a) An example of the primary event log.



(b) A view on the movement of the patients.

Figure 2: Primary results of tracking patients by RTLS.

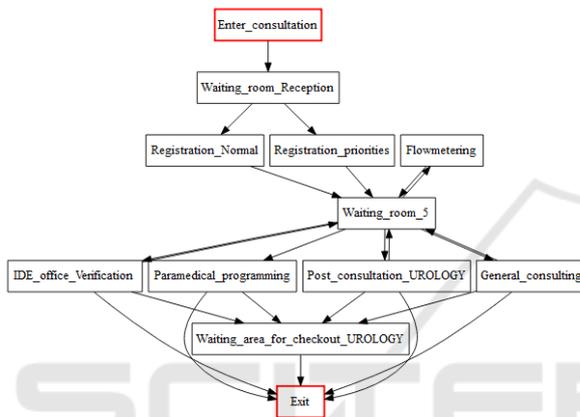


Figure 3: The extracted process model by R.IO-DIAG.

from the norm value. Consequently, this issue led to an instability in the process.

The length of patient's pathway is a practical indicator for the capability of a hospital in providing efficient services. The reason could be seen as the effect on the length of stay in the hospital. Also, it influences the efficiency of providing emergency treatments to the exact location of a patient with a critical status. Thus, this research work seeks different means to analyze the capability of processes based on the defined CTQ (Critical To Quality) specifications. These specifications (USL, target, LSL) either are defined by the health professionals or could be adapted from the way patients are moving in the hospital. Figure 5 presents the results on process capability ratio (C_p) of the experiment.

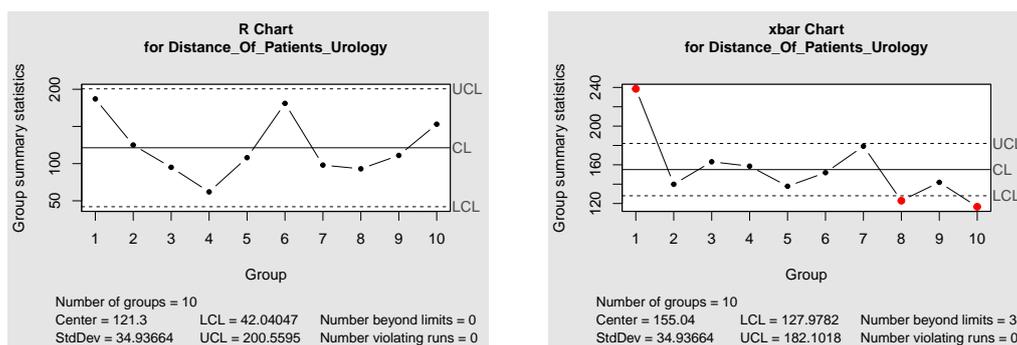
There are several points that should be inferred from the PCR analysis; first of all, this chart consists of two specification limits that show the possible margins for the length of pathways. As it has been shown in figure 5, there are processes that have the distance higher than USL and lower than LSL. These processes are not satisfying the targeted specifications. The hospital's experts have defined an USL of 220 meters

provisionally. The LSL has been adapted automatically from the current state of processes with the defined formula in 6. The C_p is less than 1 which shows that processes are being executed in an inefficient way and they are not meeting the expectations. As it has been shown, approximately 5 % of patients are walking more than the upper specification limit. Note that C_{pl} , and C_{pu} , demonstrate the performance of processes near to the lower specification limit and the upper specification limit. C_{pk} illustrates the capability of the process when the average of the samples are not well-centered. C_{pm} measure could be useful if we want to use the average value as the target to reach. C_{pm} could be applied to monitor the difference between the target value and the average of the results. For instance, the target in this chart has been identified as 135.115 meters. However, the average length of pathways is different ($Center = 155.04$). As process average moves off of the target, C_{pm} grows greater. This would help to further analyses and diagnoses. The same analysis are applicable and have been done for the duration of processes which their explanations are beyond the scope of this article. Relevant to the "enhancement" activities of process mining paradigm, the presented methods can provide applicable means in order to detect the process variations from the outcomes of the processes.

5 CONCLUSIONS

This paper aimed at addressing several issues regarding monitoring of patients' processes in healthcare organizations.

- Firstly, it emphasizes the need to consider patients' pathways as business processes. Then it provides technical and technological means to monitor these pathways. The advantage of this approach is related to a shorter period of data extraction and the accuracy level of extracted data.



(a) Analyzing the variations of pathways' length

(b) The \bar{x} -chart to monitoring the stability of patients' pathways

Figure 4: Control charts analysis related to the distance of pathways.

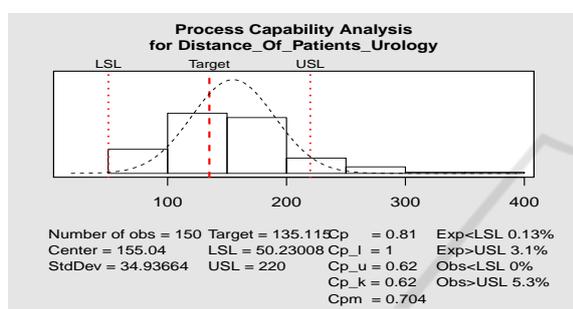


Figure 5: Evaluation of the process capability by the length of patients' pathways.

- Moreover, the presented approach provides a new sight to evaluate the capability of patients processes based on the traveled distance of their pathways. To the best of our knowledge regarding the literature, analyzing the stability and capability of patients' processes by the traveled distance of their pathways have never been done.
- Most of process mining research works are focusing on a challenge of how to discover a proper and fitting process model, which is an inevitable task. However, in this paper we tried to give an extra attention to the enhancement phase of process mining which we believe is necessary in order for process mining to become more applicable for healthcare organizations.

In this paper care-flow and performance perspectives are provided thanks to the framework of R.IO-DIAG which is exclusively being used to extract business process models from location data.

As it has been introduced, R.IO-DIAG applies different statistical process monitoring techniques which are not being used in other process mining packages. Thanks to these techniques, it is possible to monitor and reduce the variability of patients' pathways, which could lead to the quality improvement of pro-

cesses.

After analyzing the patients' pathways and highlighting the shortcomings of the processes, the future perspective of this research work is to provide an automatic diagnosis of the processes.

REFERENCES

Aalst, W. M. P. v. d. (2016). *Process Mining: Data Science in Action*. Springer. Google-Books-ID: hUEGDAQAQBAJ.

Araghi, S. N., Fontanili, F., Lamine, E., Tancerel, L., and Benaben, F. (2018). Applying Process Mining and RTLS for Modeling, and Analyzing Patients' Pathways. pages 540–547.

Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F. M., Marrella, A., Mecella, M., and Soo, A. (2017). Automated Discovery of Process Models from Event Logs: Review and Benchmark. *arXiv:1705.02288 [cs]*. arXiv: 1705.02288.

Cotera, P., Velazquez, M., Cruz, D., Medina, L., and Bandala, M. (2016). Indoor Robot Positioning Using an Enhanced Trilateration Algorithm. *International Journal of Advanced Robotic Systems*, 13(3):110.

Dumas, M., Rosa, M. L., Mendling, J., and Reijers, H. A. (2013). Introduction to Business Process Management. In *Fundamentals of Business Process Management*, pages 1–31. Springer, Berlin, Heidelberg.

Fernandez-Llatas, C., Lizondo, A., Monton, E., Benedi, J.-M., and Traver, V. (2015). Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems. *Sensors*, 15(12):29821–29840. WOS:000367539100004.

Joglekar, A. M. (2005). Appendix G: Table of Control Chart Constants. In *Statistical Methods for Six Sigma: In R&D and Manufacturing*, pages 311–311. John Wiley & Sons, Inc.

Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.

- Namaki Araghi, S. (2018). *ACS/IEEE International Conference on Computer Systems and Applications AICCSA*.
- Partington, A., Wynn, M. T., Suriadi, S., Ouyang, C., and Karnon, J. (2015). Process mining for clinical processes: A comparative analysis of four Australian hospitals. *ACM Transactions on Management Information Systems*, 5(4):19:1–19:18.
- Pinchin, J. (2015). Getting lost in hospitals costs the NHS and patients. *The Guardian*.
- Rebuge, I. and Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, 37(2):99–116.
- Rojas, E., Munoz-Gama, J., Sepúlveda, M., and Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, 61:224–236.
- Thor, J., Lundberg, J., Ask, J., Olsson, J., Carli, C., Härenstam, K. P., and Brommels, M. (2007). Application of statistical process control in healthcare improvement: systematic review. *Quality & Safety in Health Care*, 16(5):387–399.
- Tiwari, A., Turner, C. J., and Majeed, B. (2008). A review of business process mining: state-of-the-art and future trends. *Business Process Management Journal*, 14(1):5–22.
- Vanhaecht, K., Sermeus, W., Peers, J., Lodewijckx, C., De-neckere, S., Leigheb, F., Decramer, M., and Panella, M. (2010). The impact of care pathways for exacerbation of Chronic Obstructive Pulmonary Disease: rationale and design of a cluster randomized controlled trial. *Trials*, 11:111.
- Yang, W. and Su, Q. (2014). Process mining for clinical pathway: Literature review and future directions. In *2014 11th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–5.